PHYSIOLOGICAL BASIS FOR PREDATOR ESCAPE IN *SALMONELLA*

by

Kristen A. Butela

B.S. in Biology, Seton Hill University, 2003

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Kristen A. Butela

It was defended on

June 14, 2011

and approved by

Graham Hatfull, PhD; Professor, Department of Biological Sciences

Susan Kalisz, PhD; Professor, Department of Biological Sciences

Valerie Oke, PhD; Lecturer, Department of Biological Sciences

James Carroll, PhD; Senior Staff Scientist, National Institutes of Health

Dissertation Advisor: Jeffrey G. Lawrence, PhD; Professor, Department of Biological Sciences

PHYSIOLOGICAL BASIS FOR PREDATOR ESCAPE IN *SALMONELLA*

Kristen A. Butela, PhD

University of Pittsburgh, 2011

Salmonella enterica, a major causative agent of gastrointestinal illness, exhibits a host-specific pattern of infection, with certain serovars predominantly infecting particular hosts. Extensive variation is observed at the *Salmonella rfb* locus, which makes up serovar-defining O-antigen. Unlike other pathogens, this diversity cannot be explained by selective pressure from the host immune system. Here, I implicate the O-antigen to the physiological basis for escape from protozoan predators. These predators have differential feeding preferences on *Salmonella* and may be responsible for maintaining O-antigen diversity, controlling which serovars are able to survive predation to potentially cause disease. I demonstrated that the O-antigen plays a strong role in mediating predator escape and uncovered a trade-off that may exist between O-antigen identity and chain length regulation in response to the dual selective pressures of evading host intestinal predators and successful interaction with the host immune system. To complete these experiments, I developed two new techniques: a) genetic manipulation of non-Typhimurium *Salmonella* and b) multicolor flow cytometry for assessment of microbes in natural, complex environments. These results link variation at virulence loci to environmental selective pressures other than the host immune system and provide an explanation for the role of the *rfb* locus in the fragmented speciation process in *Salmonella*.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Previously published material has been incorporated into this document:

Chapter 1

Butela, K.A. and J.G. Lawrence. (2010). *Population genetics of Salmonella: Selection for antigenic diversity*, in *Bacterial Population Genetics in Infectious Disease*, D.A. Robinson, D. Falush, and E.J. Feil, Editors. John Wiley and Sons: Hoboken, NJ. p. 287-319. Available online.

Acknowledgements:

First and foremost, I would like to thank my dissertation advisor, Dr. Jeffery Lawrence, for pushing me to my absolute limits of intellectual development and never allowing me to settle for anything less than my absolute best work. Jeffrey provided an environment that challenged me to do work that changes the way people think about science. While there certainly were tough times, the Lawrence lab was always a place in which I could be myself. Although I will rarely admit it, I believe that Jeffery's sense of appreciating the subtle joy of proving outdated ideas wrong through science has certainly become for me one of the more fun aspects of research.

The broad scope of my work required a dissertation committee comprised of individuals with great expertise in genetics, microbiology, evolution, ecology, and immunology; thus, I must acknowledge the help of Dr. Graham Hatfull, Dr. Valerie Oke, Dr. Susan Kalisz, and Dr. Jay Carroll for all of their thoughtful advice and support throughout my graduate career. I would also

like to acknowledge the efforts of the staff of the Department of Biological Sciences, in particular Cathy Barr, Natalie Marinzel, Crystal Petrone, Kathy Hoffman, Patty Henry, Lynn Weber, Meredith Lindelof, Deanna DeKlaven, Ellie Caligiuri, Roxanne Scarano, Hernan Brizuela, Dr. Eric Polinko, Pat Dean, Tom Harper, Jay Bashor, Dave Malicki, and Frank Vincunas, for taking care of many of the tasks that keep our department running smoothly.

I am grateful to have worked with a great group of fellow graduate students in the Lawrence lab. Dr. Hans Wildschutte introduced me to the wonderful world of protozoan predation; upon joining the lab, Hans immediately took me under his wing and taught me everything I needed to know to get started on my research. Although his research centered more on computational biology, Dr. Adam Retchless was a source of invaluable feedback on experiments, data analysis, and discussion on many scientific ideas. Dr. Heather Hendrickson and Dr. Rajeev Azad, a previous research assistant professor in the lab, provided helpful advice on both my project and on my scientific career. Working with former lab manager Tom Seiflein was a lot of fun, and I would like to thank him for all of his support on many topics throughout my time at the University of Pittsburgh.

During my time in the Lawrence lab, I worked alongside and mentored a small army of undergraduates and technicians who assisted me in the construction of the hundreds of *Salmonella* strains I used throughout the course of my work as well as on a myriad of small projects related to my dissertation research. I would like to thank Bryan Goddard, Jessica Cheek, Ben Cross, Mark Brown, Sarah Hainer, Nikeva Silverton, Brittany Rogers, Jessica Ravenscroft, and David Convissar for all of their hard work and dedication to the protozoan predation project and for all of their help with strain construction. I consider these individuals as fantastic coworkers and as good friends. Although Aletheia Atzinger and Alexis Fitzgerald joined the lab

during my final semester as a graduate student, I appreciated all of their support and advice throughout the process of writing and defending my dissertation.

The 3rd floor of Crawford Hall was an excellent place to work, and I would like to thank the labs of Dr. Roger Hendrix, Dr. Craig Peebles, and Dr. Graham Hatfull for resource sharing. In particular, I would like to thank Brian Firek, Dr. Welkin Pope, Dan Russell, Ching-Chung Ko, Jen Houtz, Andrew Hyrckowian, Charlie Bowman, Kaitlin Mitchell, Lauren Oldfield, Christina Ferreira, Amrita Balachandran, and Dr. Craig Peebles for all of their support and assistance. Also, the Biological Sciences faculty provided an intellectually challenging environment that broadened my horizons on work outside of my specific field.

The Department of Biological Sciences at the University of Pittsburgh has a phenomenal dedication to teaching, and I received excellent training as an educator during my time here. My teaching mentors, Dr. Susan Godfrey and Dr. Tony Bledsoe, were excellent sources of information for teaching, and I thank them for their efforts. I would also like to acknowledge Dr. Valerie Oke, Dr. Melanie Popa, David Hornack, and Debbie Jacobs-Sera for all of their thoughtful assistance in my development as an educator. Additionally, some of my fondest memories of graduate school lie in my work with Dr. Alison Slinskey-Legg, Dr. Lew Jacobson, and Brian DiRienzo as part of the Outreach Program.

As a first generation college graduate, navigating higher education was at times quite challenging. I am fortunate to have received excellent mentorship during my earlier education, in particular at Seton Hill University as an undergraduate and at the Upward Bound Program at California University of Pennsylvania as a high school student. I would like to thank my undergraduate mentors Sr. Ann Infanger, PhD and Dr. Steven Bassett for all of their support and encouragement throughout my undergraduate career; their efforts prepared me for success in

graduate training. I would also like to thank Mrs. Geraldine Jones, Mr. Gary Seelye, Dr. Alton Powe, and all of the staff of the Upward Bound Program at California University of Pennsylvania for their mentorship and guidance; their efforts helped me immeasurably in making a college education a reality.

Finally, I would like to thank my friends and family for all of their support during the long, arduous process known as graduate school. Their support and encouragement certainly kept me sane.

1.0 INTRODUCTION

The Salmonellae are gram-negative, rod-shaped γ-proteobacteria. As members of the Enterobacteriaceae, they are related to the animal pathogens *Escherichia coli, Shigella dysenteria, Klebsiella pneumonia* and *Yersinia pestis* and to the plant pathogens *Erwinia carotovora* and *Dickeya zeae* [388]. Salmonellae are pathogens of many mammalian species, commonly associated with both foodborne illness and enteric fever [131]. In humans, *Salmonella* causes typhoid and paratyphoid fevers and is one of the primary causes of foodborne bacterial illness, resulting in at least 1.5 million cases each year [214]. Salmonellosis typically results in abdominal cramps, diarrhea, nausea, vomiting, and fever [131]. Symptoms occur within 6 to 72 hours after ingestion of an infectious dose, and while most infected individuals recover after 5 to 7 days, those with severe diarrhea may develop to life-threatening dehydration or experience spread of the *Salmonella* to the blood and other body tissues [131]. In rare cases, salmonellosis can lead to Reiter's Syndrome, a condition characterized by chronic inflammation and pain in the joints, eyes, and urethra [78]. The financial costs of food-borne illness in humans in the United States caused by the six most common bacterial pathogens were estimated to be $2.9 to $6.7 billion per year a decade ago [45], a figure that continues to rise [44]. The high prevalence and significant financial impact of *Salmonella* infection led Voetsch [370] to conclude that *Salmonella* "presents a major ongoing burden to public health." Despite being pathogenic to

many hosts, *Salmonella* often adopts a commensal lifestyle in the intestines of some reptiles [39, 51, 109], birds [136, 274] and small mammals [116, 173, 347].

Two species are currently recognized within the genus *Salmonella*: *S. bongori* and *S. enterica*. *S. enterica* is further divided into six subspecies (Table 1); warm-blooded animals are the primary hosts for subspecies *enterica* and *salmae*, whereas cold-blooded animals and the environment are reservoirs for all other subspecies [106, 183-184, 353]. The genus *Salmonella* is further divided into 2579 antigenically-distinct strains, or serovars, according to the Kauffmann-White scheme [106]; 2557 of these serovars belong to the *S. enterica* group and 22 serovars from the *S. bongori* group [106]. Subspecies *enterica* (group I) includes the strains responsible for the majority of cases of mammalian illness [35]; this well-studied group contains 1531 serovars, although this relative over-representation may reflect the collective intense interest in mammalian disease more than the distribution of serovars among natural isolates. Serovars are defined by two major highly diverse surface molecules, the O- and H-antigens [106, 353]. While the term "serovar" refers to a strain of *Salmonella* as defined by the nature of its O- and H-antigens, the term "serotype" refers to the antigenic formula designation assigned to particular O- and H-antigens [106]. For example, *S. enterica* serovar Cholerasuis has an O-serotype of (6,7) and an H-serotype of ([c]:1,5). Classifying *Salmonella* based on the serotypes of the O- and H-antigens is mainly done for epidemiological purposes [125], as the identity of particular serovars are associated with causing disease in particular hosts [264].

Table 1. Classification of the genus *Salmonella*

Information adapted from Tindall *et al* [353] and Grimont and Weill [106].

Species	Subspecies	Serotypes
S. bongori		22
S. enterica	*enterica* (I)	1531
S. enterica	*salmae* (II)	505
S. enterica	*arizonae* (IIIa)	99
S. enterica	*diarizonae* (IIIb)	336
S. enterica	*houtenae* (IV)	73
S. enterica	*indica* (VI)	13

The O-antigen is the outer-most portion of *Salmonella*'s lipopolysaccharide (LPS) layer (Figure 1); it is a constitutively-produced repeating polysaccharide unit and is the most abundant molecule on the cell surface [293]. The variable component of the *Salmonella* O-antigen polysaccharide is produced by the *Salmonella rfb* genes, which encode various sugar synthases and transferases that assemble the repeating polysaccharide units of the O-antigen [293]. These units are assembled into long chains by the Rfc O-antigen polymerase [293]. At the inner surface of the inner membrane, the O-antigen polysaccharide is then linked by the WaaL O-antigen ligase to lipid A, a core oligosaccharide that is highly conserved across enteric bacteria [123, 293, 305]. The Wzx translocase then "flips" the entire lipopolysaccharide antigen onto the outside of the cell [205]. While the O-antigen may be modified, such as in the case of the acetylation of the *Salmonella* LT2 O-antigen by the unlinked *oafA* gene [328], variability in saccharide composition and linkage is conferred by variable gene content at the *rfb* locus.

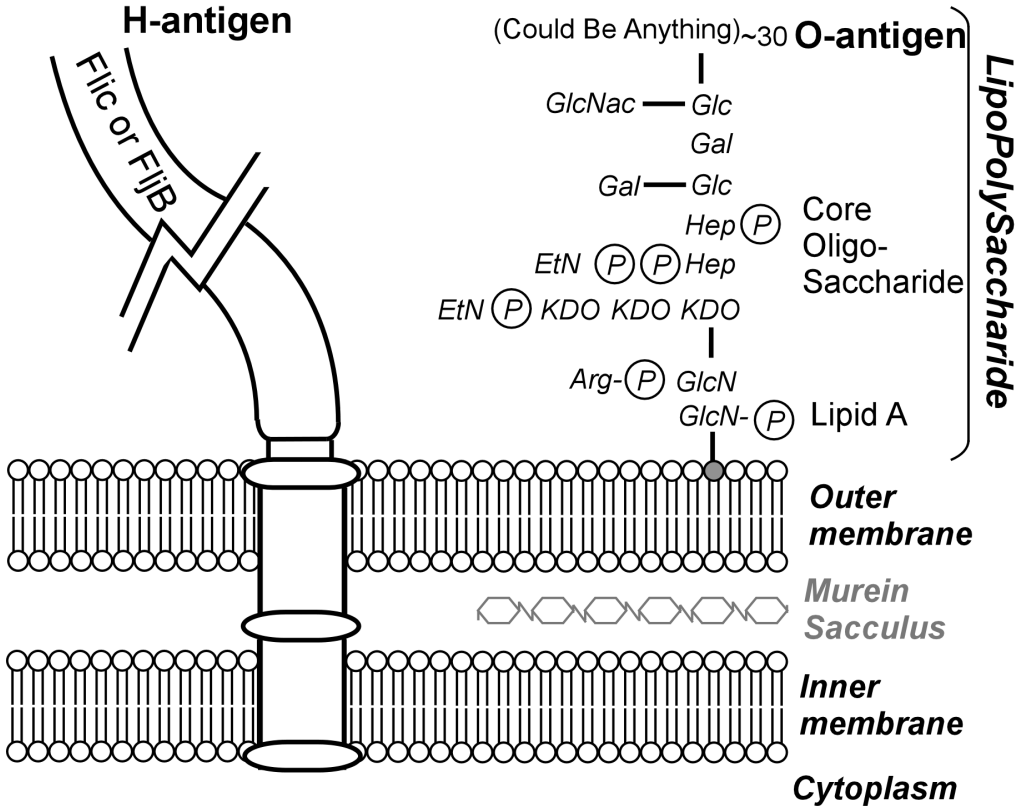

Figure 1. Schematic diagram of the O- and H-antigens of *Salmonella*

The H-antigen is conferred by the filament of *Salmonella*'s peritrichious flagella, which are used for swimming. Unlike the O-antigen, the H-antigen is only expressed under certain environmental conditions [55]. In most serovars, phase switching occurs between one of two phases of the H-antigen, which are encoded by separate genes [55, 186-187]. Serovars also differ at other antigenic proteins not included in the classification scheme, including the major outer membrane porins OmpC [326], OmpD [299, 325], and PhoE [327, 336]. Porins are transmembrane proteins that permit the diffusion of large molecules through the cell's outer membrane; they are typically highly conserved at transmembrane regions and but highly diverse at exposed, external loops [242-243].

This chapter will focus on diversity at the *Salmonella* O-antigen. Below, I discuss how genetic variability at the O-antigen-encoding *rfb* operon cannot be explained by conventional models and develop a framework for the maintenance of antigenic variability in *Salmonella* populations. It is the goal of this thesis to establish a framework that explains the generation and maintenance of high levels of antigenic diversity within *Salmonella*.

1.1 FREQUENCY DEPENDENT SELECTION AND ANTIGENIC DIVERSITY

An interesting property of loci encoding antigenic determinants is that many are hypervariable; that is, given the expectations of diversity afforded by neutral variation [165], chromosomal loci encoding antigens (or loci linked to those that do) are far more variable than one would expect. This is true for the *Salmonella*'s H-antigen-encoding *fliC* gene [213, 330-331, 334], and the O-antigen-encoding *rfb* operon [40-41, 153, 188-189, 198-200, 366-367, 374, 391-392]. For loci in many bacteria, antigenic diversity can be explained by frequency dependent

selection (Figure 2A; [192]). Here, the fitness contribution of any given antigen-encoding allele depends not only on the function of its product but on its frequency in the entire population. There are two general models one can consider where rarity is advantageous.

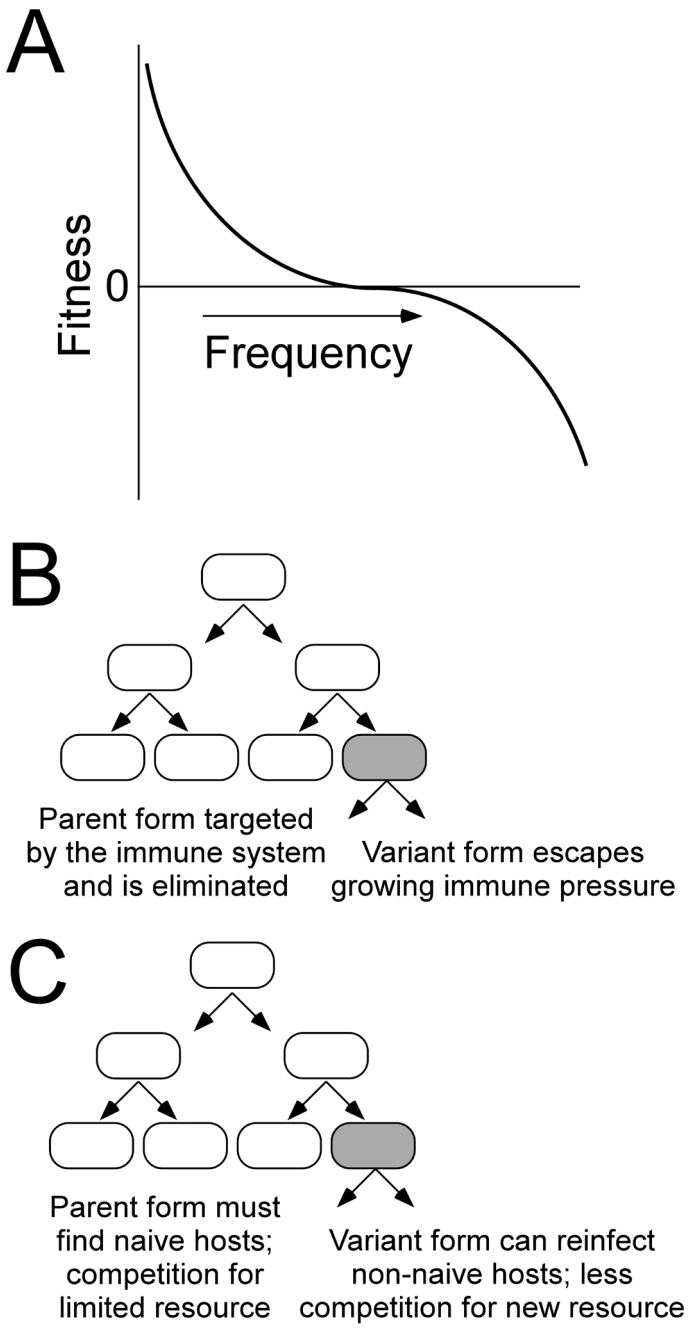

Figure 2. Frequency dependent selection

A. An overview of frequency dependent selection. **B**. Spatial heterogeneity model for frequency dependent selection; here, variant daughter cells gain an advantage by being able to infect hosts that have been exposed to their parents' antigens. **C**. Temporal heterogeneity model for frequency dependent selection; here, variant daughter cells persist in a single host that has begun to respond to their parent cells' antigens.

First, the fitness of a particular antigen varies over space, or the likelihood of encountering a host without previous exposure to that antigen. Thus, the spatial fitness of a given antigen is dependent on the distribution of hosts having susceptibility to infection by cells expressing this antigen and the probability of encountering these hosts (Figure 2B). When an antigen is common in a microbial population, hosts are more likely to encounter and mount an immune response against this antigen. Thus, cells having common antigens are far more likely to encounter hosts having prior exposure to these common antigens; these hosts are considered non-naïve to these common antigens and consequently the spatial fitness of common antigens is relatively low. For example, cells having common antigens are more likely to encounter hosts having previous exposure to these frequently occurring antigens; these hosts are considered non-naïve for the antigen of interest. Cells having rare antigens are more likely to encounter and successfully infect hosts without previous exposure to these rare antigens. As these cells preferentially multiply, their antigenic profile becomes more common in the bacterial population, leading to a lower population of naïve, susceptible hosts (Figure 2B). Therefore, frequency-dependent selection can favor pathogens that possess mechanisms to maintain population-level antigenic diversity over those that are unable to maintain such antigenic diversity.

Alternatively, rare antigens can provide an advantage in fitness over time rather than over space on a temporal scale. Once it has successfully infected a host, a pathogen can continue to evade the host's immune system to prolong the infection and increase reproductive capability only if it alters its antigenic profile. Pathogens that are able to switch their antigenic profile every few generations have a distinct advantage in prolonging infection, engaging in an "arms race" with the immune system of the host organism (Figure 2C). Continual presentation of the same antigenic profile would allow the host adaptive immune system to mount defenses against the

invading organism, whereas organisms that have the capacity to switch antigenic profiles have much better chances at evading the host immune system, reproducing, and maintaining the infection.

Because environments are dynamic, no one antigenic profile will have a sustained high fitness level over both space and time [192]. Thus, traditional host-pathogen dynamics generally favor selection of organisms that generate heritable, random, and reversible antigenic variation. As discussed below, many pathogens possess molecular mechanisms which produce generation-time-scale diversity, whereby daughter cells are often antigenically distinct from their parents. As a result, the population as a whole will be diverse at antigen-encoding loci as a result of constant change at short time scales.

1.2 CONVENTIONAL WISDOM FAILS TO EXPLAIN ANTIGENIC DIVERSITY IN *SALMONELLA*

Neither the spatial nor the temporal model of selection discussed in Chapter 1.1 explaining host-pathogen interaction appear to apply to highly variable strains of *Salmonella enterica*. Here, organisms are not infected by recently unencountered serotypes, and *Salmonella* does not alter its antigenic profile during the course of infection. Rather, each host of *Salmonella* is infected by only a small subset of *Salmonella* serovars [264], the composition of which is specific to each host (Figure 3). For example, swine are commonly infected with serovar Cholerasuis, cattle with serovar Dublin, poultry with serovar Gallinarium, sheep by serovar Abortus-ovis, horses by serovar Abortus-equi, and so on (Figure 3). Beyond the sample of one outbreak, this pattern is consistent across both time (several decades) and space (several countries). This poorly-

understood pattern of infection is referred to as host-serovar specificity. Both of the antigens used to classify serovars have been implicated in *Salmonella* virulence [24, 65, 107, 162, 233, 249, 304, 350], and efforts to link *Salmonella* antigenic diversity to host-serovar specificity have mainly focused on the relationship between *Salmonella* and the host immune system [32, 216, 358]. Whereas diversity at the H-antigen-encoding loci and other minor antigenic loci has been linked to immune system evasion [5, 18, 65, 77, 137-138, 143, 162, 244, 249, 304, 311, 378], the exact role of the O-antigen in determining host-serovar specificity remains unclear. Regardless of mechanism, any persistent relationship between O-antigen serotype and infected host belies explanation by frequency-dependent selection.

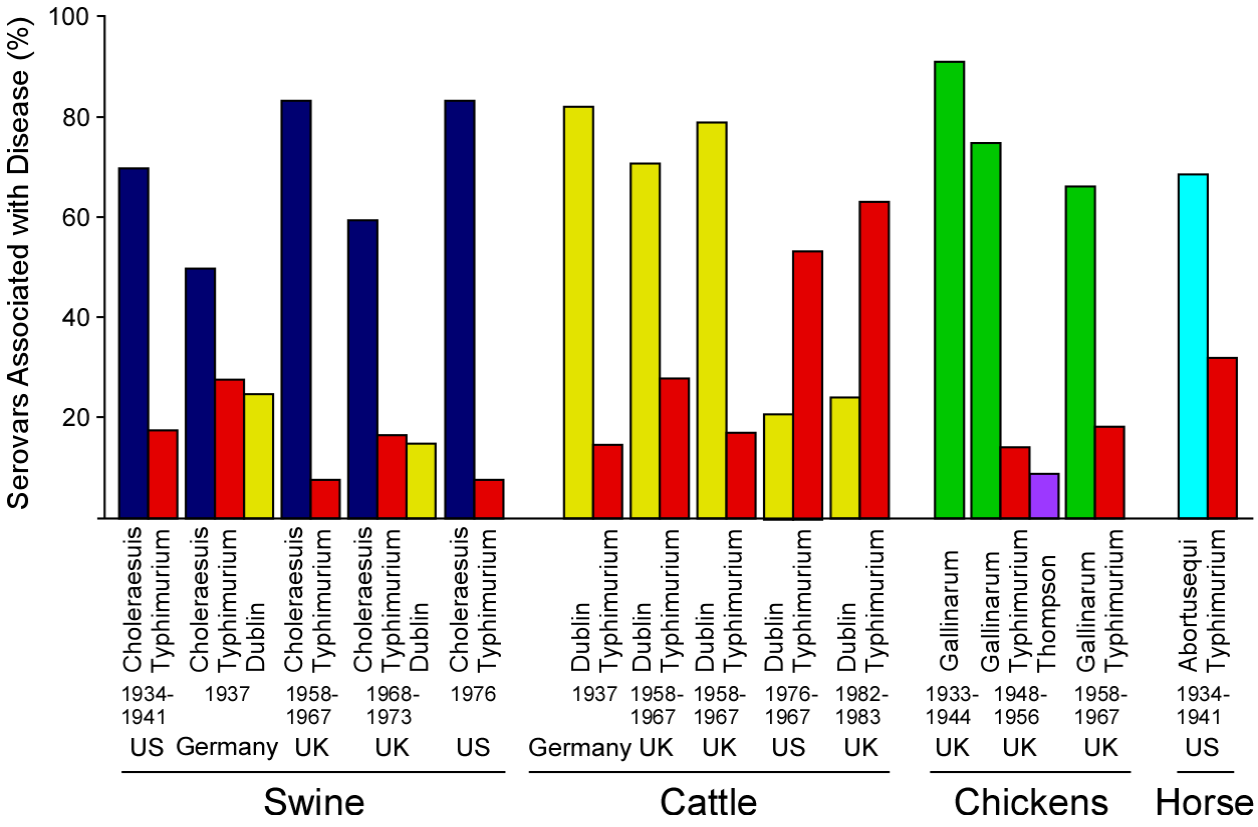

Figure 3. Host-serovar specificity in *Salmonella*

The proportion of different serovars is shown for outbreaks in different mammalian hosts. Data adapted from Rabsch *et al* [264].

Variability at any antigenic locus reflects the sum of selective pressures acting on it. As I discuss below, the nature of the variability at the *Salmonella rfb* locus reflects a suite of selective pressures experienced by this pathogen that differs from those influencing antigenic loci in other species. I propose an alternative mechanism explaining the maintenance of *Salmonella* O-antigen diversity that focuses on *Salmonella*'s lifestyle in the broader intestinal ecosystem. Here, the benefits of generation-time-scale constancy – that is, avoiding variability – outweigh any benefits of generation-time-scale diversity. As a result, frequency-dependent selection cannot be invoked to explain antigenic diversity in *Salmonella*. Rather, diversity at the population level must be maintained by diversifying selection acting on the different antigenic types. To outline this model, I begin by contrasting antigenic diversity in a number of pathogens to highlight both the differences between the selective regimes faced by *Salmonella* and other pathogens and the differences in the mechanisms by which diversity is generated and maintained.

1.3 GENERATION TIME-SCALE ANTIGENIC DIVERSIFICATION

For many pathogens, population-level diversity reflects the production of variant antigenic types on a generational time scale. That is, diversity at the species level results from the collection of highly variable cells that are produced at the cellular level. Critically, these organisms possess clear mechanisms that produce variant daughter cells. Three of these organisms, each possessing different mechanisms that reflect three different selective regimes, are discussed below. The lack of selective sweeps purging the variability reflects the difference in the time scales over which the two processes act.

1.3.1 *Haemophilus*: extending persistence times in hosts

Pathogens may continually change surface antigens by stochastic activation or inactivation of constituent genes. This process is mediated by DNA polymerase slippage on repeated nucleotide tracts termed contingency loci [224-225]. This frequent slipped-strand mispairing may occur during DNA replication and repair, resulting in the insertion or deletion of a single nucleotide or repeated nucleotide group [225, 389]. Thus, antigenically diverse daughter cells are continually produced. If located within the coding sequence of a gene, these contingency loci can cause changes in gene expression at the translational level by changing the gene's reading frame and the location of the stop codon [193]. If located within the promoter of a gene, they may alter the RNA polymerase binding sites or facilitate premature transcription termination [193, 389]. Contingency loci permit heritable, reversible, and random genotypic changes to antigen-encoding genes, changing the expression of these genes nearly every generation [19].

Notable contingency loci in the human commensal *Haemophilus influenzae* include the *lic1*, *lic2*, and *lic3* genes responsible for the major surface antigen lipopolysaccharide (LPS) biosynthesis [127-128, 133-134, 281-282, 379-380]. Expression of different O-antigen serotypes in *Haemophilus* is regulated by polymerase slippage at the 5' region of these genes, which contain a variable number of CAAT repeats [379-380]. Polymerase slippage at these sequence repeats alters the translation of *lic* genes by placing the genes in or out of the proper reading frame [379]. O-serotype variation in *Haemophilus* is critical for avoiding attack by the host adaptive immune system and maintenance of colonization within the restricted niche of the upper respiratory tract in individual hosts [381]. *Haemophilus* can also switch from commensal to opportunistic pathogen, causing meningitis and septicemia in infected hosts [218, 223, 403]. The

lifestyle switch from commensal to pathogen is dependent on O-antigen variability at both intrastrain [355] and interstrain [145-146] levels.

The ability of *Haemophilus* to continually vary its antigenic profile through polymerase slippage gives it the ability to avoid recognition by the host immune system, prolonging colonization in one particular host. Frequency-dependent selection arising from selective pressure from the host immune system explains the maintenance of antigenic diversity in the *Haemophilus* population, as no one surface antigen confers a distinct temporal advantage over the others. When surface antigens are recognized, immune defenses are produced by the host organism that can eliminate *Haemophilus* from that host. *Haemophilus* has evolved contingency loci to vary surface antigens randomly at a high frequency, which decreases the probability of an effective immune system defense and prolongs the *Haemophilus* within-host life cycle (Figure 2C). Thus the mechanism generating phenotypic variability provides insight into the selective forces acting on this species.

1.3.2 *Neisseria*: Infecting non-naïve hosts

Like *Haemophilus*, *Neisseria meningitidis* is a commensal of the human upper respiratory tract [48]and has the capacity to adopt a pathogenic lifestyle, causing meningitis and septicemia [47]. Contingency loci have been associated with the ability of *Neisseria* to maintain its commensal lifestyle, regulating many genes involved in the biosynthesis of antigens [333, 345] such as lipopolysaccharide [176], pili [354], opacity proteins [338], capsular polysaccharides [113, 181], and the PorA outer membrane protein [362]. Expression of the PorA protein, which is used to identify the serosubtype of *Neisseria* strains, is regulated by variation in the number of nucleotide repeats both in the promoter region and within the coding sequence of the gene.

Insertions or deletions caused by slipped-strand mispairing regulates PorA expression at the transcriptional level in the homopolymeric G-repeat region located in the promoter [361] and at the translational level in the homopolymeric A-repeat region located within the coding sequence [362]. In addition, expression of the *lgtABE* genes responsible for lipopolysaccharide biosynthesis is regulated by slipped-strand mispairing at a homopolymeric tract of G-repeats upstream of *lgtA*, the first gene of the locus [151-152].

Antigenic diversity in the *Neisseria meningitis* type IV pili, a major target of the host immune system [47], is generated by gene conversion. The antigenic portion of the pilus is PilE, which consists of a conserved N-terminus and a variable C-terminus [261]. Variation at the PilE C-terminus arises from RecA-dependent non-reciprocal recombination between the expressed *pilE* gene and several silent *pilS* loci found up to hundreds of bases away from the *pilE* gene [255-256, 261]. Donation from a *pilS* gene to *pilE* is based on short sequence homology and occurs through several RecA-mediated crossover events between genes [172, 310]. Type IV pili are involved in attachment and colonization of *N. meningitis* to mucosal membranes [211], and variation in pili have been linked to changes in antibiotic resistance [204] and adhesion to host surfaces [104, 289, 337]. Gene conversion at the hypervariable region of *pilE* can be explained by frequency-dependent selection, as generation of novel antigenic variants is key to the ability of *Neisseria* to colonize non-naïve hosts [7].

High rates of antigenic phase switching in *Neisseria*, as in the case with PorA and lipopolysaccharide, most likely evolved as a response to the selection pressure to establish a commensal relationship with non-naïve hosts [217, 306]. *Neisseria* is typically cleared from the host in a few days to several months after initial colonization [49], during which time the host immune system builds up a defensive response to *Neisseria* based on the recognition of bacterial

surface antigens. Phase variation of surface antigens allows *Neisseria* to re-establish a commensal relationship with non-naïve hosts that have built up adaptive immune responses from previous colonizations (Figure 2B). Frequency-dependent selection at the spatial level explains the high degree of antigenic variation maintained in the *Neisseria* population, as strains with more common antigenic profiles will encounter a greater number of non-naïve hosts. Antigenic phase variation allows *Neisseria* to continually evade host immune systems to colonize new hosts, regardless of the prior colonization status of the host. This is especially important given that *Neisseria* colonizes approximately 10% of the population at any given time in industrialized countries [91].

Interestingly, *Neisseria* virulence could be viewed as a rare consequence of phase variation, in which a commensal switches its antigenic profile to a pathogenic form, allowing tissue invasion and migration of bacteria into the host bloodstream [49, 217]. Pathogenic *Neisseria* are rarely transmitted between hosts; rather, pathogenicity arises from within a commensal population. Therefore, the selection pressure to establish commensal colonization of non-naïve hosts likely drives selection for antigenic phase variation, rather than pressure on pathogenic forms to infect non-naïve hosts. As with *Haemophilus*, the mechanism for creating phenotypic diversity sheds light on selective forces acting on this species.

1.3.3 *Bacteroides*: Avoiding innate immune response

Bacteroides fragilis, a major gram-negative bacterial inhabitant of the human intestine [194, 393], synthesizes a large number of phase-variable surface antigens using site-specific inversion [50, 180]. In this mechanism, short, inverted DNA repeats flank the invertible element, which typically contains a promoter for adjacent antigen-encoding genes [360]. These repeated

sequences are recognized and brought together in a synapse by a DNA invertase, which cleaves DNA through strand exchange, resulting in reciprocal recombination and inversion of the DNA segment flanked by the repeated sequences [360, 389]. DNA invertases belong to one of two classes based on the mechanism by which they cleave and ligate DNA: the serine site-specific recombinases (Ssr) or the tyrosine (or lambda) site-specific recombinases (Tsr) [100, 329]. Inversion results in changes in orientation of promoters for various genes, which in turn affects gene expression. Like contingency loci, invertible DNA regions produce random, heritable, and reversible changes in antigenic genotypes.

B. fragilis is able to produce eight distinct capsular polysaccharides determined by the expression of the PSA-H loci [50, 62, 180]. Expression of each capsular polysaccharide locus, with the exception of PSC, is regulated by specific inversions of DNA termed *fin* regions [251]. Promoters for the capsular polysaccharide-encoding loci (except for PSC) are located in the *fin* regions immediately upstream of each locus, with transcription of each locus dependent on the orientation of its promoter [50, 63, 180, 251]. Inversion of *fin* sites mediated by the Ssr Mpi recombinase can switch genes on or off at random [63], and the transcriptional status of each promoter is independent of the expression of other capsule-encoding loci. The only exception is the PSC locus, which produces a default capsular polysaccharide that is thought to act as a "fail-safe" in the event all seven other loci are turned off [63, 174]. Each *B. fragilis* cell has the capacity to express any suite of capsular polysaccharides simply based on the inversion status of *fin* regions, resulting in local host-level population antigenic diversity.

Production of multiple surface polysaccharides has been demonstrated for the successful long-term colonization of the intestine by *B. fragilis* [63, 197]. Because the human intestinal ecosystem is a very dynamic and competitive environment, many factors could be responsible

for maintaining diverse surface antigens in *B. fragilis*. Avoidance of bacteriophages, adhesion to changing intestinal surfaces, or competition with other commensals or pathogenic bacteria could also be involved in the maintenance of mechanisms that permit *Bacteroides* to vary its surface antigenic profile. *B. fragilis* expression of PSA has also been shown to actively protect against intestinal colitis caused by *Helicobacter hepaticus* in an animal model [208], further highlighting the complex role antigenic phase variation plays in the lifestyle of *B. fragilis*. Underlying these complex interactions is the close association *B. fragilis* forms with the intestinal mucosa. Because *Bacteroides* species form a majority of the cells in the intestinal microbiota [11, 97, 300], it is a likely target for sampling by dendritic cells which would result in IgA excretion targeting over-represented O-antigen epitopes [201-203]. Therefore, continual variation of surface antigens likely protects *B. fragilis* from attack by the host immune system [180]. Within-host frequency-dependent selection may favor a diverse array of capsular polysaccharide production by *B. fragilis*, as the continual presence of a predominant polysaccharide antigen may result in the host immune system mounting defenses against *B. fragilis* and clearing it from the intestine.

Many site-specific invertible regions found in bacteria, especially those of the Ssr family, have been imported from bacteriophages [329]. Bacteriophage P1 encodes the site-specific recombinase Cin, which mediates the recombination event that results in the phase-variable expression of tail fiber genes, altering the host range of P1 depending on which tail fibers are expressed [126, 141-142]. Host range of the temperate coliphage Mu is also dependent on a site-specific recombinase, encoded by the *gin* gene [160, 359]. *Escherichia coli* contains a recombinase that is very similar to the *gin*- and *cin*-encoded proteins, the Pin recombinase, which has been shown to rescue *gin* mutants of Mu, that is involved in flagellar phase variation [85,

259]. DNA invertases can control a wide variety of phenotypes in bacteria and bacteriophages that are under frequency-dependent selection, including host range and antigenic variation. Thus this mechanism of phase variability, used by *Bacteroides* to provide generation-time-scale diversity, has the potential to be widely distributed among bacteria.

1.4 MECHANISMS FOR GENERATING DIVERSITY REFLECT AN ORGANISM'S SELECTIVE REGIME

In all of the cases discussed above, it is beneficial for cells of one antigenic type to yield offspring of a different antigenic type. The continual switching of surface antigens presents a host-pathogen "arms race," in which antigen switching occurs in response to selective pressure from the host adaptive and innate immune systems. *Neisseria, Haemophilus* and *Bacteroides* contact the adaptive and/or innate components of the host immune system, so it is advantageous for these organisms to maintain molecular mechanisms that permit continual switching of surface antigen profiles to evade immune defenses within and among hosts. The maintenance of molecular mechanisms that permit frequent antigenic phase switching can allow microorganisms to prolong infection or colonization within an individual host and increase the likelihood of infecting non-naïve hosts. Frequency-dependent selection explains the maintenance of antigenic diversity for such organisms, as host immune responses prevent any one antigenic profile from dominating a population of infectious or commensal microorganisms for more than a brief period of time.

When a particular surface antigen becomes common in a population, the chances that host immune systems mount defenses against that antigen increase. Once an immune response is

mounted against a particular surface antigen, cells expressing that surface antigen are more likely to be eliminated by the host immune system, whereas cells that have switched antigens can continue to evade host defenses, prolonging infection within a host or spreading to non-naïve hosts. Under conditions that favor frequency-dependent selection mediated by host immune systems, populations of microorganisms that possess the capacity to generate offspring with different surface antigens than parent cells have a greater survival advantage over microorganisms that are unable to vary cell surface antigens.

Perhaps nowhere is this phenomenon better illustrated than with human immunodeficiency virus, where excess phenotypic diversity is one of the main obstacles to creating successful treatment methods and potential vaccines for HIV [267]. Here, the mechanism generating diversity is intrinsic to viral reproduction, but the concept is the same. HIV and other retroviruses are especially prone to mutation, as the molecular mechanisms used for retroviral reproduction (lack of proof-reading and short replication time) are in themselves are highly mutagenic [94, 268]. In addition to its high mutation rate, HIV undergoes approximately three recombination events per genome per replication, one of the highest recombination rates of all organisms [401]. In HIV, recombination typically occurs between two co-infecting virus particles through the actions of the viral reverse transcriptase, which can switch between the strands of the co-packaged viruses [268] and can occur between viruses of the same subtype, different subtype, or different groups [88, 148, 159, 210]. High rates of mutation and recombination create a virus population with high, continually-changing antigenic diversity which allows HIV to rapidly adapt surface antigens in response to selective pressure from host immune systems. HIV is able to prolong within-host infection through rapid changes of surface antigens and other properties, which enables it to evade host immune system defenses.

1.5 ANTIGENIC DIVERSITY IN *SALMONELLA*

As outlined above, *Salmonella* serovars are classified by their two highly variable antigens, the O-antigen polysaccharide and the H-antigen flagellar filament. As I discuss here, the mechanisms producing this diversity vary greatly from the mechanisms discussed above. Therefore, the selective regime responsible for maintenance of that diversity must be different than selective pressure from the host immune system driving antigenic diversity in *Neisseria, Haemophilus,* and *Bacteroides*. I will use this contrast to argue that factors specific to each host environment must drive the maintenance of antigenic diversity within the Salmonellae.

1.5.1 *Salmonella* H-antigen diversity

The *Salmonella* H-antigen is conferred by flagellin, the major filament of its peritrichious flagellae. Unlike many constitutively-expressed antigens, flagellae are expressed only under certain environmental conditions. Expression of flagellin is regulated by the flagellar master regulator genes *flhCD* in response to starvation conditions when locomotion is advantageous [178, 395]. The FlhCD master regulatory proteins up-regulate numerous genes for the synthesis of the flagellum, motor proteins for flagellar rotation and the chemotaxis signal-transduction system which controls the direction of rotation. The flagellar filament is the final, outermost portion of the bacterial flagellum to be synthesized and assembled; because they form the exposed portion of the flagellum, *Salmonella* flagellins are targets of both the innate and adaptive host immune systems [240, 291, 294-296]. Flagellin binds Toll-like receptor 5 (TRL5), activating a pro-inflammatory response by the innate immune system [6, 89]. Flagellin is also

recognized by memory $CD4^+$ T cells, which are involved in the clearance of *Salmonella* from infected phagocytes [24-25, 65].

The most common form of flagellin found in *Salmonella* is encoded by the *fliC* gene, which is embedded within the major flagellar gene locus. The *fliC* alleles found in different serovars of *Salmonella* are quite diverse (Figure 4A); in *E. coli* (where they have been studied more intensively), it has been proposed that strong frequency-dependent selection leads to this diversity [375]. It is important to note here that the work described in this dissertation takes the perspective that the biology of *Salmonella* is at odds with the frequency-dependent selection model traditionally used to explain the generation and maintenance of antigenic diversity within *Salmonella*. Instead, I argue that diversifying selection mediated by intestinal protozoan predators is a more valid hypothesis explaining antigenic diversity, at least with respect to the O-antigen. I discuss this hypothesis in further detail in Chapter 1.6. In addition, most *Salmonella* serovars have the capacity to produce one of two possible forms of antigenic flagellin at any one time, with the alternate (H2 antigen) flagellin encoding genes found in the unlinked *fljBA* operon [2-3, 158, 394]. Like the *fliC* gene, alleles of the *fljB* flagellin gene are also hypervariable [213]. In general, different *fljB* and *fliC* alleles are well-conserved at the 5' and 3' ends with hypervariable regions in the middle of the genes (Figure 4A and [213, 331, 375]). These regions correspond to the functionally constrained N- and C-termini and the antigenically-exposed middle domain of flagellin [213], respectively. Diversity of flagellin-encoding genes account for 114 different serotype combinations made from 99 antigenically distinct H-antigen factors [106]. In a minority of serovars, plasmid-encoded elements create production of a third flagellar phase or influence the H1 or H2 serotypes [14, 332].

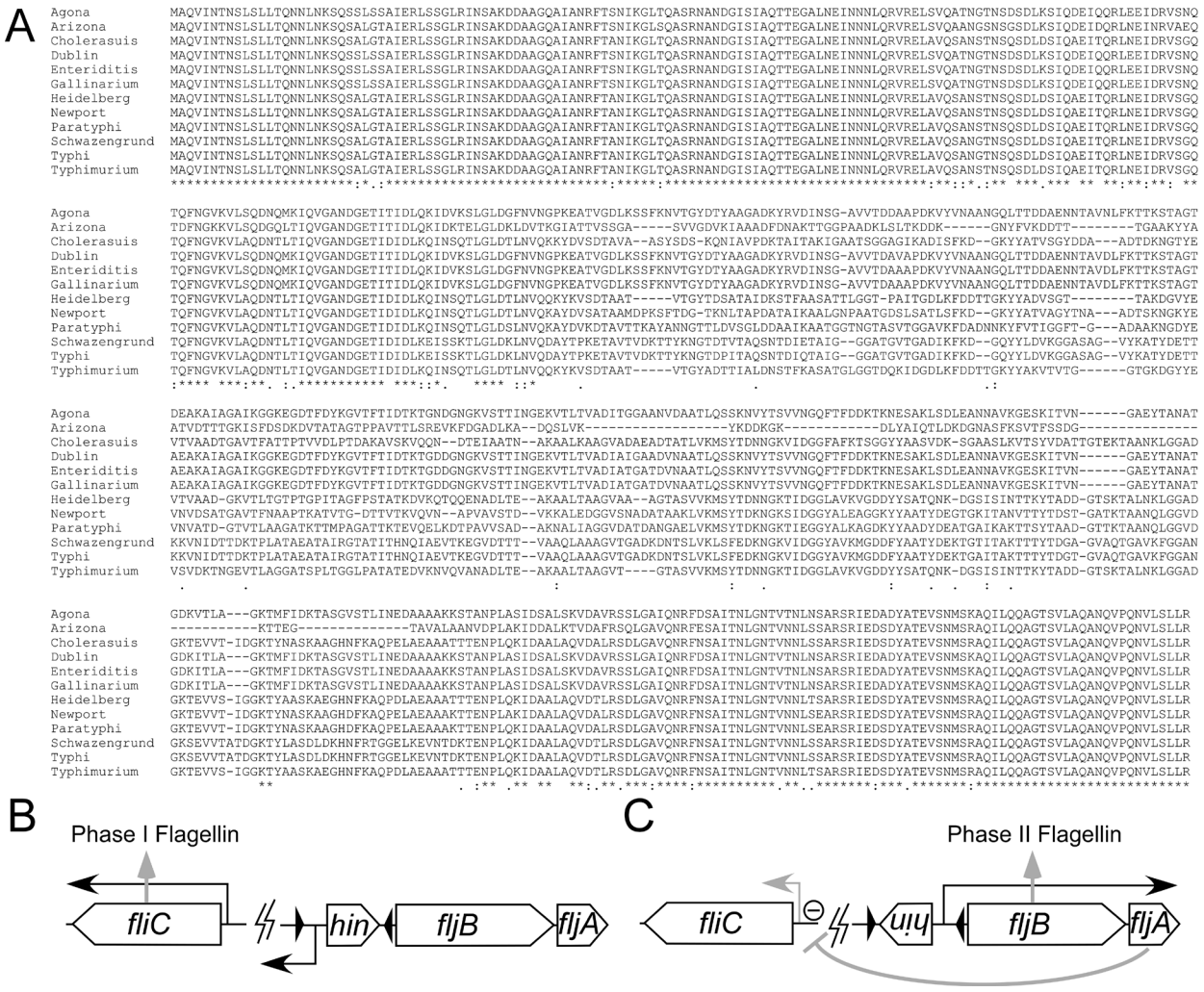

Figure 4. Diversity and phase variation of the *Salmonella* H-antigen

A. Diversity at the *Salmonella fliC* locus. **B**. Mechanism of phase switching at the *Salmonella fljBA* locus. Phase I flagellin is expressed from the *fliC* gene. **C**. When the *hin* region inverts between the two *hix* sites, Phase II flagellin is expressed from the *fljB* gene, while the FljA repressor prevents *fliC* expression.

Phase variable expression of the two flagellins is controlled by site-specific inversion of *hin*, a region of DNA most likely arising from integration of a Mu-like phage [360]; the *hin* region bears structural similarity to the *fin* regions of *B. fragilis* and the *cin* region of bacteriophage P1 [251]. Aside from the gene for the *cis*-acting invertase, the *hin* region contains the promoter for the *fljBA* operon (Figure 4BC). Under appropriate environmental conditions, expression of flagellae is turned on by the actions of the flagellar transcriptional regulatory factors FlhCD [178]. To begin, the Type III secretory system that comprises the transmembrane core of the flagellum is synthesized [55]. After this is complete, a flagellin gene is expressed, and export of the flagellin results in filament synthesis outside the cell [55]. When the *hin* gene is in one orientation (Figure 4B), the *fljBA* promoter is in the incorrect orientation for *fljBA* expression; the *fljBA* operon remains silent and the *fliC* gene is transcribed, thus producing the H1 antigen [2-3]. Stochastic expression of the *hin* gene results in site-specific recombination at the two *hix* sites which flank the *hin* gene [2, 402]. In the opposite orientation, the *fljBA* operon acquires an indirectly FlhCD-responsive promoter. FljA represses both the transcription and translation of the flagellin-encoding *fliC* gene [3, 394], while FljB encodes an alternate flagellin, presenting the H2 antigen (Figure 4C). Hence, switching between the H1 and H2 antigens is reversible, heritable, and occurs at random on a generational time scale.

While phase switching allows two H-antigens to be variably expressed, the variable H-antigen-encoding genes found in the *Salmonella* population are not variably expressed on generational timescales. That is, variable alleles are found in the population, but *Salmonella* cells produce daughter cells with the same two H-antigen flagellins as their parents. Moreover, the conservation of the central, flagellin-variable domain [213, 331, 375] suggests that mutations that alter the sequence of the flagellin are counter-selected, rather than placed under positive

selection. That is, if frequency-dependent selection acted on these proteins, one would expect an excess of nonsynonymous substitution in the variable domain of the flagellin gene; however, this is not observed. Therefore, *Salmonella* H-antigenic diversity has a generational-time-scale component in its phase switching, but population-level selection must act to maintain excess diversity at the constituent *fliC* and *fljB* loci.

1.5.2 *Salmonella* fimbrial diversity

Salmonella has several different types of fimbriae, which are involved in attachment of *Salmonella* to intestinal epithelia [5, 138]. Some of these adhesion factors undergo phase variation, although the molecular mechanism by which this is accomplished is poorly characterized compared to that of the H-antigen. The long polar fimbriae-encoding *lpf* operon undergoes generational time-scale, heritable phase variation and expression is required for *Salmonella* colonization of Peyer's patches [137]. LpfA, the major subunit of long polar fimbriae, has been shown to elicit an antigenic response in mice [137]. Although the role of fimbrial diversity present in *Salmonella* is unclear, the capacity of *Salmonella* to phase-regulate expression of the *lpf* operon demonstrates that *Salmonella* is capable of using multiple molecular mechanisms of phase variation to regulate surface antigenic diversity.

1.5.3 *Salmonella* O-antigen diversity

The O-antigen, the outermost layer of the gram-negative lipopolysaccharide, is a repeating sugar unit found on the outside of the cell and is the most abundant cell surface molecule in *Salmonella* [293, 305]. The O-antigen is synthesized by various sugar synthases and transferases encoded by

the *rfb* genes, a cluster of genes typically 10-35 kb in length located between the *gnd* and *galF* genes on the *Salmonella* physical map [37, 209, 272]. While flanking genes are little changed among *Salmonella* serovars, the *rfb* operon varies widely in gene composition among *Salmonella* serovars ([270, 272-273]; Figure 5A). For example, serovars Typhi, Typhimurium and Dublin share a common set of genes responsible for the 4, 12 and 9, 12 serotypes (the 1 epitope is conferred by a prophage and the [5] epitope is conferred by an unlinked gene). Yet the *rfb* operon conferring the 3,10 serotype on serovar Welteverden has gained and lost numerous genes relative to serovar Typhimurium. Serovar Cholerasuis is even more extreme, having no genes in common with serovar Typhimurium in the *rfb* operon (Figure 5A). This variability in gene content results in varying patterns of content, linkage, and order of sugars composing the antigen [293]. Variation at the *rfb* locus arose over time by a series of horizontal gene transfer events [41, 67, 153, 161, 195, 200, 270, 272, 366, 373-374, 376, 391-392], where the introduced genes encode enzymes for the synthesis and assembly of novel sugar configurations or compositions. With respect to the O-antigen, the only time a cell can become antigenically distinct from its parent is through the acquisition and maintenance of a new *rfb*-like gene from another bacterium.

Mutating the *rfb* locus affects the pathogenicity of *Salmonella*, but it is not clear why. Although LPS does provide a mechanism for adhesion to eukaryotic cells surfaces, SPI-1 encoded genes are responsible for pathogenicity-specific cell adhesion. *Salmonella* also contains numerous fimbrial genes involved in many forms of adhesion [138]. Therefore, *rfb*-encoded genes likely influence the efficacy of infection via a more indirect route. Similar patterns of variability in *rfb* operon composition are seen among serovars of the closely-related species *Escherichia coli* (Figure 5B). In both cases, recombination has introduced foreign genes into the operon, enabling different sugars to be synthesized and novel biochemical linkages to be created

among these various sugars. This results in differences among O-antigenic types that reflect stable gene loss and gain (Figure 5C), not mutational change, slipped-strand mispairing, gene conversion events, or site-specific recombination in invertible segments of DNA.

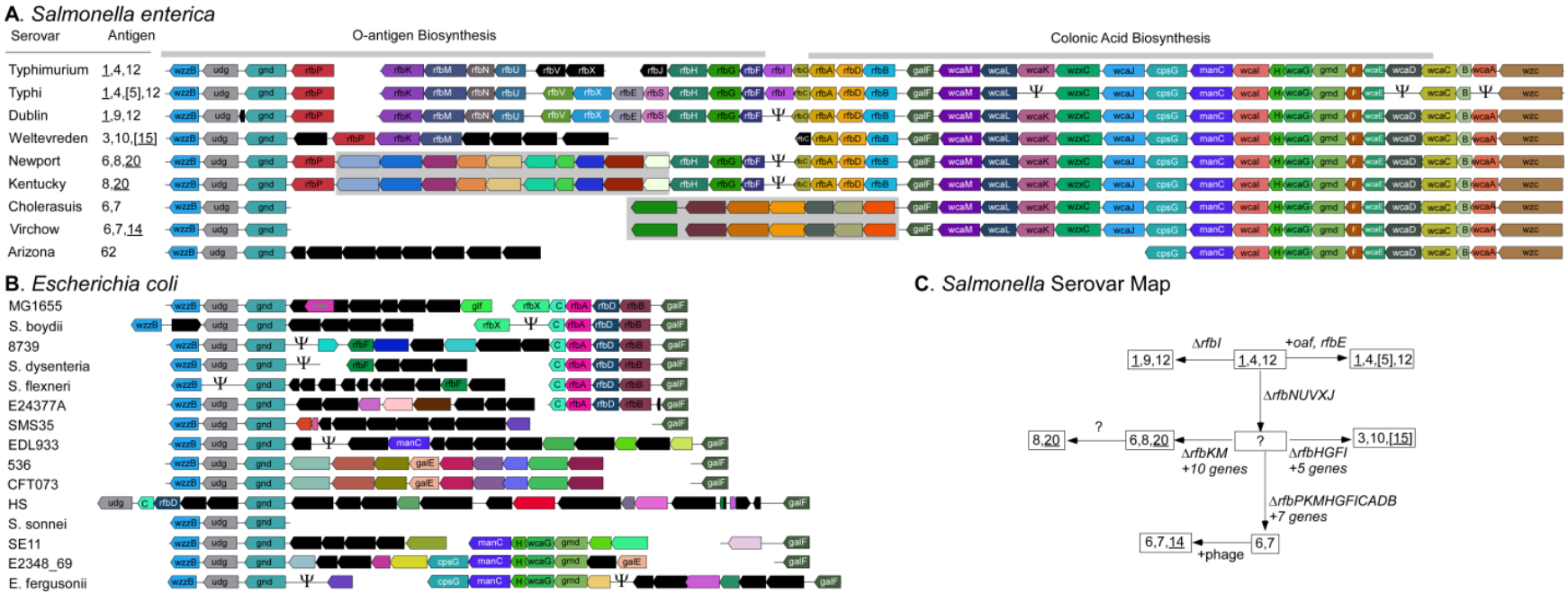

Figure 5. Diversity at the *Salmonella rfb* locus

A. Alignment of *rfb* operon regions of *Salmonella* strains. Orthologous genes are shaded with the same color. **B**. Alignment of *rfb* operon regions of *E. coli* strains. Genes with *Salmonella* orthologues are shown in cognate colors. **C**. Differences between *Salmonella* O-antigens.

Although the sugar synthases and transferases encoded by the *rfb* genes are responsible for the majority of the hypervariable region of the O-antigen, other genes have been shown to affect the serotype of the O-antigen. In serovar *Typhimurium* LT2, O-factor [5] is an acetylation of the O-antigen mediated by *oafA* [328], and O-factor 27 is due to $wzy_{a1\text{-}6}$ [373], an O-antigen chain-length regulator closely linked to *rfb*. The O-antigen is a crucial virulence determinant and most likely also protects *Salmonella* from harsh environmental conditions such as desiccation [95, 350]. It is not clear what roles are played by O-antigen modifications, and the range of modifications catalyzed – and their distributions among strains – is not well studied.

1.6 WHY ARE DIVERSE H- AND O-ANTIGENS MAINTAINED IN *SALMONELLA*?

As is suggested by Figure 5, the phenotypic diversity of *Salmonella enterica* serovars results from the structurally distinct *rfb* operons they harbor, wherein non-homologous genes encode enzymes for the synthesis of different sugars and their attachment into structurally distinct polysaccharides to be placed on the exterior of the cell. Aside from the notable changes in gene inventory at the *rfb* operon, the genes flanking this locus show elevated diversity as well (Figure 6), again suggesting that variance-purging selective sweeps do not affect this region of the chromosome. This phenomenon was first described in the late 1980s, when alleles of the *rfb*-proximal *gnd* gene (see Figure 5 for location) proved far more diverse in strains of *E. coli* than alleles of other loci [80]. Unlike genes elsewhere in the *Salmonella* or *E. coli* chromosomes, genes flanking the *rfb* locus maintain very high levels of polymorphism; this increase in variability in the *rfb* region is evident when multiple genes are assessed using complete genome sequences (Figure 6). Rather than reflecting unusual selective regimes affecting these loci

directly, the excess variation has been attributed to linkage to the *rfb* operon [219, 239]. At loci unlinked to the *rfb* operon, beneficial alleles may arise by mutational processes. Selective sweeps operate to purge diverse alleles as the beneficial allele is transferred among strains by homologous recombination [e.g., see 108]. Such selective sweeps cannot occur in the proximity of the *rfb* operon, as recombinants with variant *rfb* alleles would have decreased fitness in the environment to which the original *rfb* allele was adapted.

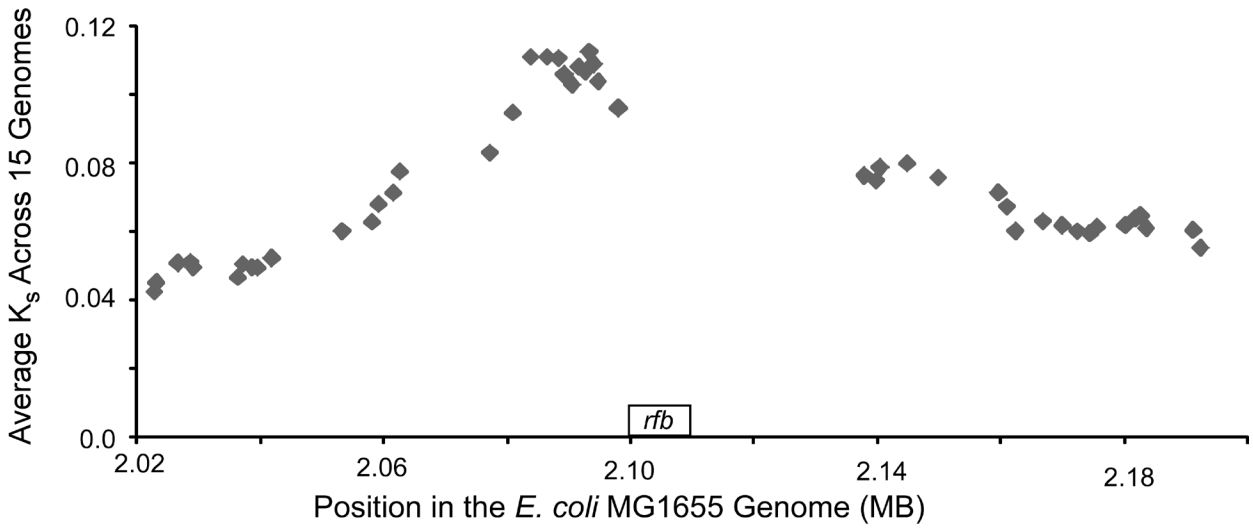

Figure 6. Genetic diversity near the *rfb* locus in *Escherichia coli*

The average divergence of synonymous sites among pairwise comparisons of genes shared among 16 completely sequenced *E. coli* genomes is plotted according to their position in the *E. coli* K12 genome. The diversity of the set of completely sequence *E. coli* strains enabled this genome-scale analysis; the poor sampling (as yet) of completely sequence *Salmonella* genomes precluded a robust analysis in that taxon at this scale. K_s = divergence at synonymous sites [196].

This excess genetic variability raises an interesting conundrum. As discussed above, many organisms have mechanisms for producing antigenic diversity at the capsular or O-antigen polysaccharide using generation-time-scale mechanisms that act upon otherwise statically-encoded information that is similar between strains. *Salmonella* utilizes such mechanisms to generate phenotypic diversity at the H-antigen and at fimbriae, where genotypically similar organisms can be phenotypically distinct. Yet *Salmonella* does not alter the O-antigen on a generational time scale. Thus, any model invoking frequency-dependent selection as a rapid response to changing environmental conditions – either to extend an infection or to infect non-naïve hosts – cannot apply to the diversity being maintained at the O-antigen-encoding *rfb* operon.

This is not to say that the H- and O-antigens do not experience such selective pressure. Like the H-antigen, the O-antigen is a target of the host innate immune system and recognized by the Toll-like receptor 4 (TRL4) [229, 288, 365]. The O-antigen, in complex with LPS, stimulates production of antibodies by the adaptive immune system, which has been shown to provide temporary protection against re-infection [74, 235, 280]. Because the O- and H-antigens elicit a host immune response, the conventional notion that frequency-dependent selective pressure from the immune system has been offered to explain *rfb* diversity. But given the lack of a mechanism to create generation-time-scale diversity, as well as the robust phenomenon of host/serovar specificity (Figure 2), the failure of many studies attempting to link O-antigen diversity to variations in pathogenicity and immune system evasion is not surprising [18, 32, 166, 219, 271, 358].

The lack of variation-purging recombination at or near the *Salmonella rfb* locus suggests that the different diverse forms are advantageous for the strains that stably harbor them. That is,

there is no *rfb* allele that is favored by all strains of *Salmonella*. Whereas many bacteria employ mechanisms for creating generation-time-scale diversity that ensures production of antigenically distinct progeny, *Salmonella* produces daughter cells that retain phenotypic identity with their parents at the O-antigen even while differing at other antigens. This suggests *Salmonella* strains that differ at the O-antigen lead significantly different lifestyles wherein each different O-antigen form provides an advantage not realized by other strains. While this argument essentially assigns the different serovars to distinct niches, I do not consider these strains to be representatives of wholly different species. Strains of *Salmonella* do experience interstrain homologous recombination, leading to species-wide selective sweeps. These sweeps simply do not occur at the *rfb* operon.

1.6.1 Diversifying selection in *Salmonella*

I suggest that the different O-antigens stably expressed by different serovars of *Salmonella* confer advantages in different environments; thus, genetic diversity above that predicted by the neutral theory would reflect the action of diversifying selection rather than frequency-dependent selection. Here, no single allele can confer a benefit in all environments, precluding a selective sweep at the *rfb* operon. Moreover, genes closely linked to the *rfb* operon also fail to experience selective sweeps since recombination there would likely result in problematic introduction of less-suited genes at the *rfb* operon (see Figure 6). Yet recombination at unlinked loci would still occur, providing strains of *Salmonella* with the genotypic and phenotypic cohesion expected of a bacterial species [80]. Thus, *Salmonella* strains could be considered different strains of the same species everywhere except at the *rfb* operon, where they carry adaptations to distinctly different environments. This raises fundamental questions regarding the role of recombination in

providing species-level genetic cohesion in *Salmonella*. In viewing the loss of recombination as a stepwise process during bacterial speciation [182, 275-276], this model suggests that the *rfb* locus may play a pioneer role in beginning the process of lineage separation.

The phenomenon of host-serovar specificity is consistent with this model. Unlike other bacterial pathogens, specific serovars of *Salmonella* are consistently associated with disease states in different host animals (Figure 2). That is, rather than requiring infecting strains present an O-antigen that is novel to the vertebrate host, *Salmonella* strains that mount successful infections consistently present the same antigen to particular hosts [264]. These data support the hypothesis that the nature of the host environment favors particular O-antigenic types of *Salmonella*.

This model is not at odds with the observation that other intestinal bacteria – for example, species of *Bacteroides* as discussed above – gain a benefit from producing daughter cells with variant O-antigens. Rather, I posit that the advantage gained by *Salmonella*'s retention of its parental O-antigen outweighs any detriment this lack of variability incurs. *Bacteroides* is a major constituent of the intestinal microflora [97, 194]; therefore, it is likely to be heavily sampled by the host immune system resulting in targeted IgA excretion [203]. Because *Salmonella* is such a rare member of the intestinal microbiota, it would not be targeted by the immune system and constant switching of the O-antigen would not be advantageous. Instead, I propose that the particular features of its parental O-antigen would provide more direct benefits.

1.6.2 Differential distribution of bacterial strains

Central to this hypothesis is the supposition that *Salmonella* is differentially distributed in natural environments. Beyond the pattern of host-serovar specificity (Figure 3), data from several

species of intestinal bacteria indicate that genotypes are differentially distributed among host species. This phenomenon has been exploited for Microbial Source Tracking, whereby the source of fecal contamination in water is preliminarily identified by virtue of the genotypes of contaminating bacteria found in the water [16, 309, 324]. Common methods for MST include rep-PCR, pulsed-field gel electrophoresis, viral typing, antibiotic-resistance profiles or multilocus sequence typing [105, 118, 236]; all methods are grounded in the observation that genotypes of intestinal bacteria are not randomly distributed in the enteric environments of mammalian hosts and exploit these patterns to infer the source of water contamination.

Differential distribution of bacteria begins at the species level, where, for example, genera of enteric bacteria are differentially distributed among major lineages of mammals showing that mammalian intestines are not all uniform environments [103]. Closely-related species within a genus are also differentially distributed, such as the strains of *Enterococcus* used in Microbial Source Tracking [383]; here, water being contaminated from untreated human waste water can be discriminated from runoff from a cattle farm by the relative abundances of *Enterococcus* strains. In addition, strains of *Bacteroides* have differential distribution among intestinal environments [61, 135, 278]. Lastly, and most importantly from my perspective, genotypically distinct strains within a single bacterial species can also be differentially distributed. For example, strains of *Escherichia coli,* another species widely used in Microbial Source Tracking [156, 266, 341], are differentially distributed among mammals, whereby mammals having different diets or dwelling in different environments harbor different genotype of *E. coli* [101-102]. The use of genotypic differences among bacteria found in different intestinal environment for Microbial Source Tracking indicates that these differences are stable, robust and repeatable.

Why do different environments favor *Salmonella* with different O-antigens? Although *Salmonella* is broadly noted for its pathogenic effects, it also dwells in intestinal environments as a harmless commensal, likely for a much greater percentage of time. For example, many captive reptiles asymptomatically shed *Salmonella* that can be transmitted to humans, especially small children or immunocompromised individuals, resulting in pathogenic infection [26, 58]. Interestingly, while the conditions in the reptile intestine favor a more commensal lifestyle for *Salmonella*, the introduction of these same cells into human intestines results in a switch to pathogenicity. Unlike *Neisseria* and *Haemophilus*, where the switch from commensal to pathogen is dependent on immune system evasion mediated by O-antigen phase variation, the same *Salmonella* cells harmlessly inhabiting reptiles cause gastroenteritis in humans without any change in the nature of the O-antigen. Particular O-antigens could contribute to differential survival in different environments independent of pathogenic behavior.

The reasons why *Salmonella* can adopt a commensal lifestyle in one organism and cause pathogenic infection in another organism are complex. Many genes could show adaptive differences in response to abiotic variation among environments, such as differences in oxygen tension, pH, ionic strength, salinity or the availability of nutrients. Yet it is difficult to attribute advantages to particular O-antigens in response to such differences. A particular O-antigen could also provide greater competitive abilities in certain environments; for example, they may mediate more effective adhesion to some intestinal mucins, resulting in a greater chance for invasion of intestinal epithelial cells. However, adhesion to intestinal mucins is very complex and most likely involves many factors, and the role of differential adhesion to mucin mediated by different O-antigens is not particularly well-tested. In addition, *Salmonella* possesses several other mechanisms for attachment, including fimbriae [54] and SPI-1 [167], making differential mucin

attachment an unlikely explanation for O-antigen diversity. Lastly, different O-antigens may provide defense against predation in the intestine.

1.7 PREDATION AS A SELECTIVE FORCE

The connection between O-antigen variation and predation is clear; the O-antigen is the most abundant molecule on the surface of the cell, thereby being a likely ligand for predator/prey interactions. Many organisms are potential predators of bacteria in intestinal environments, such as ciliates, bacteriophages and amoebae. However, the biological characteristics of both ciliates and bacteriophages make them unlikely sources of predation that could drive diversity at the *Salmonella rfb* locus, whereas amoebae have the requisite characteristics of predators that can exert selective pressure on the O-antigen.

1.7.1 Ciliates: nondiscriminatory predators

Ciliates are eukaryotic organisms that commonly prey upon bacteria in a variety of ecosystems [10, 42, 84, 169, 352, 382]. These organisms are generally classified as nondiscriminatory feeders, as they will consume any bacteria that are sufficiently small enough to pass through their feeding comb. Thus, *Salmonella* susceptibility to ciliate predation should not be impacted by the identity of the O-antigen, as these predators are more impacted by size of prey. Previous work from our lab demonstrated that while the presence of the ciliate *Tetrahymena pyriformis* reduced overall numbers of antigenically distinct *Salmonella* serovars as expected, *T. pyriformis* did not

discriminate among these prey [387]. Thus, predation from ciliates is too nonspecific in order to result in selective pressure exerting diversifying selection on the *Salmonella rfb* locus.

1.7.2 Bacteriophages: highly discriminatory predators

Bacteriophages, or viruses that infect bacteria, are important predators of bacteria in many environments [4, 27-28, 253, 318, 400]. Bacteriophages require the presence of specific receptors in the outside of host bacterial cells for successful infection [140]. Phages exploit specific host factors for intracellular reproduction; thus, it is not surprising that phage display specific recognition of prey. Point mutations that abolish specific binding epitopes on attachment proteins strongly impact the ability of phage to recognize their bacterial hosts [96, 222, 265, 340]. Such mutations cannot drive antigenic escape at *rfb*. The products encoded by the *rfb* operon are the synthases and transferases that manufacture sugars that compose the repeating O-antigen monomer. In order for the composition of the O-antigen to be altered, any given point mutation must destroy the functionality of the original synthase or transferase or confer a new function. Point mutations are unlikely to frequently result in these significant alterations to the same degree that they can confer dramatic changes in the ability of a phage to bind an outer membrane protein receptor. Although bacteriophages can use the O-antigen as a receptor, as in the case of the *Salmonella* phage P22 [147], this relationship is extremely specific, making it highly unlikely that the same phage could use more than one structurally distinct O-antigen as receptors to recognize prey. Thus, bacteriophage predation is too specific to act on the diversity observed at the *Salmonella rfb* locus.

1.7.3 Amoebae: generalized yet discriminatory predators

Amoebae are single-celled eukaryotic organisms that serve as general predators of bacteria; prey recognition here occurs primarily by cell-cell contact. Amoebae are abundant predators in water, soil, and intestinal environments [112, 283, 287]. Because phagocytotic amoebae rely on cell-cell contact to recognize their prey, then one would expect different serovars to be recognized with different efficiencies. Moreover, amoebae would be expected to be differentially distributed among vertebrate intestinal environments. As a result, amoebae could mediate diversifying selection at the O-antigen-encoding loci, allowing different serovars of *Salmonella* to gain fitness in particular environments where amoebae consume them less rapidly.

1.7.4 *Salmonella* most likely do not exhibit active responses to protozoan predation

For many reasons I posit that an active response to potential predation is not likely. *Salmonella* are non-motile in the gut, so cells do not have the capacity to swim away from predators unlike ciliates, which do exhibit a behavioral response to predation [177]. Bacteria do not have time to adapt behavioral responses to predation, as predation results in cell death. Because cell death is constantly occurring in the intestinal lumen, any chemical alarmone-mediated responses would be constitutively active independent of predation risk. The small size of bacteria prevents their escape by size refugia, unlike animals that can grow so large as to escape predation due to size [52, 228]. Defensive cell-well thickening is not seen in vegetative cells, although they are a feature of persistent spores; defensive structures – such as those seen in protozoa [93, 175, 384] or *Daphnia* [75], are neither evident nor thought to be effective since they would act on the molecular scale and do not impede chemical degradation in the food vacuole. Simply put,

bacteria are unable to adopt many of the strategies employed by higher organisms to escape predation.

A common framework for interpreting predator-prey relationships between higher organisms casts predation as a barrier that must be overcome to live in a most preferred environment; this does not necessarily apply to bacteria. Unlike higher organisms, bacteria possess a remarkable degree of metabolic diversity, making survival into other environments a less challenging task. Additionally, bacteria can rapidly adapt to new environments through the acquisition of genes via horizontal transfer, which cannot be accomplished in higher organisms. The key to the bacterial prey-predator relationship is that successful bacteria out-compete other bacterial prey, which leads to a fitness advantage in the intestine. They do not avoid predation entirely; as long as a given *Salmonella* serovar is consumed by predators less efficiently than other serovars, the given serovar has a better chance of survival in that particular environment than other serovars. Therefore, predation will impact the genetic structure of species, like *Salmonella*, which is found across multiple environments, resulting in differential distribution of serovars across environments.

1.8 PREDATION-MEDIATED DIVERSITY AT THE *SALMONELLA RFB* OPERON

Facets of this model have been established with rigor. First, it is clear that amoebae consume bacteria as prey in natural environments. Not only have bactivorous amoebae been isolated from ground water and soil, but also from intestinal environments [387]. Amoebae within the intestinal lumen consume intestinal bacteria, thus limiting both bacterial growth yield and persistence time within the lumen of any individual host [319]. I posit that amoebae are the

major general predators of bacteria found in vertebrate intestines, and that the feeding preferences of these amoebae affect the structure of intestinal bacterial populations.

For amoebae predators to influence the distribution of their prey, they cannot be randomly distributed in the environment; if they were, a prey bacterium could not persist in an environment where it could avoid all of the resident predators (since they would be constantly changing). Some survey experiments on pathogenic and commensal populations suggest that amoebae are not randomly distributed in the environment, with particular species of hosts harboring specific populations of amoebae. Differential distribution of pathogenic protozoa has been described for *Entamoeba*; *E. invadens* causes disease in reptiles [76], including ball pythons [170], whereas *E. histolytica* causes disease in humans [185, 262]. *E. suis* and *E. chattoni* infect non-human mammals, yet a related but distinct species preferentially infects birds [206]. The amoeba *Vannella platypodia* was found to infect multiple fishes [81], while *Neoparamoeba* preferentially colonizes gills [90]. The microsporidian *Encephalitozoon cuniculi* is a pathogen of rabbits and dogs, whereas *E. intestinalis*, *E. hellem,* and *E. bieneusi* are opportunistic pathogens of humans [377]. Commensal protozoa also show differential distribution among hosts. For example, the non-pathogenic amoeba *Paravahlkampfia ustiana* was isolated multiple times from the intestines of skinks [308]. Lastly, data collected in this laboratory show that there is differential distribution of *Naegleria*, *Hartmannella*, *Tetramitus* and *Acanthamoeba* among amphibian, fish and reptile intestinal tracts [386]. Based on this information, I conclude that the population of amoebae in a given host intestine is most likely stable and specific to that particular species of host. Therefore, when *Salmonella* cells enter an intestinal lumen, they are faced with predictable communities of amoebae that are encountered among all individuals of that host species.

Intestinal amoebae do not simply consume bacteria indiscriminately; rather, amoebae can discriminate between different bacterial strains provided as prey [387]. When presented different serovars of *Salmonella* as prey, amoebae will consistently consume one serovar more quickly than another, less preferred, serovar (Figure 7). In this example, the fitness of any given serovar is calculated relative to the group of serovars; those serovars that are eaten faster by a particular predator have a low relative fitness while those eaten more slowly have a higher relative fitness. This discrimination is evident even when amoebae are presented with both strains at the same time; one strain is consumed from the mixed population more quickly than its fitter competitor [387]. The basis for this discrimination is complex, but changes in the identity of O-antigen did influence predation risk [387]. Thus, I link the O-antigen to *Salmonella* susceptibility to predation.

Not only do amoebae discriminate among prey bacteria based on the nature of the O-antigen, different amoebae also have different feeding preferences (Figure 7). For example, while strain SARB36 was not readily consumed by the amoeba *Naegleria gruberi* strain NL, it was the most preferred strain when facing *Naegleria* strain F1-9 (Figure 7). The feeding preferences of amoebae shown in Figure 7 do not show any significant similarity ($R=0.06$, $P> 0.1$) and these amoebae were isolated from different environments. This leads to the possibility that *Salmonella* serovars may experience differential survival in different environments as the result of their ability to avoid the resident amoeboid predators in that environment. That is, differential susceptibility to predation by amoebae in a given host could influence which serovars from the entire *Salmonella* population maybe best suited to survive there. Because the intestinal lumen of each host species harbors different populations of predators with different feeding

preferences, no one form of the *Salmonella* O-antigen can confer enhanced resistance to predation in all environments.

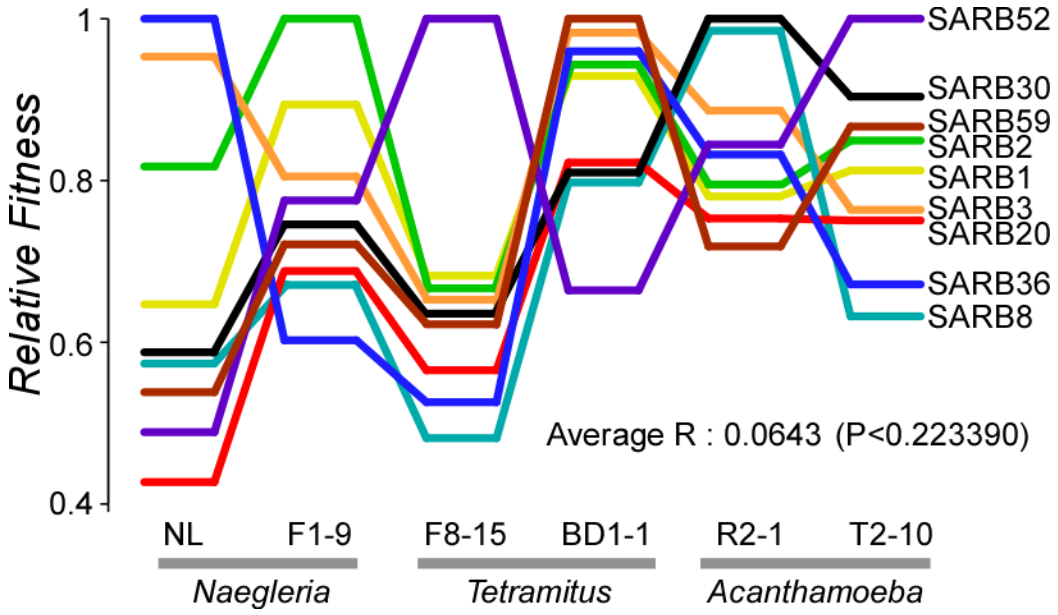

Figure 7. Relative fitness of *Salmonella* strains against protozoan predators

Each amoeba was tested against 9 SARB strains. The feeding preferences of each predator were determined separately against 9 antigenically diverse serovars of *Salmonella*. The least preferred strain was assigned a fitness of 1.0 for each predator.

One prediction of this model is that unrelated predators that inhabit a single environment must share feeding preferences, which would allow a single serovar of *Salmonella* to escape predation by the suite of predators it would face in its preferred environment. When predators were isolated from the intestinal tracts of fish, their feeding preferences were significantly more similar than one would expect [386]. In the example shown here (Figure 8A; $P < 10^{-5}$), even unrelated amoebae from two families isolated from the same host species shared a common set of feeding preferences. The similar fitness values of the 5 tested serovars against the 16 predators isolated from the same environment in Figure 8A stands in stark contrast to their varying fitness values against 6 predators from different environment shown in Figure 7 [387]. Overall, predators isolated from the same environment (including the intestinal tracts of goldfish, tadpole or turtles) shared feeding preferences (closed markers in Figure 8B), whereas those from different environments do not (open markers in Figure 8B).

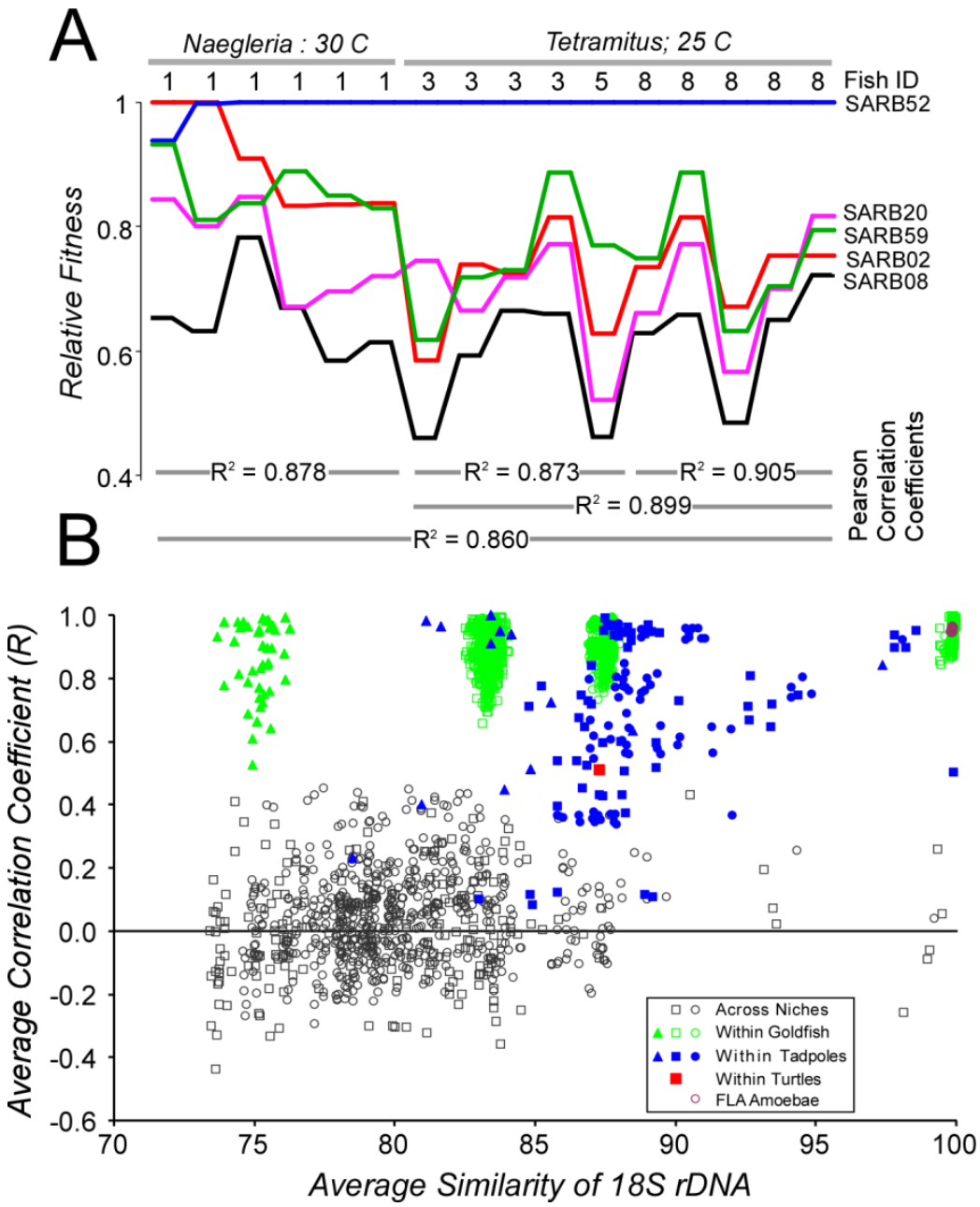

Figure 8. Protozoa can present uniform selective pressure on bacteria in a single environment

A. Fitness of *Salmonella* against 16 predator amoebae isolated from four separate goldfish. Amoebae are numbered according to the fish from which it was isolated (fish number 1, 3, 5 or 8). **B**. Relationship between feeding preference and environment. Average correlation coefficients are calculated for feeding preferences of 2, 3 and 4 different predators; these values are plotted against the average similarity of their 18S rDNA genes. Data for amoebae isolated from different environments are shown in open gray markers, and data from amoebae from the same environment are shown in black and/or closed markers.

Unlike the dynamic interactions between pathogens and host immune systems that are observed with *Neisseria* and *Haemophilus*, *Salmonella* serovars face stable selective pressure from predation each time it encounters an individual of a particular species of host. The expression of an O-antigen that confers enhanced resistance to predation in that particular environment affords that serovar with a greater chance of survival and replication within that host intestine. In contrast, other serovars are more readily consumed by the resident predators, eliminating those serovars from the environment in a manner akin to clearance from a host of *Neisseria* or *Haemophilus* strains having O-antigens that are quickly recognized by a non-naïve immune system.

The mechanistic basis for this congruence in feeding preference among unrelated amoebae isolated from the same host is not clear. Amoebae encounter two different sets of carbohydrates in their environment. The intestinal epithelium is covered in mucins, and binding to these carbohydrates allows the protozoa to remain in the lumen and avoid expulsion. The bacteria they consume are covered in LPS carbohydrates, and these experiments have demonstrated that they use the O-antigens to discriminate among prey [386-387]. Here, I propose a model to explain the mechanistic basis for protozoan feeding preference: intestinal amoebae recognize host mucins as attachment sites, and unrelated amoebae share an inclination to bind to the intestinal wall without trying to consume it [386]. This model proposes that amoebae bind differently to food than they do to intestinal mucins. If a bacterial O-antigen resembles the intestinal mucins of its host, it could act as molecular camouflage. This similarly would confound intestinal amoebae attempting to discriminate between food and housing. Such camouflaged bacteria would benefit from increased resistance to predation over serovars with O-

antigens that are different from the mucins of that host. In effect, bacteria may better escape predation by molecular mimicry of host mucins.

1.9 GOALS OF THE DISSERTATION

While previous work demonstrated a connection between the O-antigen and fitness against predation [386-387], a clear link between *rfb* diversity and fitness against predation had yet to be established. The next steps in investigating the potential for protozoan predation to drive diversifying selection at the *Salmonella rfb* locus was to a) establish the O-antigen as the major surface antigen affecting fitness against predation and b) determine if protozoa can discriminate among *Salmonella in vivo*. These issues are central to the possible validation of the hypothesis that selective pressure from protozoan predation drives diversifying selection at the *Salmonella rfb* locus. However, the techniques required to address these questions did not exist prior to the work outlined here: not only was genetic manipulation of natural *Salmonella* strains impossible, but an experimental system designed to collect large amounts of data for competition tests in complex environments did not exist. Thus, the work presented in this dissertation developed the necessary techniques to more clearly establish a link between the *Salmonella* O-antigen and fitness against protozoan predation. When accomplished, this research will provide a novel framework for the maintenance of diversity by environmental forces other than the host immune system and establish a mechanism for the initiation of fragmented speciation in bacteria [182, 275-276].

2.0 DEVELOPMENT OF BACTERIOPHAGE P1-MEDIATED GENETIC MANIPULATION OF NON-TYPHIMURIUM *SALMONELLA*

Prior research established that predators could discriminate among different bacterial prey, but it was not clear that the O-antigen was the moiety being used to discriminate among them [386-387]. To establish amoebae predation as a likely source of selection for the maintenance of *rfb* diversity in *Salmonella*, I needed to assess the relative contribution of different antigenic determinates to the susceptibility of *Salmonella* to predation. To do this, I needed to manipulate the genes encoding surface antigens among natural isolates of *Salmonella*.

Previous work attempting genetic manipulation of wild-type Salmonellae exploited naturally occurring sensitivity to bacteriophages P22 and ES18 to create near-isogenic strains of *Salmonella* that only varied at the *rfb* locus [387]. However, the use of these transducing bacteriophages was very limited due to widespread lack of host sensitivity among non-Typhimurium *Salmonella* and was limited to serovars that were very similar to *S. enterica* Typhimurium [387]. Thus, this approach was woefully insufficient to investigate the full spectrum of antigenic diversity in *Salmonella* with respect to protozoan predation. Although molecular approaches to strain construction were considered, these methods were deemed impractical for my need to transfer large amounts of DNA, including the ~30 kb *rfb* locus, among strains. Instead, I explored the possibility of adapting existing bacteriophage transduction techniques for use in natural isolates of *Salmonella*. Here, I describe the development of

bacteriophage P1-mediated genetic manipulation in non-Typhimurium strains of *Salmonella enterica*. This approach represents a major step forward for the study of *Salmonella*, and the work I present in this chapter enabled all of the experiments described in this dissertation.

2.1 BACTERIOPHAGE-MEDIATED GENETICS IN *SALMONELLA*

Use of *Salmonella* as a model organism for the study of bacterial genetics, physiology and pathogenicity has been dramatically impacted by the relative ease of genetic manipulation possible in serovar Typhimurium. In particular, the laboratory serovar Typhimurium LT2 and its corresponding transducing bacteriophage P22 [302] have been widely used to explore the biology of *Salmonella*. Phage P22 uses the *rfb*-encoded O-antigen polysaccharide as an attachment site, and loss of the O-antigen via mutations in the *rfb* locus renders cells immune to phage P22. Successful phage P22 infection apparently requires cells to display Typhimurium-like (1,4,[5],12) O-antigen serotypes. While temporary disabling of restriction endonucleases via heat-shock renders other serovars of the Typhimurium subgroup bearing this O-antigen (*e.g.,* SaintPaul, Heidelberg, Paratyphi B) sensitive to P22 [83], infection does not occur in strains outside of the Typhimurium subgroup. Although naturally immune to phage P22, the closely related bacterium *Escherichia coli* can be rendered sensitive to phage P22 by the introduction of the Typhimurium *rfb* operon [238], forcing cells to express the proper O-antigen structure permissive for infection. Thus, while P22 is an effective tool for genetic manipulation of serovar Typhimurium, it cannot be used among natural isolates of *Salmonella* bearing different O-antigens, including serovars that persistently cause disease in agriculturally important animals.

Bacteriophage P1 has been used extensively for the genetic manipulation of common laboratory strains of *Escherichia coli* and *Klebsiella aerogenes* [23, 171, 191, 207, 312-313, 343, 346, 349, 357, 399]. Despite the close relatedness of *Salmonella* to *E. coli* and *K. aerogenes*, both laboratory and wild-type strains of *Salmonella* are naturally resistant to P1. However, mutations in the *galE* gene confer sensitivity to phage P1 in serovar Typhimurium LT2 [248]; these mutants are unable to epimerize UDP-glucose to UDP-galactose, a necessary building block for the lipopolysaccharide (LPS) core antigen that forms the foundation for anchoring the O-antigen to the outer membrane (Figure 1). UDP-galactose is also required by many serovars of *Salmonella* to initiate O-antigen biosynthesis [293]. As a result, *galE* mutants cannot present O-antigens on the outside of the cell, thus exposing the membrane-proximal portion of the LPS core antigen. Although its exact role in determining sensitivity to phage P1 is unclear, the LPS core antigen influences sensitivity to phage P1 [248]. Producing phage P1 sensitivity in *Salmonella* through the *galE* mutation has been used to move DNA between serovar Typhimurium and *E. coli* [215, 357], and to manipulate strains of Typhimurium that are resistant to P22 [86]; otherwise, phage P1 has not been extensively used in *Salmonella*.

Despite the relative paucity of use of phage P1 as compared to phage P22 in *Salmonella*, both phages have distinct advantages over one another for various types of experiments. Phage P22 may be directly used in serovar Typhimurium LT2, whereas phage P1 requires a mutation in *galE* for use in *Salmonella* that may require an additional repair step in strain construction. For transducing large amounts of DNA between strains, phage P1 has a clear advantage over phage P22, as phage P1 reliably transduces more than twice the amount of DNA than phage P22. The much larger capsid size of phage P1 enables the packaging of more than 100 kb of DNA in a transducing particle [117, 339], whereas phage P22 only encapsulates ~44 kb of DNA in

transducing particles [82, 150]. Thus, phage P1 is the better choice for the manipulation of large genomic regions, such as large pathogenicity islands like the O-antigen producing *rfb* locus, among strains. The advantage in DNA packaging of phage P1 is offset by the lower rate of production of transducing particles; P1 lysates contain ~0.1% transducing particles, compared to ~40% for commonly-used terminase mutants of P22 [302]. Therefore, phage P22 is best suited for experiments requiring large numbers of transductants within serovar Typhimurium. Both phages can be useful in the genetic manipulation of *Salmonella enteric* Typhimurium, although the ease of use of phage P22 has established it as the primary tool for *Salmonella* genetics.

Although phage P1 has the potential to permit the rapid, reliable transfer of large genomic regions among strains, the utility of phage P1 has not been examined in other, non-Typhimurium serovars of *Salmonella*, likely for two reasons. First, the receptors for bacteriophage attachment are notoriously hypervariable and experience strong selection for variant forms conferring phage resistance. For example, variation in the LamB maltose receptor among strains of *E. coli* results in strain-specific resistance to phage λ [57, 130, 348]. Second, coimmune prophages often leave cells resistant to phages of interest, even if cells present the appropriate receptors for phage attachment. For these reasons, one would not expect any one bacteriophage to infect diverse members of any bacterial species.

The lack of sensitivity to phage P1 among non-Typhimurium *Salmonella* has been casually attributed to these two reasons, although no direct experimental evidence suggests that non-Typhimurium *Salmonella* either lack the appropriate receptor for phage P1 attachment or contain coimmune prophages that confer resistance to phage P1. Phage P1 infects a wide variety of Gram-negative bacteria, including *Salmonella*'s close relatives *Escherichia coli* and *Klebsiella aerogenes* [191, 343]. The broad host range implies that the receptor for phage P1, likely the

LPS core antigen, is conserved enough within the Enterobacteriaceae to permit infection in a variety of hosts. Serovar Typhimurium LT2 must possess the correct receptor for phage P1 attachment, because *galE* mutants of *Salmonella* gain sensitivity to phage P1. If such a large group of bacteria are able to display the receptor for phage P1, then it is highly likely that non-Typhimurium *Salmonella* also possess the phage P1 receptor, which is unmasked with the use of *galE* mutations. I further discuss the nature of the receptor for phage P1 in *Salmonella* in Chapters 2.3 and 2.9.

Second, coimmune prophages that confer resistance to phage P1 are not evident in the genomes of *Salmonella* sequenced to date [13, 38, 53, 56, 132, 149, 209, 250, 277, 351, 390], suggesting that coimmunity to phage P1 may not be an issue in non-Typhimurium strains. Thus, it is reasonable to suspect *galE* mutations will confer sensitivity to phage P1 in non-Typhimurium serovars of *Salmonella*. In this chapter, I discuss the development of genetic manipulation of non-Typhimurium *Salmonella* using bacteriophage P1.

2.2 MATERIALS AND METHODS

2.2.1 Media and growth conditions

All bacterial strains were routinely propagated overnight at 37°C in LB medium prepared as 1% tryptone, 0.5% yeast extract, and 0.5% NaCl; plates were made using 1.2% agar. E medium [371] supplemented with 0.2% glucose and carbon-free E medium (NCE; prepared as E salts without sodium citrate) supplemented with 10 mM $MgSO_4$ and 0.2% galactose are the minimal defined media used in these experiments. Biotin and nicotinic acid were added at 0.002% and

0.001%, respectively. Green Indicator Plates for screening strains for the presence of P22 lysogens were prepared as 0.8% tryptone, 0.1% yeast extract, 0.5% NaCl, 6.5 mg/mL aniline blue, 62.2 mg/mL alizarin yellow, 1.5% agar, and 0.84% glucose. Transducing Broth was prepared as LB supplemented with 1X E salts and 0.2% glucose. P1 Dilution Buffer is 0.9% NaCl supplemented with 5 mM $CaCl_2$ and 10 mM $MgSO_4$. P1 Top Agar was prepared as LB medium supplemented with 0.7% agar, 5 mM $CaCl_2$, and 10 mM $MgSO_4$. Kanamycin was used at 20 μg/mL for routine propagation and at 30 μg/mL for selection of directed gene knockouts, tetracycline at 20 μg/mL for selection of plasmid-borne markers and 10 μg/mL for selection of chromosomal markers, chloramphenicol at 20 μg/mL, spectinomycin at 100 μg/mL, hygromycin at 100 μg/mL, gentamycin at 30 μg/mL, and ampicillin at 100 μg/mL. Non-Typhimurium strains used in these experiments were obtained from the *Salmonella* Reference Collection B (SARB) [35]; SARB strains 1, 3, 8, and 36 were found to be defective for purine biosynthesis and require the addition of 0.008% guanosine to minimal media to support growth.

2.2.2 Bacteriophage P22 and ES18 propagation and transduction

Lysates of bacteriophages P22-HT *int*-205, P22-H5, and ES18 were prepared using the protocols of Davis *et al* [69] and Schmieger and Schicklmaier [303]; overnight cultures of *S. enterica* in LB broth were diluted 1:6 in Transducing Broth, and phage was added to a final titer of 10^8 PFU/mL. The lysate was incubated overnight with shaking at 37°C, sterilized with chloroform, and stored with chloroform at 4°C. Transductions with phages P22 and ES18 were performed by mixing 100 μl of an overnight bacterial culture with 100 μl of phage lysate diluted to a titer of 10^8 PFU/mL, incubating with shaking at 37°C for 1 hour, and plating the transduction mixture directly to selective media. For phage P22 transductions, selective media was supplemented with

10 mM EGTA as a magnesium chelator to decrease phage infection on plates. All transductants obtained via P22 were struck on Green Indicator Plates and assayed for sensitivity to P22-H5 in order to screen for the presence of P22 lysogens. For use of P22 with strains containing *galE* mutations, LB broth was supplemented with 0.2% glucose and 0.02% galactose, and all other growth media was supplemented with 0.02% galactose.

2.2.3 Bacteriophage P1 propagation and transduction

Lysates of bacteriophage P1 *vir* were prepared using an alteration of the protocol of Kolko *et al* [171]; overnight cultures of *S. enterica galE* or *rfb* mutant strains were diluted 1:100 in LB supplemented with 5 mM $CaCl_2$ and 0.2% glucose and incubated with shaking for 30 minutes at 37°C. Following initial incubation, P1 *vir* was added to a final titer of 10^7 PFU/mL, and the lysate was allowed to incubate for up to 6 hours. Lysates that cleared within 6 hours were sterilized by chloroform addition and stored at 4°C. Recipient cells were prepared for transduction as follows: grow to mid-log phase, pellet cells by centrifugation, resuspend in ¼ volume of P1 Dilution Buffer, and incubate on ice for 30-45 minutes. Following incubation, 100 µl of recipient cells was added to 100 µl of P1 *vir* lysate at neat and 10^{-1} dilutions. Transduction mixtures were incubated with shaking at 30°C for 60 min and plated directly to the appropriate medium following addition of 150 µl of 1 M sodium citrate to each transduction mixture as a calcium chelator to reduce further infection of cells by P1 *vir*.

2.2.4 Bacteriophage P1 *vir* plaque assay

Strains were assessed for sensitivity to phage P1 using two methods: lysate preparation in liquid media and plaque assay on solid media. Plaque assays were conducted by diluting 500 µL of cells grown overnight at 37°C into 5 mL LB and incubating with shaking for 60 minutes at 37°C. Cells were pelleted with centrifugation and resuspended in 2.5 mL P1 Dilution Buffer; 100 µL cells were mixed with 2.5 mL P1 Top Agar and plated to a room-temperature LB plate. P1 *vir* lysate was serially diluted to 10^{-8} in P1 Dilution Buffer and spotted onto plates in 10 µL aliquots; plates were allowed to dry at room temperature. Plaque formation was assessed after overnight incubation at 37°C.

2.2.5 Directed gene replacements in *Salmonella enterica* serovar Typhimurium LT2

Gene replacements followed the procedure of Murphy *et al* [230], Poteete and Fenton [260] and Datsenko and Wanner [68]. Primers were designed so that their 3' ends were complementary to sequences flanking the antibiotic resistance cassettes *cat* (chloramphenicol), *aph* (kanamycin), *tet* (tetracycline), *hyg* (hygromycin) and 45 bases at their 5' ends complementary to sequences within the *S. enterica* gene targeted for replacement. PCR amplification of these cassettes produced an antibiotic resistance gene flanked on both ends with at least 40-45 bp of sequence identity to the *S. enterica* gene of interest. Allelic replacement of genes was performed by electroporation of the PCR fragment into a host cell containing either pTP223 [230, 260] or pKD46 [68], which both contain an experimentally inducible version of the λ*red* recombinase system. Wild-type or *galE* mutant strains containing these plasmids were used as recipient cells

for directed gene replacements; the *galE* mutant background was used to facilitate backcrossing of constructs to strains lacking the λ*red* recombinase system.

Competent host cells containing pTP223 were prepared by diluting cells 1:200 in LB broth containing tetracycline and 1 mM IPTG to induce the λ*red* recombinase system and incubating at 37°C with shaking to late-log phase. Competent host cells containing pKD46 were prepared by diluting cells 1:200 in LB broth containing ampicillin and 10 mM L-arabinose to induce the λ*red* recombinase system and incubating at 30°C with shaking to late-log phase. Cells were concentrated with centrifugation and washed at least three times with ice-cold sterile deionized water. Electroporation was performed by adding up to 10 μL of DNA to 40 μL competent cells in an electroporation cuvette with a 0.2 cm gap. 500 μL of ice-cold Transducing Broth was added to each cuvette after electroporation, and cuvettes were incubated on ice for at least 5 minutes. Electroporation mixtures were transferred to culture tubes and incubated at 37°C with shaking for 1-2 hours. Following recovery incubation, cells were plated directly to selective media and incubated overnight at 37°C. Replacement of chromosomal genes with antibiotic resistance cassettes was confirmed by PCR amplification of the target locus, associated phenotypes, and DNA sequencing when possible. For the creation of the *galE*-6867*::aadA* strain KAB705, primers were designed with 3' end complementary sequence to the spectinomycin resistance cassette *aadA* and 5' end complementary sequence to the kanamycin resistance cassette *aph* so that the resulting PCR-amplified fragment targeted the *aph* gene for replacement, simply replacing one antibiotic resistance gene for another. All constructs were backcrossed into the LT2 laboratory strain of *S. enterica* via P1 or P22 bacteriophage transduction.

2.2.6 Technical acknowledgements

I designed all primers used to create and verify directed gene replacements of all strains listed in this chapter except Δ*orgC-prgH*-5800::*cat* and Δ*rfbP-rfbB*-2772::*cat*, which were created by Hans Wildschutte. Under my supervision, undergraduate researchers Ben Cross constructed strains KAB701-704 and Jessica Cheek constructed strain KAB705. Research technician Jessica Ravenscroft performed the following directed gene replacements used in the construction of other strains: Δ*rfbP-rfbB*-2773::*hph*, Δ*nadA*-0468::*aph* and Δ*bioABCDF*-1100::*cat*.

2.3 P1 SENSITIVITY IS CONFERRED BY MUTATIONS IN THE *GALE* AND *RFB* LOCI IN SEROVAR TYPHIMURIUM LT2

The exact identity of the receptor for phage P1 is unclear, although it has been suggested that the terminal glucose of the LPS core antigen is involved in phage P1 adsorption to *E. coli* K12 [297-298]. Mutations in *galU* produce LPS core antigens depleted in glucose and result in resistance to phage P1 in *E. coli* K12 [92]. However, both mutations in either *galU* or *galE*, which causes the formation of galactose-depleted LPS core antigens, confer sensitivity to phage P1 in *Salmonella enterica* Typhimurium LT2 [248]. These mutations produce truncated forms of LPS core antigens in both *E. coli* and *Salmonella* and result in the inability to display O-antigens on the outside of cells (Figure 9A). It has been argued that the presence of the O-antigen causes a steric inhibition of phage P1 attachment to its purported polysaccharide receptor within the LPS core antigen in *Salmonella* [398]. However, it is unknown why retention of the O-antigen permits sensitivity to phage P1 in *E. coli* K12 but not in *Salmonella*, considering that the

polysaccharide structures of the *E. coli* and *Salmonella* LPS core antigens are very similar [157]. Although these genetic data link the LPS core antigen to phage P1 infection, it is unknown if these polysaccharides are actual attachment sites for phage P1. Alteration of the LPS core antigen could result in the formation of a structure that mimics the genuine receptor for phage P1 present on the *E. coli* K12 O-antigen.

Because the nature of the receptor for phage P1 is especially ambiguous for *Salmonella*, I investigated the role of O-antigen chain length in sensitivity to phage P1. It is possible that a less severe truncation of the O-antigen than what results from a *galE* mutation could confer sensitivity to phage P1 while still permitting some presentation of the O-antigen on the outside of cells. If cells having O-antigens shorter in length than wild-type are sensitive to phage P1, then it is possible to generate phage P1-sensitive strains of non-Typhimurium *Salmonella* without completely disrupting the O-antigen. Because the O-antigen is a key *Salmonella* pathogenicity factor and is crucial to my line of investigation into how surface antigens contribute to fitness against protozoan predation, determination of the maximum O-antigen chain length that permits phage P1 infection can reduce the number of steps required for genetic manipulation of my strains.

I constructed a series of strains in the serovar Typhimurium LT2 background bearing O-antigens of different lengths, and I tested the sensitivity of these strains to phages P22 and P1 by transducing a neutral marker unrelated to O-antigen chain length into these strains. As shown in Figure 9A, strains with mutations in *galE* and *rfb* completely lack the presence of an O-antigen on the outside of the cell, with *galE* strains producing a slightly truncated form of the LPS core antigen while *rfb* strains produce a fully intact core antigen. I experimentally controlled chain length in cells displaying O-antigens through the introduction of mutations in the O-antigen

chain length determining proteins Rfc, WzzB, and FepE (Figure 9B, Table 2). O-antigen chain length distribution is bimodal in serovars Typhimurium and Enteriditis [99, 124, 231], with the Rfc polymerase generating O-antigens with 1-16 repeats in length [59], upon which the WzzB chain length determinant adds additional monomers to the O-antigen creating antigens of 16-35 repeats in length [17]; FepE regulates the production of O-antigens of 100+ repeats in length [231]. Because both donor and recipient cells are genetically identical in terms of host restriction systems, any differences in transduction efficiency can be attributed to differences in the ability of phages P1 or P22 to infect each cell type.

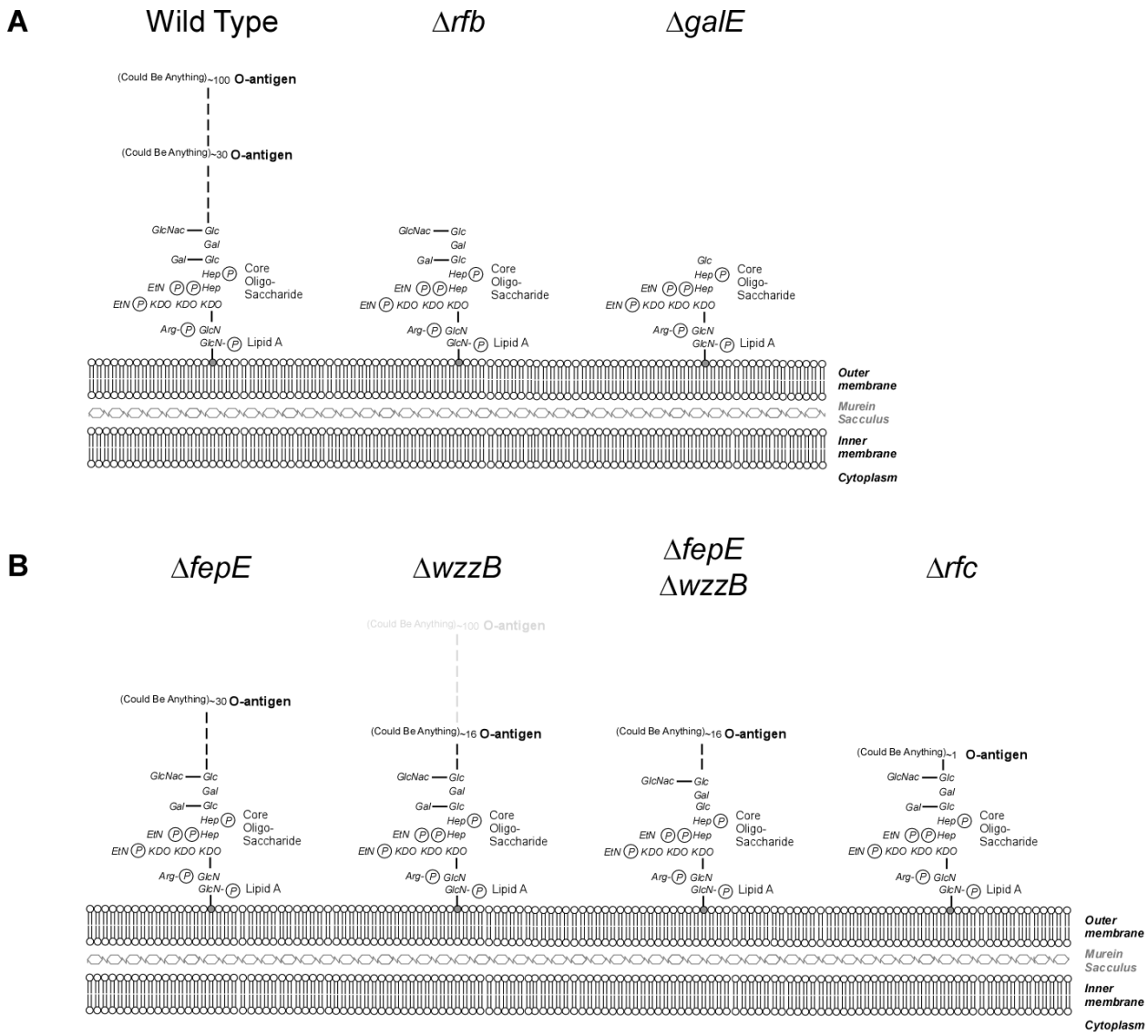

Figure 9. Lipopolysaccharide structure in *Salmonella*

A. Truncations in the lipopolysaccharide core antigen conferred by *galE* and *rfb* mutations compared to wild type. **B.** O-antigen chain length mutations conferred by mutations in *fepE*, *wzzB*, *fepE* and *wzzB*, and *rfc*. Depending on expression of *fepE* in the Δ*wzzB* background, production of O-antigens may be biased towards very large chains (100+ monomers).

Table 2. Strains used to examine the role of O-antigen chain length in sensitivity to bacteriophage P1

Strain	Genotype	O-antigen Chain Length
KAB002	Wild-type	16-35; 100+
KAB701	*wzzB*-8771::*aph*	1-16; 100+
KAB702	*fepE*-2910::*tet*	16-35
KAB703	*wzzB*-8771::*aph fepE*-2910::*tet*	1-16
KAB704	*rfc*-1727::*cat*	1
KAB030	Δ*rfbP-rfbB*-2772::*cat*	None; fully intact core antigen
KAB007	*galE*-6866::*aph*	None; truncated core antigen

Ornellas and Stocker confirmed that bacteriophage P1 is able to infect both *galE* and *rfb* mutant strains of *Salmonella*, although infection is weaker for *rfb* strains than for *galE* strains [248]. I confirmed these results in my Δ*rfbP-rfbB*-2772::*cat* and *galE*-6866::*aph* strains. Next, I assessed the ability for bacteriophages P22 and P1 to infect O-antigen chain length mutant strains of *Salmonella* listed in Table 2 with the exception of the Δ*rfbP-rfbB*-2772::*cat* strain, which was omitted in these experiments, by transducing the *galE*-6867::*aadA* marker into these strains. As shown in Table 3, only KAB007 is sensitive to transduction by bacteriophage P1, although all chain length mutants with the exception of *rfc*-1727::*cat* can be transduced using bacteriophage P22. Complete absence of the O-antigen is required for sensitivity to phage P1 in *Salmonella*, as the one O-antigen monomer present in KAB704 (*rfc*-2772::*cat*) is sufficient to block phage P1 infection. These data would argue against the sensitivity of wild-type strains of *E. coli* to phage P1 arising from the lack of steric hindrance of the phage P1 attachment site within the LPS core; rather, it seems likely that some polysaccharide portion of the *E. coli* K12 O-antigen itself is recognizable by phage P1. In *Salmonella*, the LPS core antigen must be completely exposed to permit phage P1 infection. Interestingly, both the full and truncated forms of the LPS core antigen resulting from the *rfb* and *galE* mutations respectively conferred sensitivity to phage P1, which suggests that a) removal of the O-antigen unmasks the receptor for phage P1 in *Salmonella*; and/or b) phage P1 is able to attach to different portions of the LPS core antigen.

While the single O-antigen subunit borne by *rfc* mutants was sufficient to block phage P1 attachment to *Salmonella*, it was insufficient to promote phage P22 attachment, as *rfc* mutants were nearly insensitive to both phages (Table 3). Although all strains having an O-antigen chain length greater than 1 were sensitive to phage P22, I observe a general trend towards increased sensitivity to phage P22 with increasing O-antigen chain lengths, with any chain length greater

than 16 representing maximal infectivity. The slightly reduced sensitivity of the *galE* mutant strain KAB007 was most likely due to the dependence of O-antigen production on the amount exogenously added galactose to growth media; thus O-antigen production in this strain should be considered variable. Since strains bearing partial O-antigens remain P1 resistant, I examined the ability of mutations that eliminated the O-antigen entirely to confer P1-sensitivity to non-Typhimurium serovars of *Salmonella enterica*.

Table 3. Sensitivity to bacteriophages P22 and P1 of O-antigen chain length mutants of *Salmonella* as determined by transduction.

Amount of countable transductants (30-300 colony forming units per plate) observed are indicated on a (+/-) scale in relation to the dilution of transducing bacteriophage added to cells. ++++ = 10^{-2} (P22) or 10^{-1} (P1) phage dilution; +++ = 10^{-1} P22 phage dilution; -/+ = less than 30 colonies per plate observed for undiluted P22 or P1 phage.

Strain	Genotype	O-antigen Chain Length	Sensitivity to Phage P22	Sensitivity to Phage P1
KAB002	Wild-type	16-35; 100+	++++	-
KAB701	*wzzB*-8771::*aph*	1-16; 100+	++++	-
KAB702	*fepE*-2910::*tet*	16-35	++++	-
KAB703	*wzzB*-8771::*aph* *fepE*-2910::*tet*	1-16	+++	-
KAB704	*rfc*-1727::*cat*	1	-/+	-/+
KAB007	*galE*-6866::*aph*	None; truncated core antigen	+++	++++

Although mutations in both genes conferred sensitivity to phage P1, I chose *galE* over *rfb* as the most appropriate candidate for use to explore the potential to permit phage P1-mediated genetics in non-Typhimurium *Salmonella* for three reasons. First, the *galE* gene, unlike the *rfb* locus, is well-conserved among *Salmonella*, which permits the design of molecular-based directed gene replacement strategies that should work for a large array of non-Typhimurium strains (Figure 10). Second, the *galE* mutation is easily repaired after completion of other genetic manipulations. Because *rfb* mutant strains are also sensitive to phage P1, a strain with an intact copy of *galE* and a mutant *rfb* locus can be used as a donor in a phage P1 transduction to repair the *galE* mutation and thus restore wild-type display of the O-antigen. This transduction is easily accomplished by selecting transductants with the ability to degrade galactose on minimal defined media. Third, I am reluctant to eliminate and restore the entire O-antigen region in wild-type strains, the target of my analysis, since that would increase the opportunity for unforeseen rearrangements to alter the phenotypes of these cells.

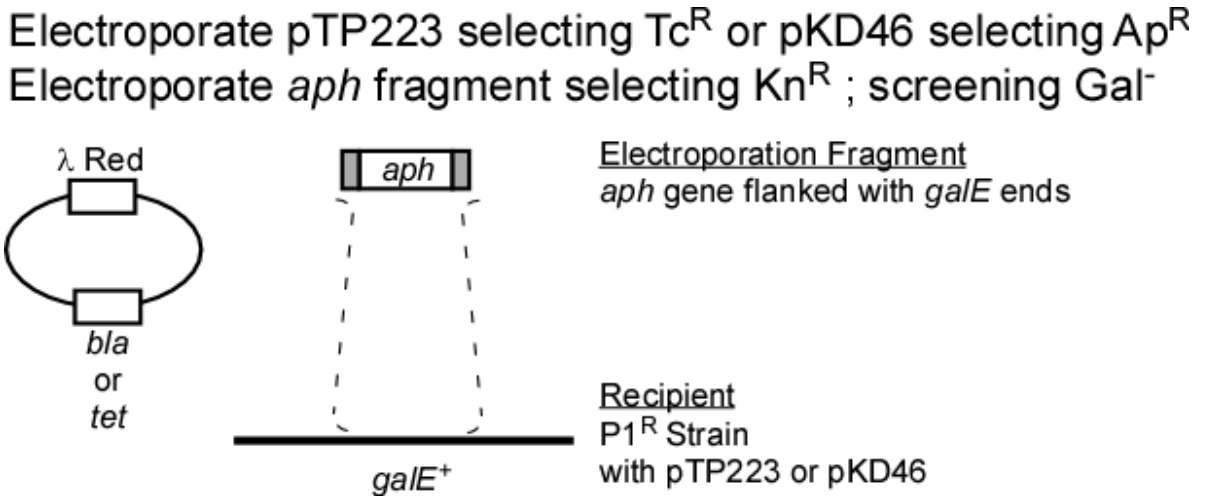

Figure 10. Conferring sensitivity to bacteriophage P1 in *Salmonella*

Directed gene replacement of *galE* mediated by λ *red* recombinase activity of pTP223 [230, 260] or pKD46 [68]. Polymerase chain reaction is used to generate the kanamycin resistance cassette *aph* flanked with complementary sequences to the *Salmonella enterica galE* gene. The intact copy of *galE* is replaced with the kanamycin resistance cassette *aph*. Dashed lines indicate approximate sites of recombination. Tc = tetracycline; Kn = kanamycin; Gal$^-$ = inability to degrade galactose; R = resistant; S = sensitive.

2.4 REVERSAL OF THE *GALE* MUTATION

All serovars of *Salmonella* must acquire a mutation in either the *galE* or *rfb* locus to permit sensitivity to phage P1. Once genetic manipulations have been completed, these mutations must be reverted in order to permit proper display of the O-antigen. I engineered a serovar Typhimurium LT2 donor strain KAB371, which lacks an intact *rfb* locus (Δ*rfbP*-*rfbB*-2773::*hph*) and wherein the intact *galE* gene was flanked by mutations in the *bio* operon (Δ*bioABCDF*-1100::*cat*) and *nadA* gene (*nadA*-0468::*aph*) (Figure 11). Using this strain as a donor, *galE* mutant cells can be transduced to Gal^+ on minimal media lacking biotin and nicotinamide and containing galactose as a sole carbon source. Use of the flanking markers Δ*bioABCDF*-1100::*cat* and *nadA*-0468::*aph* restricts the size of the incoming DNA fragment, thereby limiting the region of DNA transduced among strains. Experimental limitation of the amount of DNA transduced among strains is crucial for use with phage P1 due to the relatively large amount of DNA transduced by this phage. Because genes involved in the regulation and synthesis of cell surface structures are linked to *galE* in serovar Typhimurium LT2 with respect to the transducing distance of phage P1, I designed KAB371 to repair the *galE* mutation without transferring any of these genes.

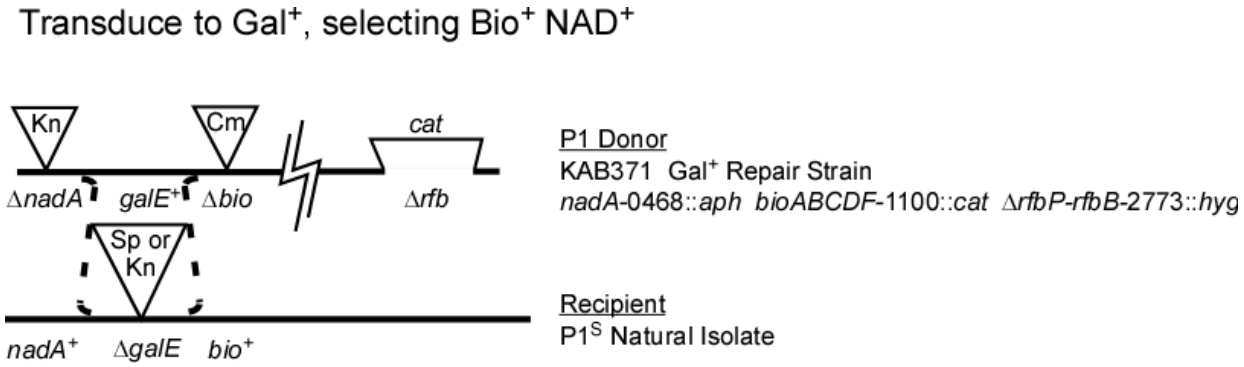

Figure 11. Repair of the *galE* mutation

The donor strain KAB371 contains an intact copy of *galE* flanked by auxotrophic markers *nadA*-0468::*aph* and Δ*bioABCDF*-1100::*cat* and is used to repair strains made sensitive to phage P1 using the *galE* mutation. Transductants are selected for the ability to degrade galactose, indicating repair of the *galE*-6866::*aph* or *galE*-6867::*aadA* mutations; selection for growth in the absence of biotin or nicotinamide indicates limiting of the amount of transduced DNA to only include the region bounded by the Δ*nadA* and Δ*bioABCDF* markers. Dashed lines indicate approximate locations of recombination. Bio^+ = able to grow on minimal media in the absence of biotin; NAD^+ = able to grow on minimal media in the absence of nicotinamide; Gal^+ = able to degrade galactose; $P1^R$ = resistant to phage P1, $P1^S$ = sensitive to phage P1.

2.5 THE *GALE* MUTATION CONFFERS SENSITIVITY TO PHAGE P1 IN NON-TYPHIMURIUM *SALMONELLA*

The results discussed above using serovar Typhimurium LT2 led me to investigate the capacity for the *galE*-6866::*aph* construct to confer sensitivity to phage P1 in non-Typhimurium strains. To examine this, I used λ Red-mediated directed allelic replacement [68, 230, 260] to introduce the *galE*-6866::*aph* construct into 13 non-Typhimurium strains from the *Salmonella* Reference Collection B (SARB) [35] (Table 4). The original *galE*-6866::*aph* construct was designed to target the most conserved portions of the *galE* locus in different *Salmonella*. To identify appropriate sequences to direct the *galE* deletion, the complete genome sequences for 16 strains of *Salmonella* were downloaded from NCBI [237] and the regions surrounding the *gal* operon were aligned. The *galE* gene was well conserved in these genomes with the exception of serovar Arizonae (Figure 12), which appears to have a truncated allele of *galE* (Figure 12A), allowing for the identification of two 40 bp regions that were nearly identical in all 16 strains (Figure 12BC); these regions are contained within the PCR primers used to generate the *galE*-6866::*aph* gene replacement cassette.

Table 4. Non-Typhimurium *Salmonella galE*-6866::*aph* strains made sensitive to bacteriophage P1

SARB Strain	O-antigen Serotype	H-antigen Serotype		*galE*-6866::*aph* Strain
		H1	H2	
1	1,4,12	f,g,s	--	KAB027
2	3,10	e,h	1,6	KAB028
3	1,4,12	l,v	e,n,z_{15}	KAB029
8	6,7	c	1,5	KAB059
20	8,20	g,m,s	--	KAB024
30	6,7	g,m,[p],s	[1,2,7]	KAB025
34	6,8	d	1,2	KAB168
36	6,8	e,h	1,2	KAB054
52	1,9,12	--	--	KAB060
54	11	r	e,n,x	KAB169
55	1,4,[5],12	e,h	1,2	KAB170
59	1,3,19	g,[s],t	--	KAB055
60	1,4,[5],12,27	d	1,2	KAB222

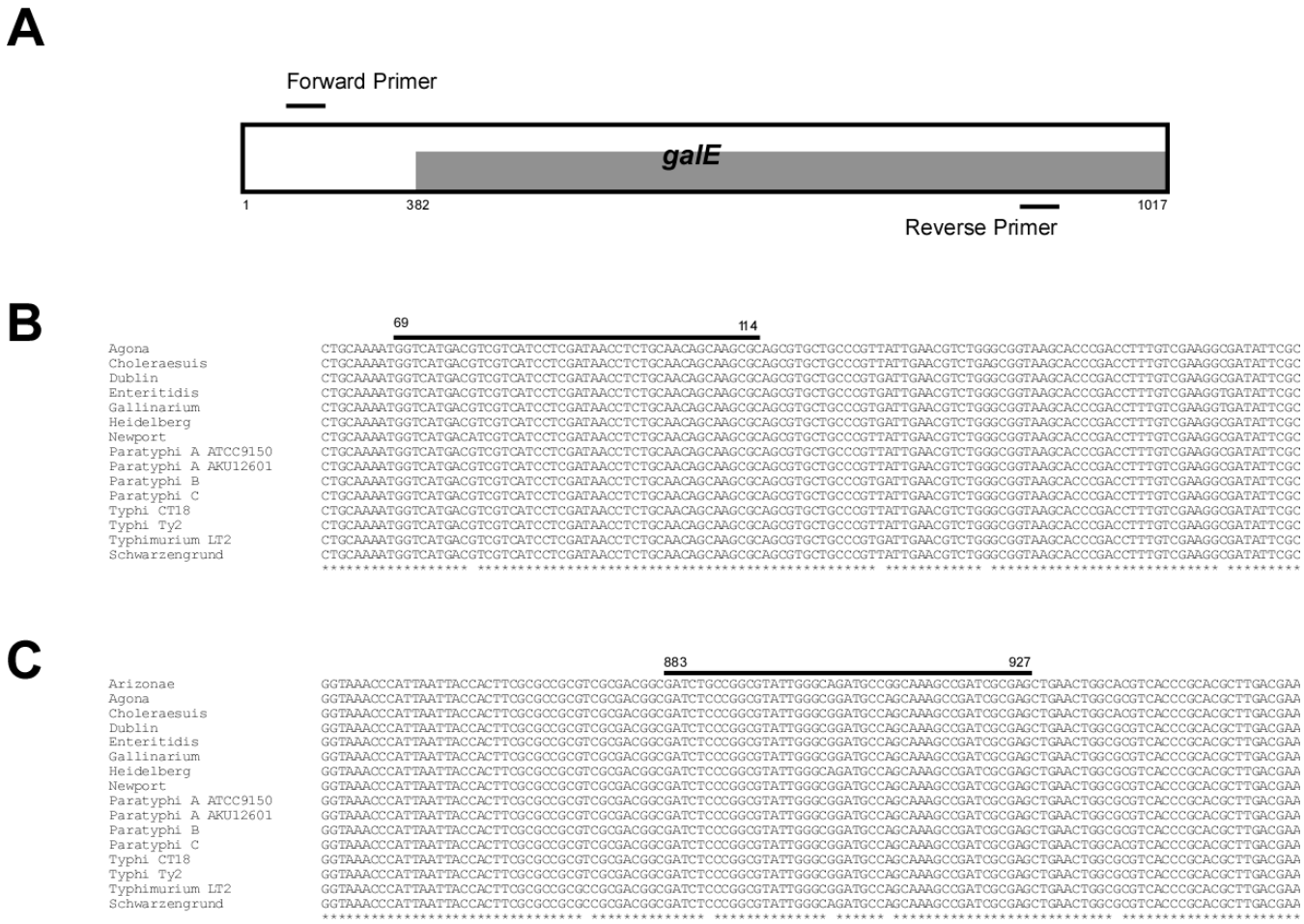

Figure 12. Codon alignment of *galE* from sixteen *Salmonella* genomes and design of primers for directed replacement of *galE*

A. Diagram of *Salmonella galE* gene and location of primers used for directed gene replacement. Gray insert indicates length of *galE* gene from *S. enterica* subspecies *arizonae* compared to other serovars. **B.** Location of forward primer designed to a conserved region of *galE*. This region is missing from *S. enterica* subspecies *arizonae*. **C.** Location of reverse primer designed to a conserved region of *galE*.

The resulting collection of 13 *galE*-6866::*aph* SARB strains (Table 4) were then tested for sensitivity to phage P1 using plaque assays on solid media and for the ability to amplify phage P1 in liquid media. In all cases, the resulting strains displayed varying degrees of sensitivity to phage P1 in both assays.

2.6 PHAGE P1 TRANSDUCES DNA AMONG TYPHIMURIUM AND NON-TYPHIMURIUM *SALMONELLA*

My line of investigation into the role of surface antigens in shaping *Salmonella* fitness against protozoan predation required the experimental control of genes involved in surface antigen regulation and biosynthesis. Thus, I needed a technical approach that permitted the movement of DNA to, from, and among non-Typhimurium strains.

2.6.1 Transduction from Typhimurium LT2 donors into non-Typhimurium recipients

I tested the ability for *galE*-6866::*aph* SARB strains to act as recipients for phage P1-mediated genetic manipulation by transducing them with a variety of markers from serovar Typhimurium LT2. This procedure had the highest likelihood of success since I only required that transducing particles attach to the recipient cell and inject their DNA. Due to my interest in elucidating the role of surface antigens to *Salmonella* fitness against protozoan predation, I chose several markers involved in the biosynthesis of these antigens, including an additional *galE* mutation bearing a different antibiotic resistance marker that permitted additional flexibility in use of antibiotic cassettes in constructing strains with multiple markers (*galE*-6867::*aadA*); genes

involved in the regulation and biosynthesis of the H-antigen (*fljBA*-4400::*tet* and *flhDC*-4820::*cat*), and genes involved in O-antigen biosynthesis (Δ*rfbP-rfbB*-2772::*cat* and Δ*rfbP-rfbB*-2773::*hph*). Because *in vivo* fitness assays require strains to be non-pathogenic, I also chose to manipulate the *Salmonella* Pathogenicity Island 1 (SPI1), a Type III secretion system required for virulence in host intestines [167], with the construct Δ*orgC-prgH*-5800::*cat*.

I successfully transduced these constructs into nearly all of the *galE*-6866::*aph* SARB strains (Table 5), although transduction was occasionally limited due to the natural level of antibiotic resistance and antibiotic tolerance of some strains. In particular, I was limited in manipulating KAB168 (SARB34 *galE*-6866::*aph*) and KAB222 (SARB60 *galE*-6866::*aph*) due to natural spectinomycin resistance in both strains and tetracycline resistance in KAB222. Successful transduction of multiple constructs from serovar Typhimurium LT2 into a diverse collection of SARB strains suggests that the host restriction systems in SARB strains do not significantly impact transfer of LT2-derived DNA. I demonstrate here that genetic constructs can be easily created in the serovar Typhimurium LT2 background using well-established protocols [68, 230, 260] and transferred to non-Typhimurium strains via phage P1 transduction, eliminating the need to perform multiple directed gene knockouts in varying strain backgrounds.

Table 5. Constructs transduced into *galE*-6866::*aph* SARB strains using bacteriophage P1

Y: attempted transduction; successful. N: attempted transduction; unsuccessful. ND: transduction not attempted. *Natural drug resistance of recipient strain prevented transduction of construct.

SARB	Strain	*galE*-6867::*aadA*	*fljBA*-4400::*tet*	*flhDC*-4820::*cat*	Δ*rfbP*-*rfbB*-2772::*cat*	Δ*rfbP*-*rfbB*-2773::*hph*	Δ*orgC*-*prgH*-5800::*cat*
1	KAB027	Y	Y	Y	Y	Y	Y
2	KAB028	Y	Y	Y	Y	Y	ND
3	KAB029	Y	Y	Y	Y	Y	Y
8	KAB059	Y	Y	Y	Y	Y	ND
20	KAB024	Y	Y	Y	Y	Y	Y
30	KAB025	Y	Y	Y	Y	Y	Y
34	KAB168	ND*	ND	Y	Y	ND	ND
36	KAB054	Y	Y	Y	Y	Y	Y
52	KAB060	Y	Y	Y	Y	Y	ND
54	KAB169	Y	Y	Y	Y	Y	Y
55	KAB170	Y	Y	Y	Y	Y	ND
59	KAB055	Y	Y	Y	Y	Y	Y
60	KAB222	ND*	ND	ND	ND	Y	ND

2.6.2 Transduction from non-Typhimurium donors into Typhimurium LT2 recipients

Next, I tested the ability of *galE*-6866::*aph* SARB strains to act as donors for phage P1 transduction using serovar Typhimurium LT2 as a recipient. This is a more difficult test as phage P1 must replicate successfully in natural isolates of *Salmonella*, which may not harbor all necessary host factors required for phage replication. The *rfb* operon was selected as the target locus for transduction due to my interest in genetic manipulation of the loci responsible for production of surface antigens. Moreover, transducing this region requires large incoming DNA fragment to escape cleavage by Typhimurium's restriction endonucleases. The *rfb* operons are hypervariable in *Salmonella*, so the *rfb* directed gene replacements were designed by directing the deletion between the conserved *gnd* and *galF* genes (Figure 5), replacing this region with the chloramphenicol resistance cassette *cat*. I also generated a second *Rfb*$^-$ construct Δ*rfbP-rfbB*-2773::*hph*, replacing the region between the *gnd* and *galF* genes with the hygromycin resistance cassette *hph* for flexibility in creating strains with multiple directed gene replacements used for later experiments.

I created the serovar Typhimurium LT2 recipient strain KAB037 (*galE*-6866::*aph hisD*9953::Mu*d*J Δ*rfbP-rfbB*-2772::*cat*), transducing the region from the *his* operon through the adjacent *rfb* operon from *galE*-6866::*aph* SARB strains into KAB037, selecting for histidine prototrophy. Transductants were screened for chloramphenicol sensitivity, indicating the repair of the Δ*rfbP-rfbB*-2772::*cat* construct with a SARB-derived intact *rfb* locus (Figure 13; Table6). I was unable to transduce the *rfb* region from KAB222 (SARB60 *galE*-6866::*aph*) into KAB037, although KAB222 is sensitive to phage P1 and is able to grow on minimal media lacking histidine. The most likely explanation for the inability to recover transductants using KAB222 as a donor is that the hypervariable nature of the *rfb* locus itself prevented its transfer from SARB60

into serovar Typhimurium LT2. I noted that linkage between the *hisD*9953::Mu*d*J and Δ*rfbP*-*rfbB*-2772::*cat* markers fell into the range of 70% - 90% for most strains, corresponding to a physical distance of 3-10 kb, consistent with the size of the inter-operonic region in complete genome sequences [13, 38, 53, 56, 132, 149, 209, 250, 277, 351, 390]. Therefore, I conclude that restriction endonucleases are not interfering with transduction in this system; if it were, then cotransduction frequencies would have been far smaller than I observed.

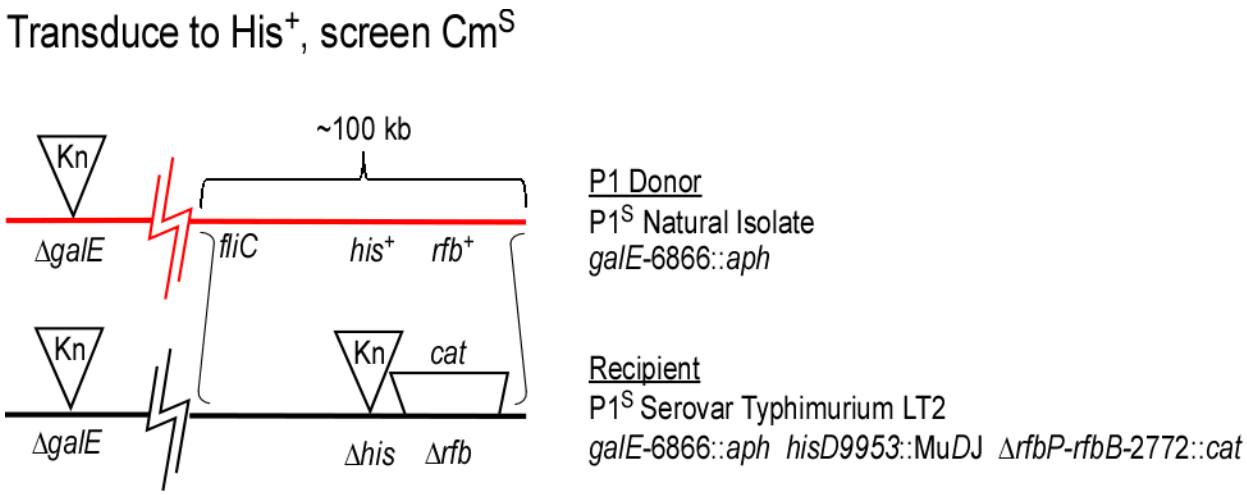

Figure 13. Transfer of the *rfb* locus from a natural isolate of *Salmonella* into serovar Typhimurium LT2

Because phage P1 is able to transduce up to ~100 kb of DNA, the H1-antigen encoding *fliC* gene may be transferred from a natural isolate of *Salmonella* along with the *rfb* locus. Dashed lines indicate approximate locations of recombination. Red lines indicate natural isolate DNA; black lines indicate serovar Typhimurium LT2 DNA. Cm = chloramphenicol; His = histidine; S= sensitive.

Table 6. Serovar Typhimurium LT2 strains containing the *rfb* loci from natural *Salmonella* isolates

SARB Strain	*galE*-6866::*aph* SARB strain	*galE*-6866::*aph* SARB *rfb* LT2 background
1	KAB027	KAB065
2	KAB028	KAB062
3	KAB029	KAB063
8	KAB059	KAB067
20	KAB024	KAB064
30	KAB025	KAB043
34	KAB168	KAB175
36	KAB054	KAB057
52	KAB060	KAB068
54	KAB169	KAB176
55	KAB170	KAB177
59	KAB055	KAB066
60	KAB222	Failed

2.6.3 Transduction among non-Typhimurium donors and recipients

To demonstrate that effective P1 transduction in Salmonella does not require Typhimurium LT2 as either a donor or a recipient, I moved genetic material between natural isolates, here transducing the *rfb* locus between *galE*::*aph* SARB strains. I created SARB *rfb* recipient strains (*galE*-6866::*aph hisD*9953::Mu*d*J Δ*rfbP-rfbB*-2772::*cat*), transducing the region encompassing the *his* and *rfb* operons from *galE*-6866::*aph* SARB strains into the SARB *rfb* recipient strains (Table 7), again selecting for histidine prototrophy and screening for chloramphenicol sensitivity. Limiting the amount of DNA transduced upstream of the *his* operon was not necessary here, as I only sought to demonstrate transfer of DNA among natural isolates of *Salmonella* via phage P1 transduction. The Rfb^+ phenotype was verified with antibody agglutination using cells grown in the presence of galactose. Therefore, I conclude that *galE* mutants of non-Typhimurium serovars of *Salmonella* can be readily manipulated using bacteriophage P1. Critically, it is not impeded by restriction endonucleases and easily moves more than 50 kb of DNA.

Table 7. The *rfb* operon was transduced among non-Typhimurium *Salmonella* via bacteriophage P1

Donor Strain ***galE*::*aph***	**Recipient Strain** ***hisD*9953::Mu*dJ*** **Δ*rfbP-rfbB*::*cat***	**O-serotype conversion**
KAB055 (SARB59)	KAB135 (SARB20)	(8,20) to (1,3,19)
KAB029 (SARB3)	KAB135 (SARB20)	(8,20) to (1,4,12)
KAB028 (SARB2)	KAB137 (SARB30)	(6,7) to (3,10)
KAB059 (SARB8)	KAB139 (SARB59)	(1,3,19) to (6,7)

2.7 DIFFERENTIAL P1 SENSITIVITY IN *GALE* AND *RFB* MUTANTS

Both the *galE* and *rfb* mutant strains possess LPS core antigens that lack the O-antigen, but as discussed above, there is a notable difference between the two in terms of O-antigen biosynthesis and display. The O-antigen is attached to the LPS core, which is in turn mounted onto the lipid A of the outer membrane of the *Salmonella* cell [293, 305]. While *rfb* mutants do not synthesize an O-antigen at all and produce fully intact LPS core antigens [305], *galE* mutants are able to synthesize O-antigens but are unable to present them on the outside of the cell, as they do not complete synthesis of the core antigen itself and thus lack the O-antigen attachment site [305]. As a result, the surfaces of these cells are subtly different: *galE* mutants possess a slightly truncated form of the LPS core antigen while *rfb* mutants possess a fully formed LPS core antigen. I have found that bacteriophage P1 behaves differently when amplified on these two mutants.

Bacteriophage P1 has two different infection phases as determined by expression of alternative tail fiber genes. This is achieved through the *cin* recombinase, which catalyzes random inversion of the genes responsible for expression of the tail fibers, altering the host range of P1 depending on which tail fibers are expressed [126, 141-142]. Thus, the lysates of one phase will contain approximately 0.001% of particles with the alternative tail fiber phase. The low relative titer of the alternative phase results in several rounds of amplification when moving between hosts that are recognized by the alternate phage. This is most commonly characterized with the hosts *Escherichia coli* and *Klebsiella aerogenes*, which represent the two infection phases of phage P1. In practice, a lysate prepared on *E. coli* that is used to amplify phage P1 on *K. aerogenes* only contains an effective concentration of 10^4 PFU/mL of phage P1 that express the suite of tail fibers enabling infection of *K. aerogenes*, and *vice versa*. Because amplification

of phage P1 requires seeding host cells with a starter lysate of phage P1 at 10^7 PFU/mL, multiple rounds of amplification are required to produce phage P1 lysates when moving between alternative hosts such as *E. coli* and *K. aerogenes*; this process takes three successive rounds of amplification to produce a high-titer transducing lysate of phage P1 on its new host.

I noticed that phage P1 amplification specifically on the *rfb* mutant using phage prepared on a *galE* mutant also required several rounds of amplification to produce high titer lysates, consistent with infection with alternative tail fiber phases. These results suggest that one of the P1 tail fiber phases recognizes the outer portion of the LPS core antigen that is present in a *rfb* mutant, but which is absent from the *galE* mutant. If so, then the attachment site would be very broadly distributed, not only among strains of *Salmonella* but among enteric bacteria in general. Moreover, it would not be susceptible to rapid evolutionary change so that phage resistant forms with unrecognizable attachment sites would not arise rapidly.

To examine the difference in P1 infectivity in *galE* and *rfb* mutant strains, I conducted reciprocal lysate amplification assays in serovar Typhimurium LT2 in which the *galE*-6866::*aph* (KAB007) and *galE*-6867::*aadA* (KAB705) constructs were used to seed lysates of *rfb* mutant strains, either Δ*rfbP-rfbB*-2772::*cat* (KAB030) or Δ*rfbP-rfbB*-2773::*hph* (KAB561) and *vice versa*. Similar to my previous observations in lysate production, only lysates made using the *rfb* mutant strains KAB030 and KAB561 were able to clear cultures of each other (Table 8), suggesting that the *rfb* mutant strain could represent a different infection phase of phage P1 as compared to the *galE* mutant strain. Interestingly, *galE* mutant strains KAB007 and KAB705 were equally cleared by lysates produced on both *galE* and *rfb* constructs (Table 8).

Table 8. Time of clearing of *galE* and *rfb* mutant strains of *Salmonella*

Bacteriophage P1 lysates prepared on KAB007 (*galE*-6866::*aph*), KAB705(*galE*-6867::*aadA*), KAB030 (Δ*rfbP-rfbB*-2772::*cat*), and KAB561 (Δ*rfbP-rfbB*-2773::*hph*) were used to seed self- and reciprocal lysates production. Number of + symbols indicate relative degree of clearing of lysate after five hours of incubation with shaking at 37°C.

Host	**Bacteriophage P1 Lysate**			
	KAB007 *galE*-6866::*aph*	**KAB705 *galE*-6867::*aadA***	**KAB030 Δ*rfbP-rfbB*-2772::*cat***	**KAB561 Δ*rfbP-rfbB*-2773::*hph***
KAB007	++++	++++	++++	++++
KAB705	++++	++++	++++	++++
KAB030	+	+	++++	++++
KAB561	+	+	++++	++++

Although there are differences in the behavior of phage P1 when grown on *galE vs. rfb* mutant hosts, I elected to not pursue this issue in further detail for two reasons. First, I observed ambiguities in follow-up experiments similar to the original genetic data suggesting conflicting results for the identity of the receptor for phage P1 in *E. coli* and *Salmonella*, indicating that more sophisticated approaches are required to examine the mechanism of phage P1 adsorption to its hosts. Second, elucidation of the mechanism of phage P1 adsorption to *Salmonella* is beyond the scope of this dissertation. The primary goal of this chapter was to develop a genetic system for the manipulation of non-Typhimurium *Salmonella* to create strains that enable the quantification of the contributions to surface antigens against protozoan predation. I addressed the differential behavior of phage P1 when grown on different mutant *Salmonella* hosts here simply to indicate that, while phase differences do exist, P1 lysates grown on *rfb* strains effectively transduce *galE* mutants, thus allowing repair of these mutants following genetic manipulation (Figure 11).

2.8 DISCUSSION AND FUTURE DIRECTIONS

Here, I described a toolkit for the rapid and reversible use of bacteriophage P1 for genetic manipulation of natural isolates of *Salmonella*. Conserved oligonucleotide sequences allow for high-efficiency introduction of a deletion of either the *galE* gene or *rfb* operon, thus conferring sensitivity to phage P1 with high confidence. Because both the *galE* and *rfb* mutations confer phage P1 sensitivity, transduction may be used to repair the mutation used to confer phage P1 sensitivity. The *galE*-6866*::aph* mutation confers phage P1 sensitivity in a group of SARB strains representing diversity within Subspecies 1 of *Salmonella* [35], indicating its potential for

broad applicability within *Salmonella*. One of the P1 tail fiber phases appears to recognize the LPS core antigen, ensuring widespread distribution of the attachment site. The ease, transparency, widespread distribution of receptor, robustness to variability in restriction endonucleases, and apparent lack of coimmune prophages makes this toolkit valuable to those interested in performing classical genetics in natural isolates of *Salmonella*. Moreover, this strategy will likely be useful in *E. coli*, wherein some strains can be made sensitive to phage P1 in a similar way [129], and other enteric bacteria sharing the LPS core antigen.

This toolkit provides a mechanism for easy genetic manipulation of natural isolates of *Salmonella* that has several advantages over previously described approaches. In one earlier approach, transient heat shock was used to briefly inactivate host restriction enzymes, increasing transduction efficiency in some strains previously unable to be transduced with phage P22 [83]. However, this method does not allow for transduction into serovars outside the Typhimurium subgroup, as they lack the required (1,4,[5],12) serotype for phage P22 infection. In another approach, *E. coli* was made sensitive to phage P22 via the introduction of a conjugative plasmid bearing the Typhimurium *rfb* operon [238], potentially permitting the use of phage P22 in otherwise resistant strains. However, this plasmid is unstable, carries a large number of foreign genes, and is unpredictable in its ability to confer sensitivity in other strains wherein the native and foreign *rfb*-encoded enzymes may interact to create O-antigens unrecognizable by phage P22. Lastly, phage P1 packages more than 100 kb of DNA, more than twice as much as phage P22, making it a good candidate for moving pathogenicity islands and other large regions of DNA that are specific to particular serovars.

3.0 O-ANTIGEN IDENTITY IS A MAJOR INFLUENCE ON *SALMONELLA* FITNESS AGAINST PROTOZOAN PREDATION

I propose that *rfb* diversity in *Salmonella* is maintained by selective pressure from protozoan predation. In this case, I expect the *rfb*-produced O-antigen to have a major role in shaping the feeding preferences of protozoa. With respect to *Salmonella*, changes in O-antigen identity should result in major shifts in relative fitness against predation if avoiding protozoan predation is the primary driver of genetic diversity at the *rfb* locus. The experiments described in this chapter tested the null hypothesis that the O-antigen does not contribute to *Salmonella* fitness against predation, and thus protozoan predation is not a viable explanation for *rfb* diversity.

3.1 IDENTIFICATION OF THE MAJOR SURFACE ANTIGEN AFFECTING *SALMONELLA* FITNESS AGAINST PREDATION

While previous results showed that differences in the O-antigen are sufficient to allow amoebae to discriminate among *Salmonella* prey [387], the major antigens recognized by these predators was unclear. The microbiology of *Salmonella* – including its abundance and constitutive expression - suggests that the O-antigen is the major contributing factor to fitness against predation.

If O-antigen diversity in *Salmonella* is maintained by selective pressure from protozoan predation, then I would expect the *rfb* genes, which are responsible for production of the major monomeric polysaccharide O-antigen subunit [293], to have a major role in shaping *Salmonella* fitness against predation. To address this issue, I created a large series of strains that were either genetically near-identical except for the *rfb* locus (*rfb* near-isogenic) or genetically dissimilar except for the *rfb* locus (*rfb* near-exogenic). I specifically manipulated one antigen in two separate genetic backgrounds, affording a high degree of experimental control not present in previous experiments. These strains enable the performance of experiments to determine if O-antigen identity influences protozoan predation to a sufficiently high degree that a large-scale experiment quantifying the relative contributions of O-antigen and H-antigen identities should be undertaken. If changes in O-antigen identity fail to alter protozoan feeding preferences to a significant degree, then predation is unlikely to play a major role in shaping *rfb* diversity in *Salmonella*.

3.2 MATERIALS AND METHODS

3.2.1 Media and growth conditions

Amoebae were propagated at 33°C on either NM medium (15.5 mM potassium phosphate pH 7.5; 0.2% peptone; 0.2% glucose; 2.0% agar) or NM-LG (low glucose) medium (15.5 mM potassium phosphate pH7.5; 0.2% peptone; 0.02% glucose; 1.5% agar) as reformulated by Wildschutte and Lawrence [386]. Strains were routinely propagated on LB media supplemented with appropriate antibiotic(s) and grown at 37°C. All antibiotics were used at concentrations

described in Section 2.2.1. The minimal media used were: 1X E salts [371] supplemented with 0.2% glucose as a carbon source and 1.2% agar for plates; 1X NCE was prepared as described in Section 2.2.1. SARB strains 1, 3, 8, and 36 are defective for purine biosynthesis and require the addition of 0.008% guanosine to minimal media to support growth.

3.2.2 Propagation and acid-base treatment of amoebae

Amoebae propagation plates were prepared by spreading NM or NM-LG plates with 100 µL of an overnight culture of feeder bacteria and grown overnight at 37°C; *Klebsiella aerogenes* W70 (KAB003), *Escherichia coli* K12 (KAB047), or *Salmonella enterica* serovar Typhimurium LT2 (KAB002) were used as feeder strains. Amoebae were seeded onto the centers of feeder plates and grown at 33°C for 3-4 days; cysts were harvested in 2 mM Tris-Cl pH 7.6 and pelleted via centrifugation at 7500 rpm for 10 minutes. Cysts were exposed to an acid-base treatment to remove feeder bacteria. After elution from plates and pelleting, cysts were resuspended in an equal volume of 10 mM glycine pH 1.5 and incubated for 1 hour at room temperature. Cysts were recovered via centrifugation at 7500 rpm for 10 minutes and rinsed with an equal volume of 50 mM Tris-Cl pH 9.0 followed by an additional rinse with an equal volume of 50 mM Tris-Cl pH 7.0. Cysts were again recovered via centrifugation at 7500 rpm for 10 minutes, resuspended in an equal volume of 10 mM K_2HPO_4 pH 12.5, and incubated for 1 hour at room temperature. Cysts were rinsed as described above, except the order of the Tris-Cl rinse buffers were reversed. After final rinsing, cysts were pelleted and resuspended in 2 mM Tris-Cl pH 7.6 and enumerated with direct counting on a hemocytometer. Working stocks of cysts were kept at a concentration of 10^6 cysts/mL and stored at room temperature.

3.2.3 Construction of *rfb* near-isogenic and near-exogenic strains

Strains having SARB-derived *rfb* loci in the LT2 chromosomal background were engineered as described in Chapter 2.6.2 (Figure 13) and listed in Table 6. The *fliC* gene encoding the hypervariable portion of the H-antigen is located within ~100 kb of the *rfb* locus [209], which can be co-transduced with *rfb* using phage P1. To limit the amount of SARB DNA transduced into serovar Typhimurium LT2, I twice backcrossed the SARB *rfb* locus into additional serovar Typhimurium LT2 recipient strains (Figure 14). First, the SARB-derived *rfb* regions from KAB strains (Table 6) were transduced into serovar Typhimurium LT2 strain KAB046 (*galE*-6866::*aph phs*-209::Tn*10d*Gn *hisD*9953::Mu*d*J Δ*rfbP-rfbB*-2772::*cat*); transductants were selected for histidine prototrophy and were screened for sensitivity to chloramphenicol, and resistance to gentamycin, indicating replacement of Δ*rfbP-rfbB*-2772::*cat* while retaining LT2 DNA upstream of the *phs*-209::Tn*10d*Gn marker (Figure 14A; Table 9). The second backcross involved the transduction of the SARB-derived *rfb* regions from KAB strains (Table 9) into the recipient serovar Typhimurium LT2 strain KAB030 (*hisD*9953::*Mu*dJ Δ*rfbP-rfbB*-2772::*cat*); transductants were again selected for histidine prototrophy and screened for sensitivity to chloramphenicol and to gentamycin, demonstrating the transfer of the SARB-derived *rfb* region without encompassing the region upstream of and including the *phs*-209::Tn*10d*Gn marker (Figure 14B; Table 9). KAB030 contains an intact copy of *galE* and relies strictly on Δ*rfbP-rfbB*-2772::*cat* to confer sensitivity to phage P1, so a separate *galE* repair is unnecessary when using this strain background. *Rfb*$^+$ phenotypes of resulting strains were verified with antibody agglutination assays, indicating the expression of SARB-derived O-antigens in the Typhimurium LT2 genetic background.

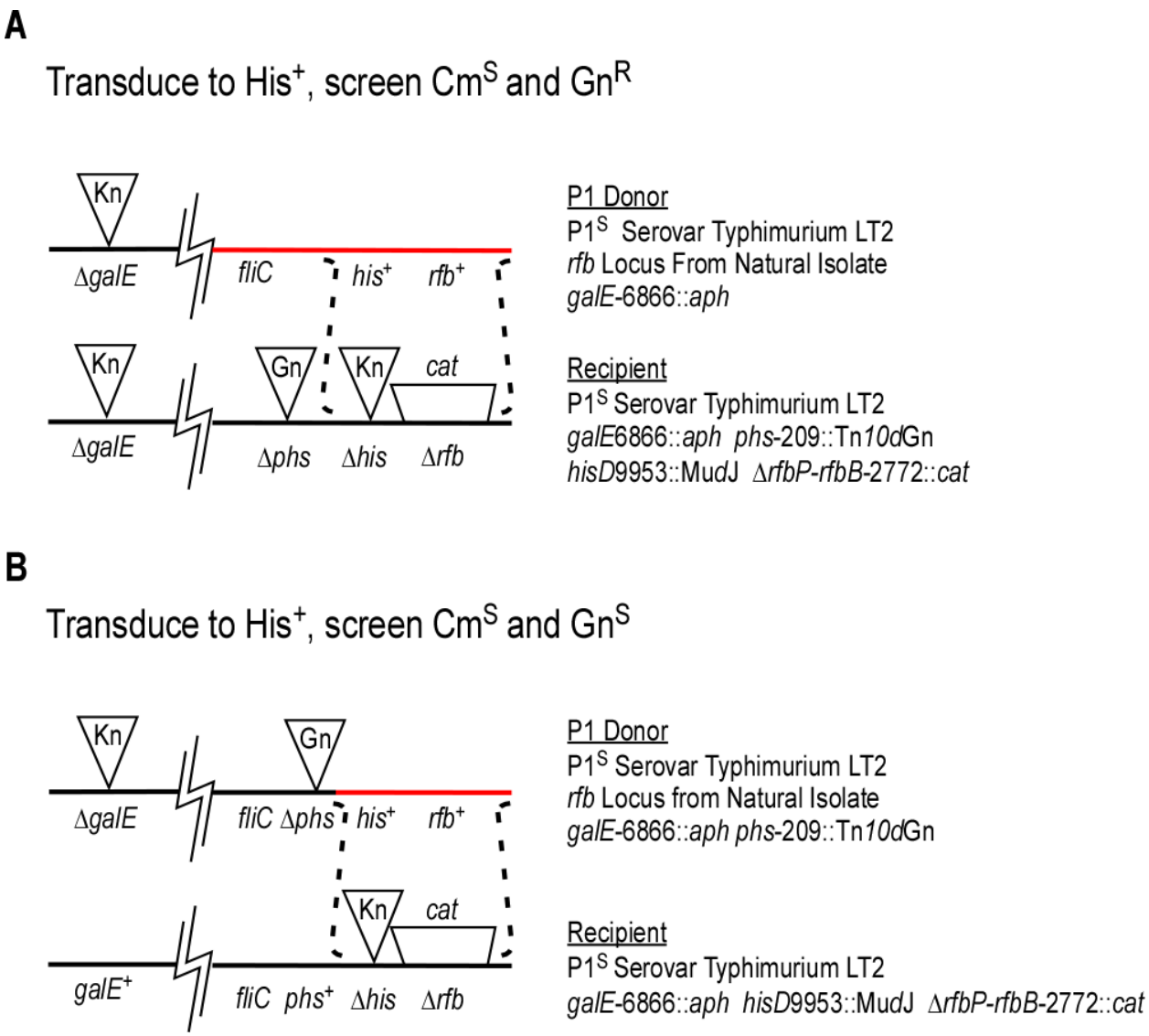

Figure 14. Construction of *rfb* near-isogenic strains

A. Limitation of the amount of DNA transduced by bacteriophage P1 through the use of a selective marker. The *rfb* region from the donor strain was transferred via phage P1 transduction into a recipient strain containing the *phs*-209::Tn*10d*Gn marker; retention of this marker limits transduced DNA downstream of *phs*. **B**. Final phage P1 transduction of the region limited to the area near and including the *rfb* locus into a *galE*$^+$ strain. Red lines indicate SARB-derived DNA; black lines indicate DNA from serovar Typhimurium LT2.

Table 9. Strains made near-isogenic at *rfb* to assess the contribution of O-antigen identity to *Salmonella* fitness against protozoan predation

Donor strain A: SARB strain containing the *galE*-6866::*aph* construct. Donor strain B: LT2 strain containing a SARB-derived *rfb* locus and the *galE*-6866::*aph* construct. Donor strain C: LT2 strain containing a SARB-derived *rfb* locus, the *phs*-209::Tn*10d*Gn construct to limit the amount of DNA cotransduced with *rfb* and the *galE*-6866::*aph* construct. ND = not determined.

SARB#	Donor Strain A	Donor Strain B	Donor Strain C	*rfb* Near-Isogenic Strain LT2 Background
1	KAB027	KAB065	KAB070	KAB080
2	KAB028	KAB062	KAB071	KAB081
3	KAB029	KAB063	KAB072	KAB082
8	KAB059	KAB067	KAB073	KAB083
20	KAB024	KAB064	KAB074	KAB084
30	KAB025	KAB043	KAB075	Failed
34	KAB168	KAB175	KAB183	KAB206
36	KAB054	KAB057	KAB076	KAB085
52	KAB060	KAB068	KAB077	KAB086
54	KAB169	KAB176	KAB184	KAB207
55	KAB170	KAB177	KAB185	KAB208
59	KAB055	KAB066	KAB078	KAB087
60	KAB222	Failed	ND	ND

To create *rfb* near-exogenic strains, the *hisD*9953::Mu*d*J and Δ*rfbP-rfbB*-2772::*cat* constructs were transduced from KAB037 (*galE*-6866::*aph hisD*9953::Mu*d*J Δ*rfbP-rfbB*-2772::*cat*) into SARB *galE*-6866*::aph* strains (Table 10) using phage P1; transductants were selected for chloramphenicol resistance, demonstrating the transfer of the Δ*rfbP-rfbB*-2772::*cat* construct, and screened for histidine auxotrophy, indicating transfer of *hisD*9953::Mu*d*J (resulting strains in Table 10). The *rfb* region from serovar Typhimurium LT2, obtained from KAB045 (*galE*-6866::*aph phs*-209::Tn*10d*Gn), was transduced into the strains listed in Table X; transductants were selected for histidine prototrophy, signifying repair of *hisD*9953::Mu*d*J, and screened for sensitivity to chloramphenicol and gentamycin, indicating transfer of the LT2 *rfb* operon and limiting transduced DNA to the region downstream of *phs*-209::Tn*10d*Gn. *Rfb*$^+$ phenotypes were verified by serotype agglutination assays.

Table 10. Strains made near-exogenic at *rfb* used to assess the contribution of O-antigen identity to *Salmonella* fitness against protozoan predation

ND = not determined.

SARB	*rfb* Near-Exogenic Strain LT2 *rfb* Locus in SARB Strain Background
1	ND
2	KAB201
3	KAB210
8	KAB216
20	KAB202
34	KAB209
36	KAB211
52	ND
54	ND
55	KAB205
59	KAB204

3.2.4 Line test competition assay and fitness calculations

Line tests were modified from the original protocol of Wildschutte and Lawrence [387]. Strains were struck in lines approximately 30 mm in length on square 245mm x 245 mm plates of NM-LG. To prevent cross-feeding of predators, lines were placed 30 mm apart and plates were partitioned into three equal sections with sterile wooden barriers embedded ~5 mm below the surface of the agar and extending ~5 mm above the surface of the agar. Each plate contained 3-5 replicates of each strain; a total of three plates were tested for each set of strains per predator. Plates were incubated overnight at 37 A total of 10^4 protozoan cysts in 10 µl 0.9% NaCl were added to the ends of lines; plates were incubated at 34°C. Plates were photographed every six hours and the distance of each line consumed by predators was measured. Rates of predation were determined using the distance of the feeding front of predation relative to the starting point for each line. Rates of predation were determined in mm/hr by comparing the distance of the feeding front of predation relative to the starting point for each line. Regressions were calculated for distance of each line consumed by predators *vs*. time (R^2 typically > 0.95). Overall consumption rates were calculated as mean slopes for each strain on each individual plate and then averaged for that of all plates. Fitness was normalized to the value of the least-preferred strain.

3.3 O-ANTIGEN IDENTITY INFLUENCES FITNESS AGAINST PREDATION

Previous experiments using strains made near-isogenic at *rfb* suggested that the identity of the O-antigen can confer fitness against predation; however the collection of strains used to test this

hypothesis was limited to only three strains representing two similar and one dissimilar O-antigen serotype [387]. A more comprehensive analysis of the role of O-antigen identity in *Salmonella* fitness against predation was hindered due to the lack of genetic manipulation techniques for non-Typhimurium strains. I expanded upon this experiment by using the bacteriophage P1-mediated genetic manipulation technique described in Chapter 2 to construct a substantially larger set of near-isogenic *rfb* strains that only differ in terms of O-antigen identity. To this end, I created a collection of eleven *rfb* near-isogenic strains representing nine distinct O-antigen serotypes representative of diversity within *S. enterica* Subspecies I (Table 9). This comprehensive strain collection allowed me to more robustly refute the null hypothesis that O-antigen identity does not influence fitness against predation.

If O-antigen identity is a minor contributor to fitness against predation, then the fitness of *rfb* near-isogenic strains should be very similar regardless of the predator being faced. To test this prediction, I assessed the fitness of these *rfb* near-isogenic strains against predation by three genetically distinct amoebae: *Naegleria gruberi* NL, *Acanthamoeba* sp. R2-1, and *Tetramitus* sp. BD1-1 [386]. The *rfb* near-isogenic strains do not have similar relative fitness against predation when challenged by any of the three amoebae (Figure 15). Fitness values ranged from 1.0 to 0.55, indicating that the identity of the O-antigen alone increased risk to predation by two-fold. Given the large bacterial population sizes within the intestinal lumen, this difference in predation susceptibility is sufficient to drive differential survivorship of different serovars.

Moreover, if the O-antigen did not significantly contribute to fitness, I would expect that the fitness values of all near-isogenic strains against all predators to have a very high degree of correlation, with R^2 values of strain fitness between predators being very close to a value of 1. However, this is not the case. While I observed a moderate correlation of fitness against

predation between *Tetramitus* sp. BD1-1 and *N. gruberi* NL, R^2 = 0.5413, the R^2 values for fitness of all isogenic strains between two predator groups are actually negative: *Acanthamoeba* sp. R2-1 and *Tetramitus* sp. BD1-1 R^2 = -0.3850 and *Acanthamoeba* sp. R2-1 and *N. gruberi* NL R^2 = -0.0631. Because predators can discriminate among strains that only vary at the O-antigen and the fitness against predation for the *rfb* near-isogenic strains varied considerably among predators, I conclude that the O-antigen must be a major determinant of protozoan feeding preference.

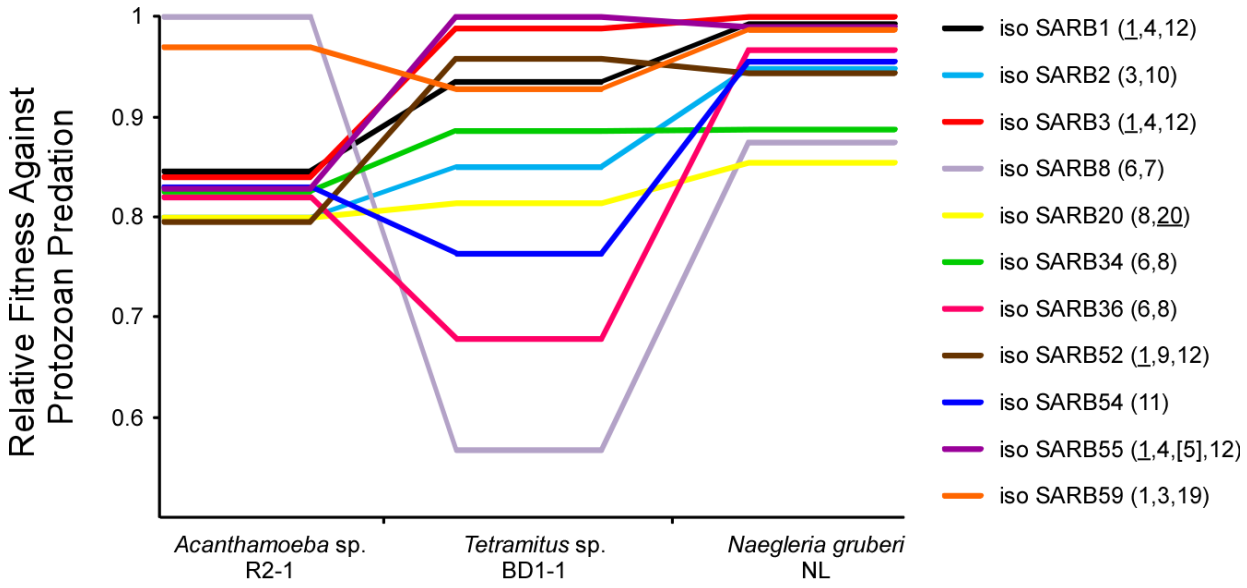

Figure 15. Protozoan predators can discriminate among *Salmonella* strains that only differ at the *rfb* locus

Relative fitness values for strains made near-isogenic at *rfb* against predation by three different amoebae. Relative fitness values were calculated by determining the rate of predation for multiple lines of each strain, correcting for strain growth rate and normalizing to the slowest-consumed (most fit) strain. The strain having the highest fitness against predation was assigned a value of 1; the fitness values of all other strains were normalized to this value. O-antigen serotypes are listed in parentheses.

3.4 DISRUPTION OF O-ANTIGEN IDENTITY ALTERS *SALMONELLA* FITNESS AGAINST PREDATION

The data presented above show that strains differing only at the O-antigen have different relative fitness against predation. Therefore, I predict that disruption of native O-antigen identity should alter fitness against predation. I tested the null hypothesis that disruption of O-antigen identity will not significantly change *Salmonella* fitness against predation by engineering a series of *rfb* near-exogenic strains in which the native *rfb* locus in SARB strains was disrupted via replacement with serovar Typhimurium LT2 *rfb* locus (Table 10). Each *rfb* near-exogenic strain and its wild-type parent SARB strain share all antigens except for the O-antigen. If O-antigen identity does not contribute significantly to fitness against predation, then I expect a high degree of correlation (R^2) between the fitness of each wild-type strain and its *rfb* near-exogenic counterpart. Conversely, if the O-antigen does contribute to fitness against predation, then I expect a low correlation (R^2) between fitness values of wild-type strains and their *rfb* near-exogenic derivatives.

I tested the fitness of wild-type SARB strains and their *rfb* near-exogenic counterparts (Table 10) against predation from *N. gruberi* NL, *Acanthamoeba* sp. R2-1, and *Tetramitus* sp. BD1-1. I observed a very low degree of correlation (R^2 = 0.04, 0.52, 0.15, respectively) between the fitness of wild-type SARB strains and their *rfb* exogenic counterparts against predation by all three amoebae (Figure 16). This indicates that O-antigen identity, and consequently the identity of the *rfb* locus, likely plays a major role in shaping *Salmonella* fitness against protozoan predation. A caveat to this conclusions is that the common (1,4,[5],12) O-antigen introduced from serovar Typhimurium may be differentially modified in these strains, even though it is not

native to that strain background. Therefore, a more rigorous set of experiments is required to estimate the relative contribution of O-, H-, and other antigens to fitness against predation.

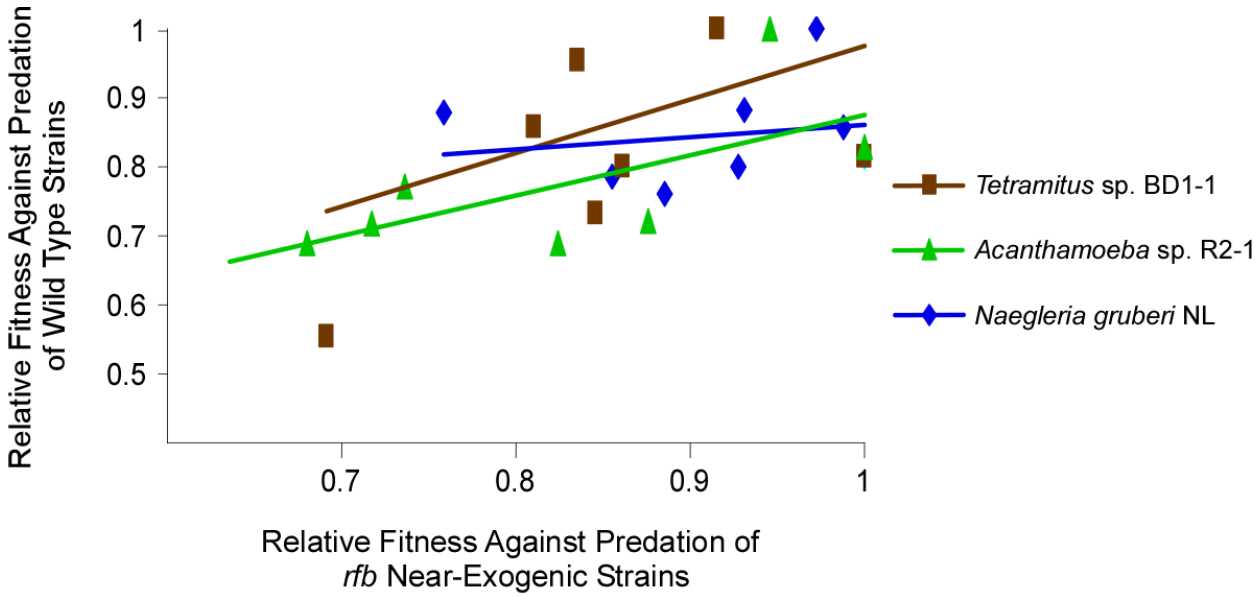

Figure 16. Disruption of the *rfb* locus alters *Salmonella* fitness against protozoan predation

Relative fitness of *rfb* near-exogenic strains *vs.* wild-type SARB strains. Relative fitness values are calculated by determining the rate of predation for multiple lines of each strain, normalizing to the slowest-consumed strain (most fit), and correction for strain growth rate. R^2 values for wild type fitness compared to *rfb* near exogenic fitness are: *Tetramitus* sp. BD1-1 $R^2 = 0.146$; *Acanthamoeba* sp. R2-1 $R^2 = 0.5215$; *Naegleria gruberi* NL $R^2 = 0.0358$.

3.5 O-ANTIGEN IDENTITY IS A MAJOR DETERMINANT OF FITNESS AGAINST PROTOZOAN PREDATION

The experiments described in this chapter establish O-antigen identity as a significant contributing factor to fitness against predation. The next step is to quantify the relative contributions of the O, H- and other surface antigens. Different protozoan predators may use the O-antigen to varying degrees in discrimination among prey [386], so a more comprehensive picture of how various protozoan predators use surface antigen identity to discriminate among prey must be determined. However, a more sophisticated approach than the experiments outlined in this chapter is required to address this question.

For many reasons, the use of line tests is insufficient to provide the robust data needed to quantify the roles of surface antigens to fitness against predation. First, the line test method is crude, relying on optical measurement of the disappearance of lines of bacteria as a result of protozoan predation. Measurements are taken in millimeter increments, impeding the accuracy of this experimental approach, as very small observer error could result in an actual difference of a few million bacterial cells. Second, plate-to-plate variation can be quite high in line tests, potentially clouding subtle differences in fitness among strains. Although the larger plate design increased the number of strains tested per plate and thus was intended to decrease plate-to-plate variation, the new design did not result in a significant improvement in plate-to-plate variation and required more labor in the experimental process as compared to the original line test design. Third, line tests are inefficient compared to the number of data points generated with single lines of bacteria measured in 6-hour intervals. Unreasonably large numbers of plates need to be assessed with line tests in order to gather more robust data sets. Last, line tests are only indirect measures of competitive ability as fitness is measured in isolation; a better test would compete

strains having different antigenic profiles against each other directly. Line tests simply lack the resolution power necessary to quantify the relative fitness contributions of different surface antigens.

Additionally, the genetic approaches used in this chapter do not permit the level of experimental control required for quantification of the contributions of the O- and H-antigens to fitness against predation. While the *rfb* near-isogenic and near-exogenic strains allowed me to conclude that O-antigen identity is a major factor influencing *Salmonella* fitness against predation, they are too crude to provide exact quantifications of this contribution. Simple transfer of the *rfb* locus is insufficient to accurately phenocopy parental O-antigens in different genetic backgrounds. Genes unlinked to the *rfb* locus are known to alter the structure of the O-antigen, as in the case with the O-antigen acetylation gene *oaf* [328] and the chain length regulator *fepE* in serovar Typhimurium LT2 [231]. It is impossible to definitively identify all potential loci unlinked to *rfb* that may play roles in O-antigen modification or chain length regulation among strains used in these experiments, as most of their genomes are not fully sequenced and the structures of all O-antigens are not solved. .

I took a two-pronged approach to facilitate the identification of the major surface antigen impacting fitness against predation and quantify its role in shaping fitness with respect to other surface antigens. First, I moved away from line tests and used a new multicolor fluorescent strain tagging system to adapt the technique of flow cytometry, a more robust assay capable of generating millions of data points from a single competition plate, to assess fitness against predation as described in Chapter 4. Second, I used the P1-mediated genetic manipulation approach I described in Chapter 2 to engineer collections of SARB strains in which the presence and absence of the O- and H-antigens is experimentally controlled, avoiding the confounding

issue of accurate phenocopying of antigens in different genetic backgrounds as encountered with the *rfb* near-exogenic and near-isogenic strains identified here. The results of this new experimental approach are discussed in Chapters 5 and 6.

4.0 NOVEL APPROACHES TO MULTICOLOR FLOW CYTOMETRY

Previous work in the Lawrence laboratory established predation as a possible mediator of diversifying on the *Salmonella* O-antigen-encoding *rfb* operon [386-387], and the experiments presented in Chapter 3 suggest that protozoan predation is influenced by the identity of the O-antigen. To rigorously test this hypothesis, I need to evaluate the relative fitness of strains differing in antigenic composition when facing protozoan predators. Fitness coefficients are measured with high accuracy for bacterial populations grown in liquid culture [72-73, 79, 220, 344]. Here, strains are placed in direct competition and changes in their relative numbers are measured over time. Selection coefficients are measured from the change in frequency over time as:

$$dP/dt = sP(1-P)$$

where P is the proportion of the cells of interest in the population.

It is not feasible to measure relative fitness of different *Salmonella* strains against predators in this fashion. First, amoeba predators consume prey by phagocytosis only when grown on solid media; in liquid culture, those amoebae which feed do so by pinocytosis [34]. Second, there is spatial structuring to solid media that prevents random sampling over time. Third, one cannot compete bacterial strains both with and without predation in the same environment.

In order to validate the hypothesis that protozoan predation shapes *Salmonella rfb* diversity, I must demonstrate that the O-antigen is the major contributing factor influencing

fitness against predation. Moreover, I must also establish that differential survivorship against predation is influenced only by the identity of the O-antigen in natural environments beyond the Petri dish, such as the intestine. While the experiments to more thoroughly test this hypothesis are clear, the techniques and tools required to answer these questions did not exist prior to beginning this course of work. In Chapter 2, I discuss development of the genetic techniques that enable construction of non-Typhimurium *Salmonella* in which the presence and absence of the O- and H-antigens are experimentally controlled. These strains will permit the quantification of the O- and H-antigens to *Salmonella* survival against protozoan predation. Although the genetic techniques I developed and described in Chapter 2 enabled the ability to construct strains designed to assess the contributions of surface antigens to *Salmonella* fitness against predation, I did not have sufficient experimental techniques to actually use these strains in competition experiments. Moreover, I did not have a suitable approach to test the fitness of *Salmonella* strains *in vivo*. Competition tests designed to establish the major surface antigen affecting fitness against predation and comprehensively examine the contribution of the identity of this antigen to *Salmonella* fitness against protozoan predation *in vivo* requires the following aspects:

- ability to discriminate simultaneously enough strains to reveal subtle differences in fitness;
- detection of cells isolated from complex environments;
- minimization of variation through quantitative, direct fitness assessment;
- high-throughput collection of large data sets.

Previous work by others in the Lawrence laboratory developed a surrogate method for measuring selection coefficients involving whole plate competition assays [387]. Here, a mixture of cells was grown either in the presence or absence of predators. The relative abundance of each

cell type was assessed on the plate lacking predator, providing an expectation of cell frequency on predator-bearing plates. Deviations of observed cell frequencies from these expected values are direct measures of strain fitness (w_s), where

$$w_s = P_{observed}/P_{expected}$$

Thus, accurate measured of fitness can be acquired if cell frequencies can be determined with accuracy; robust frequency data is obtained only from large sample sizes, especially when relative cell frequencies are low. Pairwise plate competition tests were also used to support the results of line tests in a more quantitative manner [387]. Here, two strains bearing different metabolic or antibiotic resistance phenotypes were co-plated and their fitness values in the presence and absence of predation was determined relative to each other. Because only two strains are competed simultaneously, this method will not reveal any subtle fitness differences among strains. To reveal subtle variation in strain fitness, a minimum of six to eight strains must be competed at the same time; previous approaches do not permit this level of experimental sophistication. To obtain such large sample sizes, I turned to flow cytometry to identify cells eluted from assay plates.

I believe that the technique of flow cytometry fulfills all of the requirements I need in order to fully address the contributions of the O- and H-antigens to fitness against protozoan predation and assess bacterial survivorship in *in vivo* and in other biologically complex environments. Flow cytometers permit the high-resolution enumeration and analysis of small particles, from bacterial spores to whole eukaryotic cells, based on a variety of light-scattering and fluorescent properties, native or experimentally manipulated, of the particles themselves. Using flow cytometry, these properties are recorded for each individual cell passing through the field of detection.

Flow cytometry gathers information based on the fluorescent signals of cells; a wide variety of commercially available fluorescent proteins and dyes are commonly used to tag bacterial and eukaryotic cells for cytometric analysis [66, 70, 246]. These fluorescent tags are easily introduced into cells and should not alter surface antigenic profiles because they do not affect cellular metabolism. The large variety of fluorescent labels available and the ability of cytometers to detect multiple signals from these labels should permit the simultaneous competition of a group of strains large enough to reveal subtle differences in fitness among strains and representative of diversity within Subspecies I of *Salmonella enterica*. Additionally, fluorescent signals are not expected to be present at the same levels as antibiotic resistance in bacteria from natural environments; the large variety of fluorescent tags enables minimization or avoidance of confounding autofluorescence present in complex biological environments through careful experimental design. Flow cytometry affords a high degree of experimental flexibility for the simultaneous discrimination of large numbers of strains.

Flow cytometers directly measure the properties of each single cells passing through the field of detection, so determination of fitness using flow cytometry meets my requirement for direct enumeration of cells from competition tests. Moreover, acquisition of this data is culture-independent: I can directly count cells from mixed competition plates on a cytometer without the need to perform additional culturing steps as was required with pairwise plate competition tests as performed previously [387]. Not only does this approach save time, labor and materials, but I should be able to directly quantify numbers of strains present on different competition plates. Direct counting of *Salmonella* cells yields quantitative data, enabling a more rigorous test of the hypothesis that *rfb* diversity in *Salmonella* is shaped by protozoan predation.

Finally, flow cytometers enable the high-throughput acquisition of data in very reasonable periods of time. Most modern flow cytometers are able to collect data at a rate in the thousands of events per second, making possible the counting of millions of bacterial cells in just a few minutes of time. This scale of data throughput is impossible using culture-dependent methods. The data collection capacity of flow cytometers dramatically expand the statistical power of results, providing the level of detail required to quantify the roles of surface antigens to *Salmonella* fitness against predation.

Although flow cytometry is a well-established technique for the analysis of single cells, this approach is not without significant technical challenges. Flow cytometry is most commonly used for the analysis of eukaryotic cells; microorganisms are much smaller and more difficult to detect. Use of multiple fluorescent labels is also quite frequent in flow cytometry, but the level of experimental complexity permitted is limited not only by the availability of lasers and optical filters to maximize detection and separation of fluorescent signals but also by limitations in data analysis. Difficulty of data analysis increases significantly with each additional fluorescent tag used in a given experiment. Furthermore, my research requires a more quantitative approach to data analysis than what is typically necessary for most common cytometry assays.

Here, I describe the development of a multicolor flow cytometry protocol and data analysis tools for examination of bacteria in complex biological environments. This experimental system was designed in order to permit the assessment of bacterial survival in any natural environment from pond water to intestines. Development of this technique permitted all of the experiments required for assessment of the contributions of surface antigens to *Salmonella* fitness against protozoan predation as outlined in Chapters 5 and 6 as well as future experiments proposed in Chapter 7.

4.1 USE OF FLOW CYTOMETRY IN MICROBIOLOGY

In addition to its widespread use in the field of eukaryotic cell biology, fluorescence-based flow cytometry is commonly used to address many issues within research and clinical microbiology. I discuss here the basic principles of flow cytometry, common microbiological applications for flow cytometry, and the current challenges in this field.

4.1.1 Basics of flow cytometry

While flow cytometry has been used in microbiological assays for some time, existing methods are insufficient to allow the degree of discrimination I require to measure relative fitness. There were several challenges to overcome which I describe in this chapter. I first provide an overview of flow cytometry to place these challenges in context.

The technique of flow cytometry involves the measurement of the light-scattering and fluorescence properties of single particles, typically cells or cell-sized particles, suspended in a fluid; for a comprehensive review see Shapiro [316]. A flow cytometer focuses this fluid hydrodynamically using a sheath of buffer, producing a stream of single cells that are passed across a light source. Most modern flow cytometers use one or more lasers as sources of specific wavelengths of light. At the point of intersection with the laser beam, each particle creates a scattering of the light; lasers also excite any fluorescent molecules present in the particle, either intrinsic or experimentally introduced, resulting in emission of signal according to the spectral properties of each particle or fluorochrome. The scattered and emitted photons from these particles are passed through various optical filters to split the light into several paths, filter by chosen wavelengths, and channel the resulting photons into the appropriate optical detectors.

Optical detectors then convert these light signals into electrical pulses, which undergo a series of transformations that are sent to software used to visually depict and process the cytometric data.

Critically, flow cytometers measure these optical properties as a function of each individual particle that passes through the beam of light. The light-scattering properties of such events are often used to physically characterize populations of cells. These properties include forward scatter (FSC) and side scatter (SSC). FSC is the scattering of light by an event in line with laser beam; FSC is roughly proportional to the size of an event and is often used to approximate cell size [226-227, 364]. As there are many other factors other than cell size that can influence this light scattering property, care must be taken when interpreting FSC values and approximating them to cell size [163, 314, 320-321]. SSC is the scattering of light by a particle perpendicular to the laser beam and reflects the complexity or "granularity" of the event [316]. Cell physiological properties such as membrane composition and organelle content influence SSC, and thus this parameter is often used to characterize different populations of cells in a sample [263, 292, 301]. A third event property, pulse width (PW), is often used to further characterize events although it is not a light scattering property; rather, PW reflects the "time of flight" of a particle, or the duration of the electric pulse produced by a particle passing through the laser beam [316]. PW is often used to discriminate between single particle events and multiple particle events; two cells stuck together produce longer PW values than single cells. Because many factors can influence these parameters, it is crucial that combinations of cytometric parameters are used to separate genuine cell events from debris and other unwanted populations of cells [316].

Flow cytometers also record the fluorescence properties of events through the use of one or more lasers that excite particular fluorescent dyes or proteins. Emitted signals are passed

through optical filters that separate multiple fluorescent signals and direct them to the appropriate detector among a fixed set of detectors on the cytometer; this is especially critical for experiments characterizing cells using two or more fluorescent signals in addition to light scattering properties. For example, proper optical filters channel fluorescent emissions from a cell labeled with green and red light-emitting fluorophores into the appropriate detector; "green" signals will only be sent to the "green" detector and not the "red" detector, and *vice versa*. A wide variety of dyes are available for measuring various properties of cellular physiology, such as cell viability, nucleic acid content, ion concentrations and presence of specific organelles [66]. Fluorescent antibodies are often used to label specific cell types in mixed populations; these are widely commercially available. While dyes and antibodies are used for exogenous labeling, endogenously expressed fluorescent proteins can be used for whole-cell labeling, highlighting specific cellular structures with the use of protein tags or as gene expression reporters [70, 246]. The use of fluorescent labels in multicolor experiments is mechanically limited by the wavelengths of lasers used to excite fluorophores, number of signal detectors, use of suitable optical filters, and overlap of emission spectra of fluorophores.

Modern flow cytometers enable high-throughput data collection; it is reasonable to conduct cytometry experiments that collect millions of events from a single sample in just a few minutes' time. From a microbiological perspective, this eliminates the need for culture-dependent analysis and permits acquisition of data at a scale impossible to perform with microscopy or colony counting. Assessing the fitness of multiple strains against protozoan predation at the same time requires a high degree of statistical power, which is addressed through gathering considerable amounts of data. The data collection capacity of flow cytometry makes it ideal for use for *in vivo* competition tests, as detection of experimentally introduced *Salmonella*

requires the acquisition of large numbers of events due to the confounding presence of millions of other native bacterial cells in an intestinal sample.

4.1.2 Common microbiological applications for flow cytometry

The capacity for the rapid acquisition of information from single cells makes flow cytometry ideal for the study of microorganisms, and the analytical, culture-independent power of flow cytometry permits examination of microbes outside of the artificial laboratory bench environment. For example, Sørensen and colleagues [234, 335] used green fluorescent protein (GFP) as a reporter to assess horizontal transfer of a conjugative plasmid among bacteria in the soil. Flow cytometry of GFP-labeled *Salmonella typhimurium* was also used to examine the propensity for this bacterium to evade traditional disinfection procedures used in the preparation of leafy vegetables through invasion into plant tissues [98]. Survival of *Escherichia coli* in wastewater effluent and anaerobic groundwater was determined using flow cytometry [15], and additional work demonstrated that cytometric analysis of fluorescent-tagged *E. coli* was a more sensitive indicator than culture-dependent methods of survival in aquatic environments [190]. Additionally, fluorescent reporters and flow cytometry are currently being applied to diagnostic microbiology; for example, Piuri *et al* [258] and Rondon *et al* [286] engineered a fluoromycobacteriophage-based assay to examine antibiotic resistance phenotypes of *Mycobacterium tuberculosis*. Importantly, the experiments discussed above rely mostly on the detection of one or two colors, which is discussed in further detail in Chapter 4.1.3.

The simultaneous use of more than two fluorochromes in microbial cytometry is extremely limited [114] compared to the commonality of multicolor cytometry experiments using eukaryotic cells in the laboratory environment [20, 119, 269]. For eukaryotic cells, well-

developed protocols exist for the simultaneous discrimination of three [20] and four colors [119], and protocols using eleven [71] and seventeen [254] color tagging schemes have been proposed. Although multicolor flow cytometry is used with microorganisms, these experiments are typically performed under laboratory conditions and are limited to physiological measurements [315, 317, 396] or require the use of highly specific, exogenously added probes [110-111, 307]. Due to experimental complexity, a trade-off exists between examining microbes in natural environments and use of multiple fluorescent detection parameters. For my experiments examining the role of the O-antigen to *Salmonella* survival against protozoan predation, I required the ability to discriminate among multiple fluorescent-labeled cells acquired from complex environments, which to date have not been completely reconciled in the literature. To begin, I needed to address the challenges facing multicolor flow cytometry for the analysis of microbes.

4.1.3 Challenges of multicolor flow cytometry in microbiology

Why are microbial cytometric experiments performed in complex environment so limited when compared to the state of the art for experiments using eukaryotic cells? Such experiments come with unique considerations at each step of the experimental design, implementation, and data analysis and thus must be approached with care. First, appropriate fluorescent tags for experimentally-introduced cells must be selected based on the autofluorescence of growth media and other substances or microorganisms present in the natural environment. Second, analysis of microbes is complicated due to the presence of fine particles which can clog flow systems and distort signals; special sample preparation protocols must be developed and used to remove fine particles from microbes for cytometry, and to eliminate abiotic particle-based events from

collected data. Experiments conducted under laboratory conditions do not require this level of careful preparation.

Use of fluorescent tags to label bacteria as opposed to eukaryotic cells is complicated by microbial cellular properties; the nature of the tagging method is highly dependent on the biology of the organism and of experimental design. Many fluorescent dyes only stain particular structures unique to eukaryotic cells and are thus unusable for prokaryotic cells [155]. Commercial availability of fluorochrome-conjugated antibodies is quite limited for microbes relative to those having specificity to eukaryotic cells and eukaryotic cell-specific proteins [155]. Moreover, antibodies must have extremely high target specificities in order to enable the discrimination of microbial cells having subtle differences. For example, generating antibodies to differentially label a series of bacterial strains varying only with respect to auxotrophies is impractical if not completely impossible. Interestingly, simultaneous use of more than two fluorescent protein tags in bacteria is relatively unexplored despite the broad appeal and versatility of fluorescent proteins for tagging both prokaryotic and eukaryotic cells [70, 246].

Choice of fluorochromes for cell tagging is complicated by the mechanical limitations of the flow cytometer. Although a very large number of fluorescent probes with a wide variety of emission and excitation spectra are commercially available [66, 70, 155, 246], specific lasers are required to excite these proteins and appropriate optical filters must be used to enable detection of fluorochrome emissions. For example, the ubiquitous use of green fluorescent protein in microbial experiments conducted in complex environments as discussed above is not surprising, as GFP is maximally excited by the 488 nm laser [120-122], which is very commonly used in most commercially available flow cytometers. Although much more complicated fluorochrome tagging schemes have been developed and proposed for future consideration [71, 254], these

designs are limited in that they require extremely specific probes and specialized cytometry equipment and software not commonly available to most research facilities [71, 254]. Moreover, these complicated schemes are designed to be performed under ideal laboratory conditions, which is not appropriate for many microbiological experiments.

The largest challenge facing flow cytometry is data analysis, especially that arising from multicolor experiments [254]. Cytometric data analysis is an extremely time-consuming process that is often defined by varying degrees of arbitrary interpretation; see Shapiro for a comprehensive review and discussion [316] on various approaches and pitfalls of cytometric data analysis. As an example, sample data are represented in Figure 17. Briefly, data are first visualized on bivariate plots (histograms or dots). Positive events, or those that comprise actual cells or the desired event, are identified by their FSC and SSC properties as defined by the investigator; additional cell-defining properties are defined as necessary (for example, a fluorescent signal from a cell-specific stain). This information is used to then "gate," or define the region comprising positive signal. This gate can then be applied to other data plots to aid in data analysis. For example, the gates used to identify positive cells on the FSC and SSC detectors can be applied to plots of fluorescent signal distributions, resulting in display of only defined positive cells.

Figure 17A depicts a bivariate dot plot of cells expressing signal on the EGFP and APC detectors. The issue here lies within the area of signal defined by the user that reflects cells positive for EGFP and APC: where should the gating be defined on this plot? Is the black circle gate drawn onto Figure 17A a fair representation of this data? How can positive events be rigorously identified? Moreover, would two different users vary in assessment of positive events? A histogram of fluorescent signal distribution of cells expressing EGFP is shown in

Figure 17B; here, where should the user define positive events? Should the entire region above a certain threshold, as defined by the bar labeled R1, be used to score all positive EGFP events? Moreover, the situation becomes even more ambiguous when examining a mixed population of cells as depicted in Figure 17C. Here, the SSC values of a mixed population of bacterial and amoebae cells are shown in a histogram. How would an investigator define bacterial cells from protozoal cells? Can these two populations be realistically and effectively separated using this parameter?

The sample data presented in Figure 17 illustrate the often arbitrary nature of flow cytometric data analysis using traditional methods. While these methods are appropriate and statistically valid for many forms of cytometric experiments that require only qualitative data, I believe they are woefully insufficient to examine relative ratios of multicolor-labeled bacteria in complex environments. Newer computational approaches to aid in the processing of more quantitative data are currently being developed, but these approaches still require exhaustive analytical processing time as they are based on traditional approaches to data analysis [254]. In this chapter, I address the challenges discussed above in terms of experimental design, data acquisition and data analysis with sufficient rigor to make possible multicolor microbial flow cytometry experiments. Without the technological advances achieved in this work, future experiments addressing the role of protozoan predation in the maintenance of *rfb* diversity in *Salmonella* are impossible.

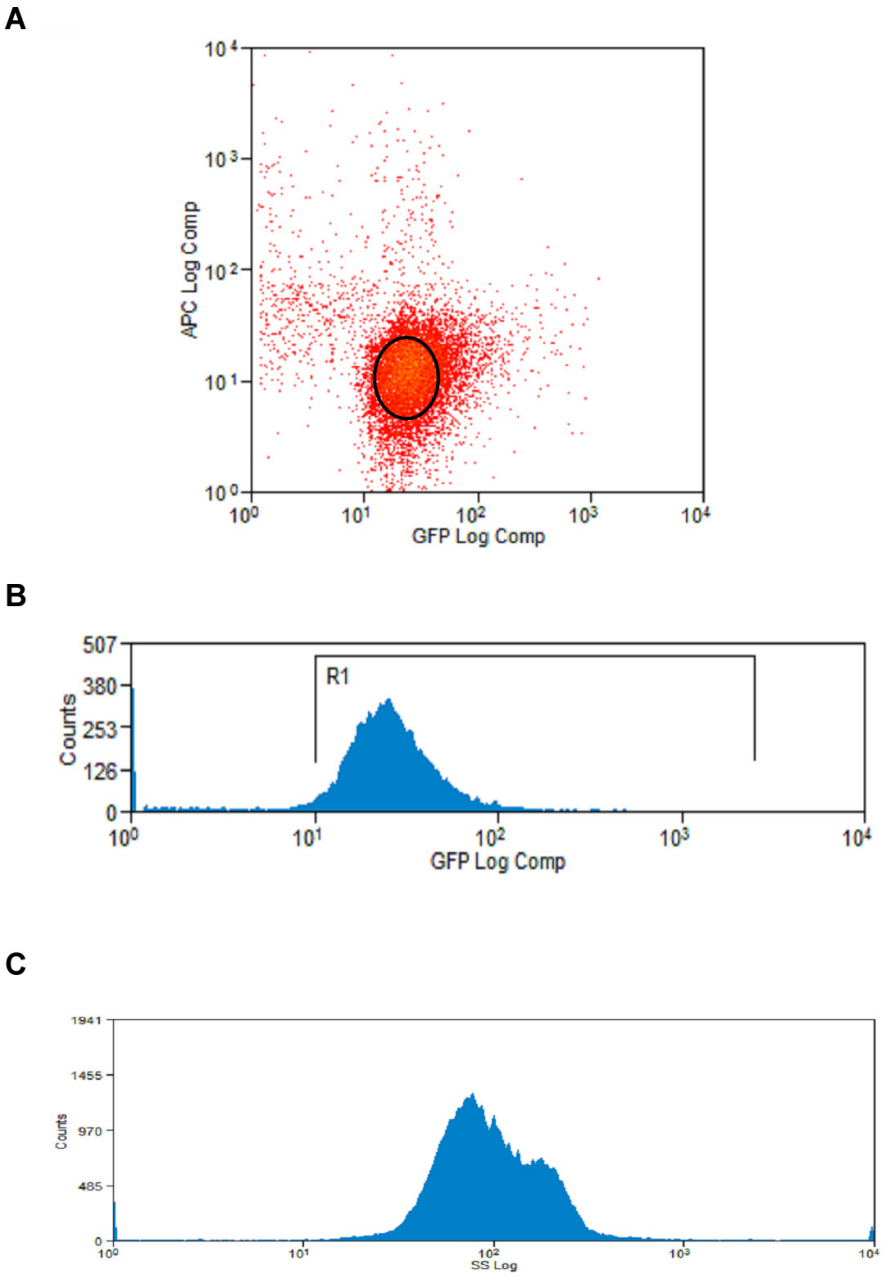

Figure 17. Sample flow cytometric data output from Summit 4.3 software

A. Sample bivariate dot plot of bacterial cells expressing EGFP and labeled with a fluorescent APC-emitting dye. Black circle represents sample gate. **B.** Sample histogram of EGFP signal of bacterial cells. **C**. Sample histogram of side scatter (SSC) of a mixed population of bacterial and amoebae cells. Figures obtained using Summit 4.3 software.

4.2 MATERIALS AND METHODS

4.2.1 Plasmids, strains and growth conditions

Bacterial strains containing fluorescent plasmids were routinely propagated on LB agar and liquid media containing ampicillin at a final concentration of 100-200 µg/mL. Plasmids pEGFP [60, 122], pEYFP [247], pECFP [60, 121-122, 221, 397], and pDsRed-Express2 [342] were commercially obtained (Clontech, Mountain View, CA) and electroporated into *Salmonella* serovar Typhimurium LT2 using protocols discussed in Chapter 2.2.5. Plasmid pRSET/BFP [221] was obtained from Invitrogen (Carlsbad, CA). Plasmid pBAD-mKalama1 [1] was obtained from Addgene (Cambridge, MA). All PCR, restriction digests, ligations, and cloning procedures were performed using well-established protocols [9]. DNA modifying enzymes were obtained from New England Biolabs (Ipswich, MA) and Fermentas Inc. (Glen Burnie, MD).

Plasmid pKAB2 was constructed by subcloning a fragment of the BFP gene from pRSET/BFP into plasmid pECFP using the multiple cloning sites contained within each plasmid; specifically, BsrGI (5' T$^{\downarrow}$GTACA 3') and NcoI (5' C$^{\downarrow}$CATGG 3'). Diagnostic restriction digests using the enzymes PvuII (5' CAG$^{\downarrow}$CTG 3'), BsrGI, EaeI (5' Y$^{\downarrow}$GGCCR 3'), and BsgI (5' GTGCAG$(N)_{16}^{\downarrow}$ 3') and sequencing were used to verify correct replacement off the ECFP gene with the BFP gene in pKAB2. Plasmid pKAB9 was constructed by subcloning a fragment of the mKalama1 gene into plasmid pEGFP using the BseRI (5' GAGGAG$(N)_{10}^{\downarrow}$ 3') and BsrGI restriction sites. Verification of replacement of the EGFP fragment with the mKalama1 fragment was performed using diagnostic restriction digests with BstYI (5' R$^{\downarrow}$GATCY 3') and Sau96I (5' G$^{\downarrow}$GNCC 3').

To construct plasmids containing two fluorescent protein encoding (XFP) genes, high fidelity PCR was used to amplify the region encompassing the P_{lac} promoter through the stop codon of EGFP, EYFP, ECFP and mKalama1 generating a fragment containing AatII (5' GACGT$^{\downarrow}$C 3') and AscI (5' GG$^{\downarrow}$CGCGCC 3') restriction sites at the 5' end and the AatII restriction site at the 3' end. These fragments and pEGFP, pEYFP, pECFP and pDsRed-Express2 were cut with AatII and ligated. Desired orientation for the second site XFP as reverse in order to prevent transcriptional or translational bias from the first XFP; insert verification and orientation were determined using diagnostic restriction digests using FspI (5' TGC$^{\downarrow}$GCA 3'). Plasmids are listed in Table 11.

Table 11. Plasmids containing two fluorescent protein encoding genes used for tagging *Salmonella* cells for flow cytometric analysis

Plasmid	Forward XFP	Reverse XFP
pKAB11	EGFP	EYFP
pKAB13	EGFP	ECFP
pKAB15	EGFP	mKalama1
pKAB19	EYFP	ECFP
pKAB21	EYFP	mKalama1
pKAB27	ECFP	mKalama1
pKAB35	DsRed-Express2	EGFP
pKAB37	DsRed-Express2	EYFP
pKAB39	DsRed-Express2	ECFP
pKAB41	DsRed-Express2	mKalama1

4.2.2 Preparation of cells for cytometric analysis

Cells grown in liquid overnight culture were prepared for cytometric analysis with dilution to a final concentration of 10^7 cells/mL in Phosphate Buffered Saline (PBS) containing 0.02% Tween20 to reduce cell clumping. Cells grown on solid media were eluted from plates with PBS 0.02% Tween20, filtered through a CellMicroSieve™ nylon mesh filter with a pore size of 5 μM (BioDesign of New York, Carmel, NY) to remove debris and protozoa and diluted to a final concentration of 10^7 cells/mL in PBS 0.02% Tween20. Prior to cytometric analysis, 50 nM of , SYTO™ 62 nucleic acid stain (Invitrogen, Carlsbad, CA) was added to samples to aid in the detection of bacterial cells from amoebae and debris.

4.2.3 Flow cytometry

Events were acquired on a Beckman-Coulter CyAn ADP (Beckman-Coulter, Brea, CA) flow cytometer at a rate of <1000 events per second using a 0.3 neutral density filter. Events were triggered on the forward scatter parameter with a threshold of 0.8 and a gain of 30.0. Side scatter was measured at 659 volts with a gain of 10.0. Commercially available software (Summit 4.3, Dako Colorado Inc.) was used for the operation of the cytometer and to determine initial baseline compensation matrices.

4.2.4 Development of Ferdinand, a flow cytometry data analysis program

Ferdinand was developed by Dr. Jeffrey Lawrence, Department of Biological Sciences, University of Pittsburgh.

4.3 ADDRESSING THE CHALLENGES OF EXPERIMENTAL DESIGN

A Beckman-Coulter CyAn ADP cytometer at the Flow Cytometry Core Facility at the Hillman University of Pittsburgh Cancer Institute was made available for use with microbiological samples. The CyAn ADP can measure up to nine colors using three different lasers (405 nm, 488 nm, and 635 nm) and is capable of acquiring data at rates of up to 30,000 events per second. Here, I discuss the strategies I used to maximize the discrimination capacity of this cytometer for use with multiple fluorescently tagged *Salmonella* cells. These modifications were required to enable collection of high-quality data from bacterial populations eluted from solid media.

4.3.1 Addressing trade-offs in fluorescent protein selection

While a highly diverse collection of fluorescent dyes and proteins are available for cell labeling, use of these labels is constrained by 1) brightness of the fluorescent label in a biological context relative to autofluorescence; 2) availability of lasers to excite these labels; 3) use of appropriate filters to channel fluorescence emission signals to the appropriate detectors and eliminating spectral overlap onto other detectors; and 4) constraints imposed by experimental approaches. To examine the fitness of *Salmonella* strains against protozoan predation using the CyAn cytometer, I must address the trade-off between choosing proteins that are a) excited by the 405/488/635 nm lasers of the CyAn and b) suitable for use with a variety of *Salmonella* strains. Signals of these labels have sufficient strength to enable their discrimination from the autofluorescence of *Salmonella* and from each other through the use of optical filters. Also, these labels must easily be introduced into strains without significant genetic manipulation requiring the use of multiple antibiotic resistance cassettes, as I nearly exhausted the spectrum of antibiotic resistance cassettes in *Salmonella* in constructing strains in which the presence and absence of major surface antigens is controlled. Finally, the labels must produce signals that are able to be cytometrically detected in the cells grown in the presence and absence of protozoan predators for up to four days at 33°C.

I investigated the practicality of a palette of six different fluorescent proteins (BFP, mKalama1, EGFP, EYFP, ECFP, and DsRed-Express2) for use in *Salmonella* based on the excitation wavelengths of the lasers available on the CyAn and corresponding emissions of the proteins. With the exception of DsRed-Express2, all of the fluorescent proteins I examined are structural variants of green fluorescent protein, originally isolated from the jellyfish *Aequorea victoria* [323], having enhanced expression properties and different excitation and emission

spectra [1, 60, 121-122, 221, 247]. DsRed-Express2 is noncytotoxic variant of the original *Discosoma* DsRed protein with a faster maturation time designed for maximal expression in rapidly dividing organisms [342]. Theoretically, BFP, mKalama1 and ECFP should be excited by the 405 nm laser while EGFP, ECFP, EYFP and DsRed-Express2 should be excited by the 488 nm laser (Figure 18); I chose to reserve the far-red 635 nm laser for additional fluorescent labels to discriminate *Salmonella* from debris and amoebae. Note that ECFP can be excited by both the 405 nm and 488 nm lasers; this is discussed later.

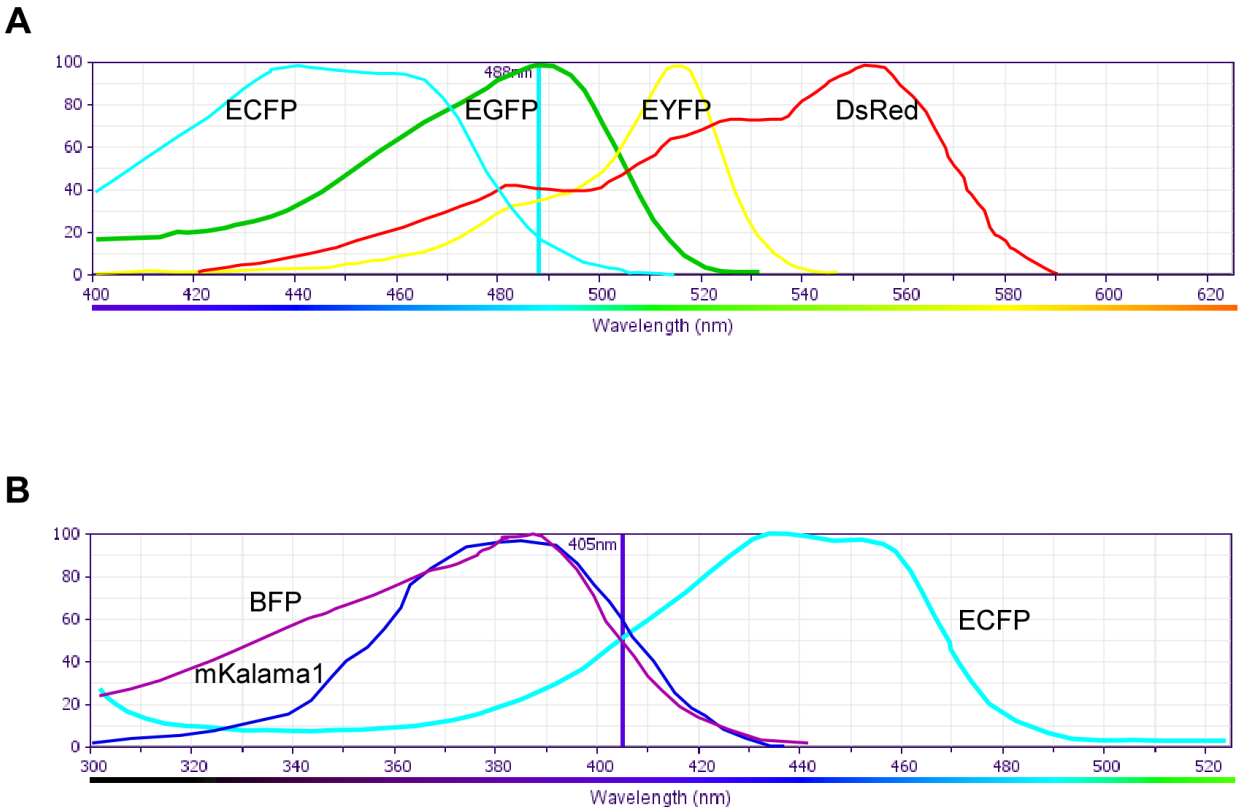

Figure 18. Excitation spectra for six fluorescent proteins

A. Excitation spectra for ECFP, EGFP, EYFP, and DsRed-Express2 by the 488 nm laser.
B. Excitation spectra for BFP, mKalama1, and ECFP by the 405 nm laser. Figures drawn using the BD Fluorescence Spectrum Viewer [29].

I chose vector-based labeling of *Salmonella* cells with these fluorescent proteins expressed from the lactose operon promoter P_{lac} in the high-copy number vector backbone pBR322 [31]. These tagging vectors are easily introduced into cells via electroporation, and the high-copy number vector backbone and constitutive expression of P_{lac} in *Salmonella* should produce large amounts of fluorescent proteins with bright signals. Additionally, the pBR322 vector backbone carries the ampicillin resistance cassette *bla* for selection; ampicillin was not used in construction of strains used to assess fitness against predation and is compatible with my strain construction strategy. Use of this vector tagging system provides a large degree of experimental flexibility over other tagging methods.

In my experiments, cells must be fluorescently labeled prior to their use in competition tests, which is easily accomplished with the use of tagging vectors, as no practical manner to label cells post-competition exists. Fluorescent dyes are of little use in my experiments beyond labeling *Salmonella* cells to enable their separation from debris, as cells are mostly physiologically identical with the exception of surface antigens. Antibodies having enough specificity to discriminate among my collection of antigenically-diverse non-Typhimurium strains do not currently exist. Generation of antibodies with fluorescent conjugates is very expensive and is most likely impractical given that surface antigens are often the targets of antibodies; the presence and absence of major *Salmonella* surface antigens are experimentally controlled in the strains used to assess fitness against predation.

Additionally, I require stable, bright signals from the fluorescent labels used in my experiments. Predation competition assays require three to four days of incubation at 33°C; fluorescent labels must have sufficient stability to permit cytometric detection of emitted signals after this period of incubation. Of the proteins that produce fluorescent signal visible to the

human eye, I found that the signal of these proteins as emitted from cells grown in both liquid and solid media remains visible for several days of growth.

I electroporated *Salmonella* serovar Typhimurium LT2 with the tagging vectors pEGFP, pEYFP, pECFP, pDsRed-Express2, Vectors pKAB2 and pKAB9 contain BFP and mKalama1 expressed from P_{lac} in the pBR322 vector backbone, respectively. I determined the ability to detect signals from these proteins using overnight cultures of these strains. I was able to discern bright signals from all strains except for the strain tagged with pKAB2; the expected signal from this protein was not visible against the background autofluorescence of *Salmonella*. Theoretically, the emission spectrum of BFP indicates that this protein should have been excited by the 405 nm laser (Figure 18). However, the propensity for photobleaching and low quantum yield of BFP as compared to EGFP [356] led me to conclude that BFP was not appropriate for use in my experiments, leaving a total of five single colors with potential expansion to ten combinations of two colors each as a strain tagging scheme.

4.3.2 Addressing trade-offs in fluorophore selection with appropriate optical filters

The use of multiple fluorescent tags excited by the same laser allows access to larger number of flow cytometers; the trade-off lies in signal discrimination. Fluorescent proteins excited from the same laser can often have significant overlap in their respective emission spectra, producing signals that "spill-over," or register false positive signals, into other detectors [316]. Because the emission signals of EGFP, EYFP, ECFP, mKalama1, and DsRed-Express2 have considerable amounts of spectral overlap (Figure 19), it is critical that I select optical filters that carefully balance the direction of signals to the appropriate detectors without significantly compromising signal strength and introducing unwanted signal spill over onto the detectors of other

fluorophores. For example, I needed to select filters to channel the EGFP signal onto its detector that minimized unwanted spill over of the EGFP signal onto the EYFP detector while simultaneously maximizing the amount of genuine signal channeled onto the EGFP detector. The filter scheme depicted in Figure 19 satisfies these requirements. Although the emission from ECFP is considerably lower on the 488 nm laser (Figure 19A) than from the 405 nm laser (Figure 19B), the spectral overlap from the mKalama1 emission signal is too close to that of ECFP as excited at 405 nm; only one protein can be excited from this laser. Thus, my design reflects a trade-off between being able to use mKalama1 with a drastically reduced ECFP signal as obtained from the 488 nm laser.

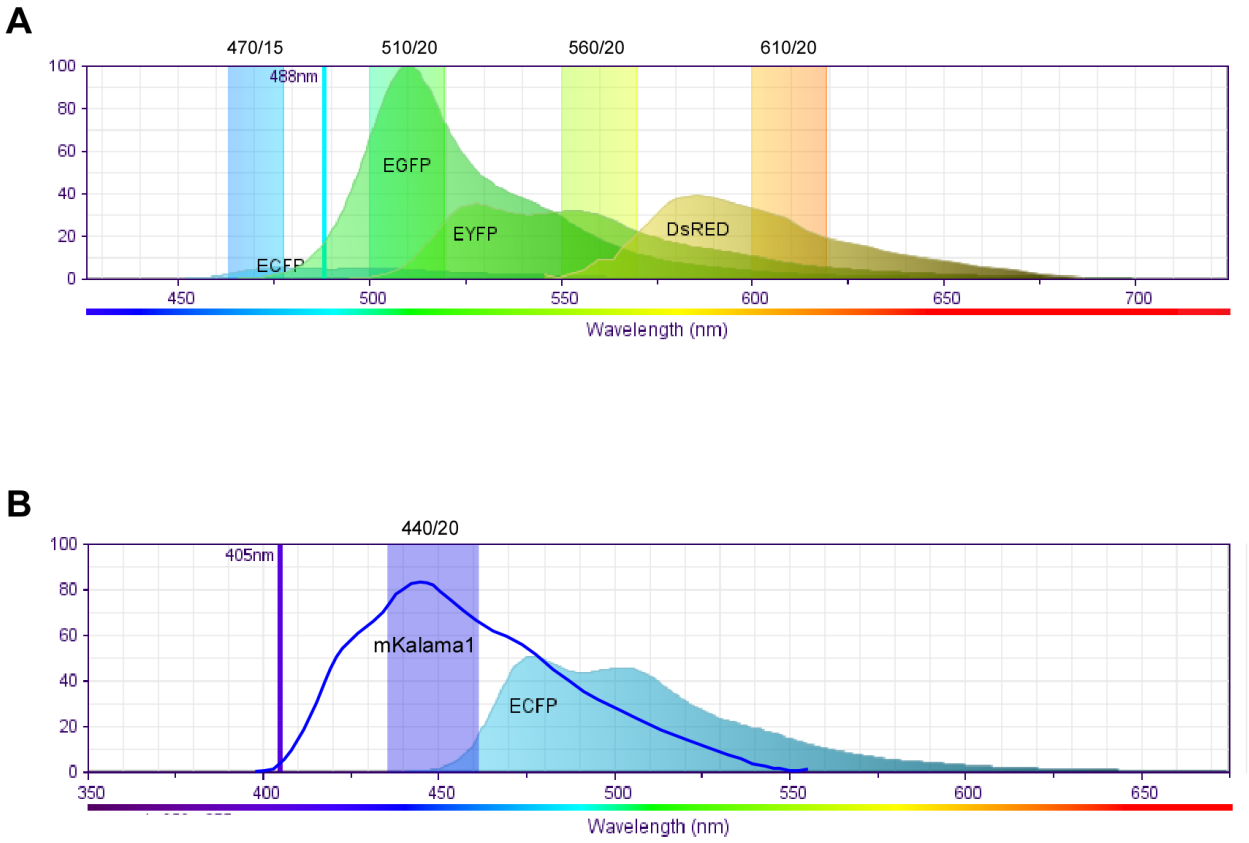

Figure 19. Emission spectra and optical filters for simultaneous detection of multiple fluorescent proteins

A. Emission spectra for fluorescent proteins ECFP, EGFP, EYFP, and DsRed-Express2 as excited by the 488 nm laser. **B.** Emission spectra for fluorescent proteins mKalama1 and ECFP as excited by the 405 nm laser. Because the emission spectra for mKalama1 has significant overlap with that of ECFP as excited at 405 nm, ECFP must be excited using the 488 nm laser despite lower emission from this laser. Optical filters highlight the area of signal emission transmitted to the appropriate detector and are listed at the top of each spectra. Figures drawn using the BD Fluorescence Spectrum Viewer [29].

4.3.3 Addressing trade-offs in signal strength

The use of selective filters to discriminate among fluorophores with similar emissions spectra reduces the overall signal being detected. This limitation can be addressed by increasing the voltage of the appropriate detector, thus making the detectors more sensitive to emitted signal. In essence, this produces "brighter" emission signals through enhanced sensitivity in detection. A trade-off is introduced, however, in that brighter detector voltages increase spillover between optical detectors.

I used cells of serovar Typhimurium LT2 strains tagged with pEGFP, pEYFP, pECFP, pKAB9, and pDsRed-Express2 grown in liquid overnight cultures to determine appropriate baseline detector voltages (Table 12). The APC detector was used to identify cells stained with the SYTO™ 62 nucleic acid dye and is unrelated to the fluorescent proteins used to tag cells. I used the principle of spillover coefficients to guide detector settings; the spillover coefficient reflects the spectral overlap of a fluorescent signal onto the detectors of other fluorescent signals that share similar emission spectra [12, 284-285]. The spillover coefficient for any given fluorescent signal is expressed as a percentage value reflecting the ratio of mean signal strength on incorrect detectors to mean signal strength on the correct detector [12, 284-285]. The spillover coefficient for any fluorophore should be less than 50% on the detectors specific for other fluorophores to ensure their optimum separation, which is what I observed (Table 13).

Table 12. Detector voltage settings for discrimination of emissions from five fluorescent proteins as expressed by cells grown in liquid culture

Detector	Voltage
EGFP	600
EYFP	573
ECFP	625
mKalama1	950
DsRed-Express2	781
APC/SYTO™ 62	950

Table 13. Spillover coefficients for emissions from five different fluorescent proteins as expressed by cells grown in liquid culture

	Detector				
Signal	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	1.00	0.40	0.01	0.02	0.07
EYFP	0.39	1.00	0.00	0.00	0.23
ECFP	0.10	0.14	1.00	0.07	0.05
mKalama1	0.00	0.00	0.26	1.00	0.00
DsRed-Express2	0.02	0.09	0.00	0.01	1.00

4.4 ADDRESSING THE CHALLENGES OF DATA ACQUISITION

Although I was able to select appropriate optical filters and voltage settings to permit the simultaneous discrimination of strong signals from five different fluorescent proteins, these conditions represent "ideal" signal expression. Cells used for the development of initial cytometer settings discussed above were grown separately overnight in well-aerated liquid cultures. The experimental conditions required to test fitness of *Salmonella* against protozoan predation are undoubtedly more complex than pure, saturated overnight cultures of cells. The competition tests proposed here involve incubating cells for several days on solid media in the presence and absence of amoebae, which influence signal strength. Thus, I needed to optimize microbiological and experimental conditions that a) permit maximal expression of fluorescent proteins after several days of incubation; and b) are conducive to the growth of both *Salmonella* and a diverse collection of amoebae. Additionally, I needed to further develop cytometric protocols and settings that maximize the cytometric discrimination capacity of cells grown under representative experimental conditions.

4.4.1 Growth condition optimization for cells grown on solid media

After baseline voltage settings were determined, modifications of these parameters were explored for cells used in protozoan predation competition experiments. Voltage determination using cells grown in liquid cultures represent optimal signal brightness; protozoan predation competition experiments require cells to be grown on solid media for 72-96 hours, leading to substantial loss in fluorescent signal brightness. Therefore, I used the voltage settings obtained under optimal conditions as a baseline upon which to build appropriate detection parameters for

experiments performed in complex environments. To explore these conditions, I grew liquid overnight cultures of strains of serovar Typhimurium LT2 KAB002 tagged with pDsRed-Express2, pEGFP, pEYFP, pKAB9-mKalama1 and pECFP and diluted them in PBS to an OD_{600} of ~1.00. I spread 100 μL of each strain onto NM-C plates; the centers of half of the plates were seeded with 10^4 *Naegleria gruberi* NL cysts. Plates were incubated for 84 hours at 33°C to permit predators to sweep the entire plate of bacteria, after which they were prepared for cytometric analysis. Because signals from cells grown on solid media for extended times were weaker than that of cells grown with aeration in liquid culture, I accordingly adjusted detector voltages (Table 14). Spillover coefficients among detectors were within appropriate ranges (Table 15).

Table 14. Detector voltage settings for discrimination of emissions from five fluorescent proteins from cells grown on NM-C solid media

Detector	Voltage
EGFP	470
EYFP	508
ECFP	636
mKalama1	880
DsRed-Express2	565
APC/SYTO™ 62	850

Table 15. Spillover coefficients for the emissions of five fluorescent proteins from cells grown on solid NM-C media

	Detector				
Signal	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	1.00	0.47	0.01	0.01	0.14
EYFP	0.38	1.00	0.01	0.01	0.39
ECFP	0.08	0.06	1.00	0.08	0.06
mKalama1	0.27	0.08	0.49	1.00	0.09
DsRed-Express2	0.06	0.12	0.18	0.28	1.00

4.4.2 Addressing trade-offs in growth conditions

While the use of fluorescent proteins avoided the tagging limitations imposed by strain-specific fluorescent antibodies or structure-specific dyes, fluorescent proteins require ready access to molecular oxygen for proper folding [356]. The requirement for access to molecular oxygen for fluorescent protein-tagged cells led to the introduction of an additional trade-off between media conditions that allow for optimal predator discrimination among prey and those that allow for optimal expression of fluorescent proteins.

NM and NM-LG media were previously used to co-culture *Salmonella* and amoebae [386-387]. A variant formulation of NM containing 200 µg/mL ampicillin for vector selection and substituting 0.2% glycerol for glucose to permit expression of EGFP as driven by P_{lac} was previously used to examine strains of *Salmonella* tagged with pEGFP in the presence of amoebae [385]. Also, the work of Wildschutte and Lawrence previously determined that the use of NM-LG better supported growth of amoebae than NM [386]. However, none of the media formulations described above were designed to support intricate competition experiments requiring strains to express multiple fluorescent proteins and maintain the growth of natural isolates of amoebae. Thus, I determined the optimum growth medium formulation and cell density of starting inoculum to maximize fluorescent protein signals, produce uniform lawns of cells and support the growth of diverse amoebae.

Using serovar Typhimurium LT2 KAB002 tagged with pDsRed-Express2 as a reporter strain, I first tested several formulations of growth media and density of starting inoculum for the ability to produce uniform lawns. I spread 100 µL of three different concentrations of an overnight culture onto NM 1.5% agar containing three different concentrations of glycerol. NM 1.5% agar was supplemented with high (1X or 0.2%), medium (0.7X or 0.14%), or low (0.1X or

0.02%) concentrations of glycerol. Starting cell densities were measured as OD_{600} values of high (1X or ~1.4), medium (0.7X or ~1.0), or low (0.1X or ~0.2). Cells were eluted from the plate and analyzed on the flow cytometer to determine the strength of the fluorescent signals.

Interestingly, I generally observed that conditions that favor bacterial cell growth (higher glycerol concentration and increased density of starting inoculum) tended to produce two populations of cells with slightly different signal strengths, indicated by deviations from a normal Gaussian distribution (Figures 20-22). Figure 20 depicts different cell densities of starting inoculum when plated to media containing high glycerol; while I observed thick, uniform bacterial lawns that are ideal for predation experiments, these plates contained visible areas of cell growth having dimmer signal than those on other areas of plates (Figure 20AB). When assessed by cytometry, cells grown under high glycerol conditions produced what appears to be two populations of cells; this is best illustrated in the "tails" at the lower end of the signal distribution (Figure 20AB). Cells grown under moderate glycerol concentrations tended to display very strong signals as visually assessed (Figure 21). However, plates with high density of starting inoculum also displayed visible regions on plates of cells having dimmed signal (Figure 22A). When assessed by cytometry, cells grown under moderate glycerol concentrations produced signals having varying degrees of deviations from normal Gaussian distributions (Figure 22). Conditions that least favor bacterial cell growth (lower glycerol concentration and decreased density of starting inoculum) tended to produce more uniform signal distributions (Figures 20-22), but a trade-off was introduced in terms of production of a uniform bacterial lawn suitable for protozoan predation experiments. Cells grown on low glycerol produced more uniform signal distributions (Figure 22), with cells grown under conditions of low glycerol and moderate density of starting inoculum produced the best normal Gaussian signal distribution of

all conditions (Figure 22B). However, these conditions also produced the most non-uniformly distributed lawns of bacteria; single colonies are visible in portions of the plates (Figure 22). Starting inoculum of low cell density across all glycerol concentrations (Figures 20C, 21C and 22C) did not produce uniform lawns of bacteria; these lawns were very thin and often displayed areas of single colonies. Regardless of signal, low concentration of starting inoculum is inappropriate for use in predation experiments. Moderate glycerol concentration and density of starting inoculum appear to best address the trade-off between cell lawn uniformity and adequate signal strength (Figure 21B).

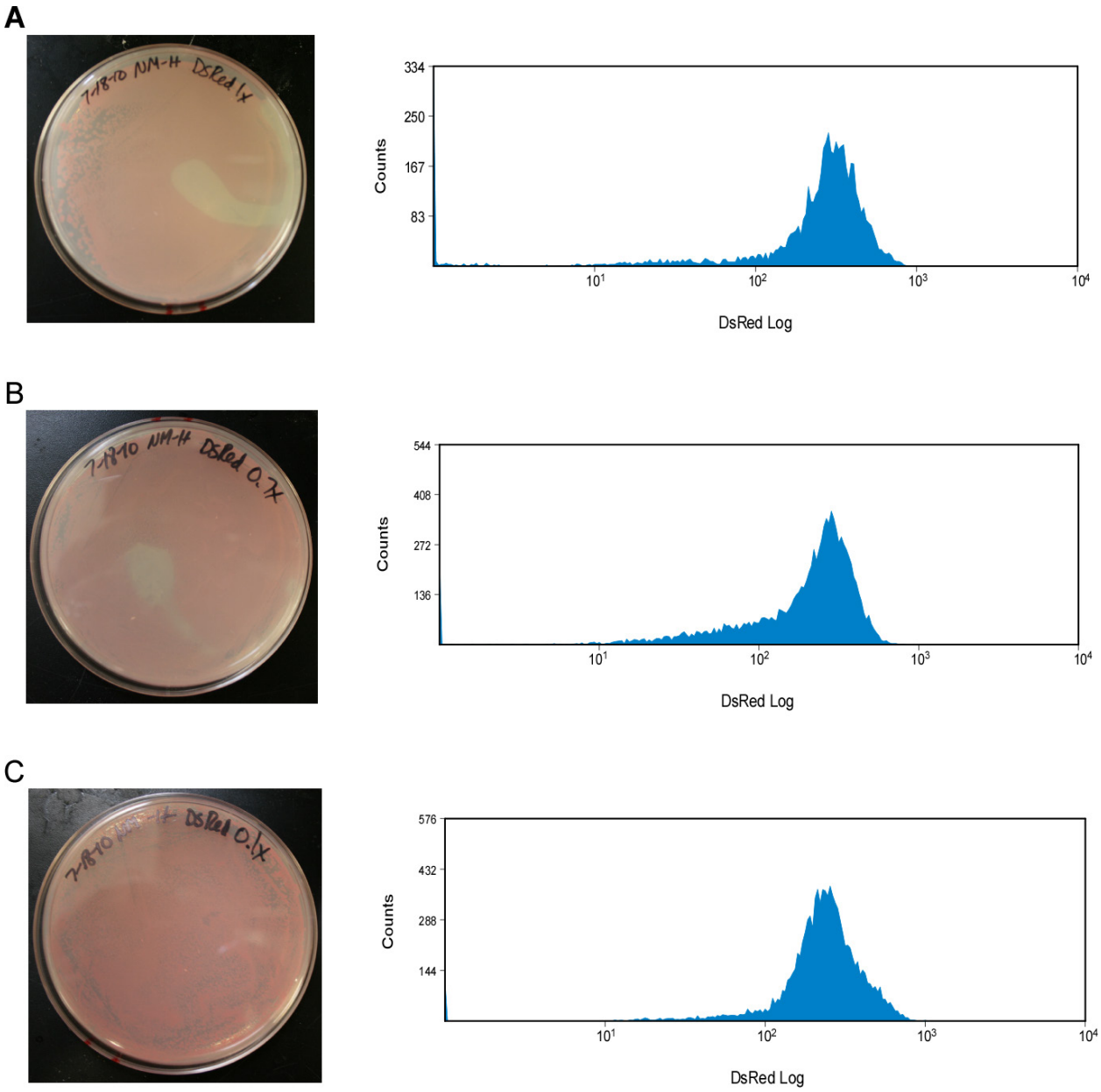

Figure 20. DsRed-Express2 expression in cells grown on solid NM-C media with 0.2% glycerol

A. High density of starting inoculum (OD_{600} ~ 1.4). **B**. Moderate density of starting inoculum (OD_{600} ~ 1.0). **C**. Low density of starting inoculum (OD_{600} ~ 0.7).

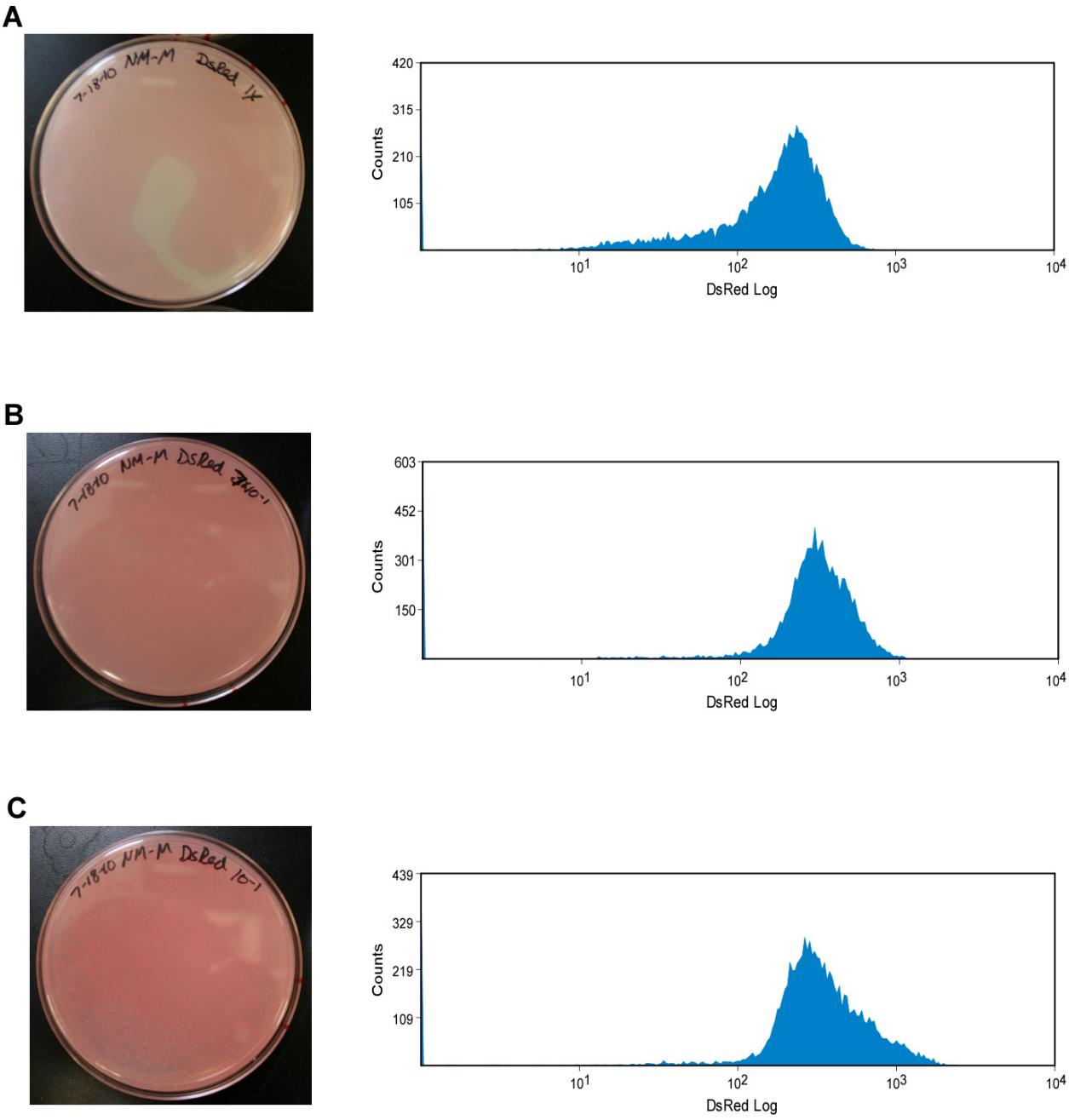

Figure 21. DsRed-Express2 expression in cells grown on solid NM-C media with 0.14% glycerol

A. High density of starting inoculum (OD_{600} ~ 1.4). **B**. Moderate density of starting inoculum (OD_{600} ~ 1.0). **C**. Low density of starting inoculum (OD_{600} ~ 0.7).

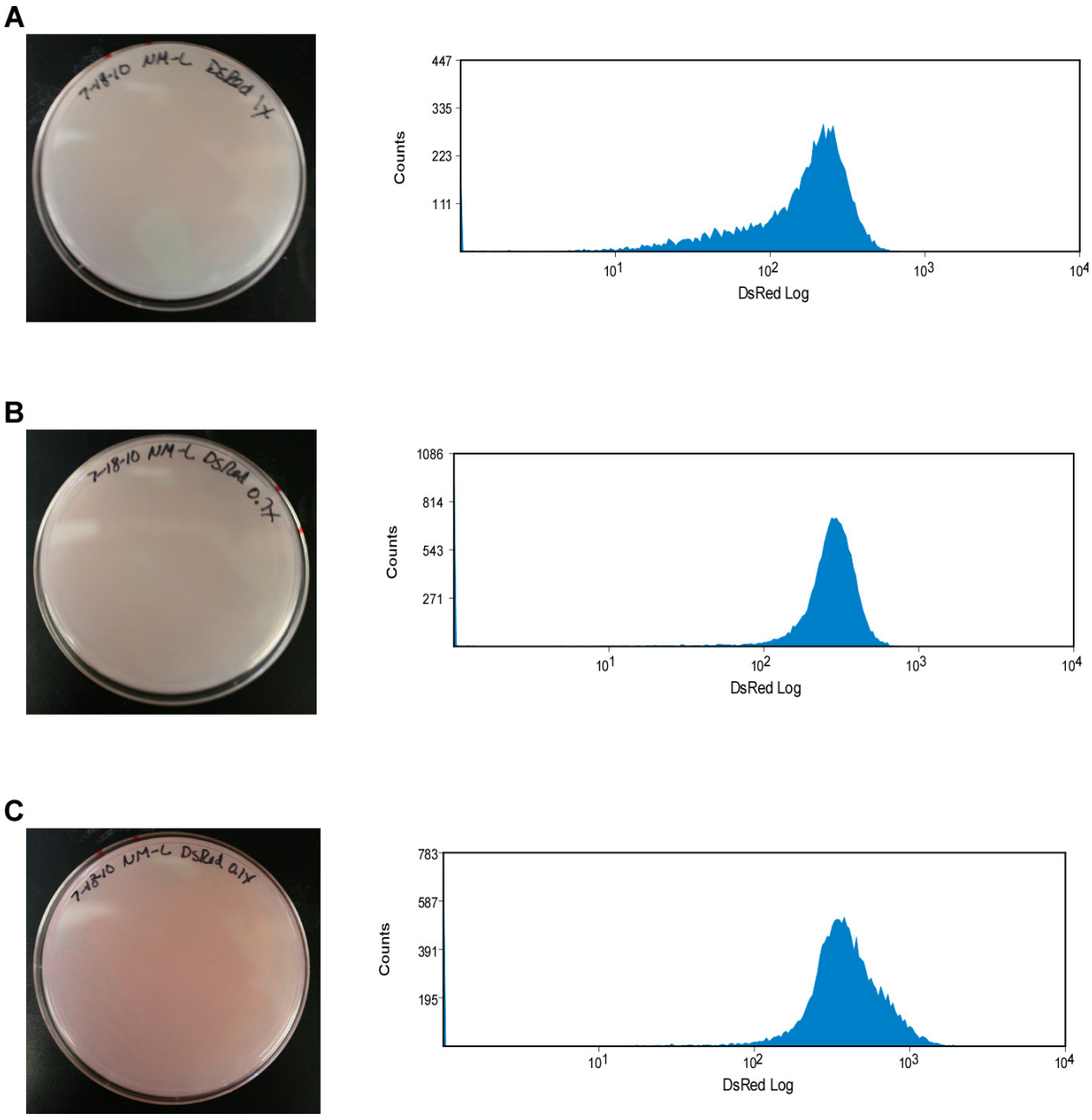

Figure 22. DsRed-Express2 expression in cells grown on solid NM-C media with 0.01% glycerol

A. High density of starting inoculum (OD_{600} ~ 1.4). **B**. Moderate density of starting inoculum (OD_{600} ~ 1.0). **C**. Low density of starting inoculum (OD_{600} ~ 0.7).

Expression of P_{lac} is constitutive in *Salmonella*; nutrient deprivation should cause upregulation of this promoter, resulting in greater expression of the fluorescent proteins. Thus, I do not think that lack of nutrients resulting from higher cell density is responsible for the dampened signal both visibly and cytometrically observed on plates with high glycerol concentration and density of starting inocula. However, the fluorescent proteins used in this assay require oxygen for proper folding; differences in oxygen availability in thicker cell lawns can explain the differences in signal strength I observed on plates with conditions favoring bacterial growth. Adequate predation competition experiments require bacterial cell lawns of uniform density and signal strength. Therefore, I developed NM-Cytometry media, or NM-C, for use with fluorescently-tagged cells: 15.5 mM potassium phosphate pH 7.5; 0.2% peptone; 0.1% glycerol; 1.5% agar). I found that using starting inoculum with OD600 values of ~1.00 on NM-C media produced uniform lawns without sacrificing signal quality. I was able to propagate seven different amoebae successfully using these growth conditions, highlighting the utility of NM-C media for a variety of protozoan predation experiments (Table 16).

Table 16. Amoebae successfully propagated on nutrient media for cytometry (NM-C)

Amoeba	Source
Acanthamoeba polyphaga I	Pondwater [387]
Acanthamoeba sp. R2-1	Red-eared slider (*Trachemys scripta*) [386]
Hartmannella sp. T3-1	Bullfrog tadpole (*Rana catesbeiana*) [387]
Naegleria gruberi NL	Laboratory strain
Naegleria sp. F1-27	Goldfish (*Cassius auratus auratus*) [386]
Tetramitus sp. BD1-1	Juvenile bearded dragon (*Pogona barbata*) [386]
Tetramitus sp. F1-15	Goldfish (*C. auratus auratus*) [386]

4.5 ADDRESSING THE CHALLENGES OF DATA ANALYSIS IN MULTICOLOR FLOW CYTOMETRY

Flow cytometry does not result in the unambiguous identification of different classes of cells; rather, the light-scattering and fluorescence properties of cells are recorded and users must evaluate these data to discriminate among cell- and non-cell-initiated events. A number of challenges were presented that are addressed herein. To accomplish this, I developed a novel program for the import and analysis of flow cytometric data. This program, Ferdinand, integrates the three tasks using cytometric data: a) the import and collation of data files generated by flow cytometers, b) the robust enumeration of the number and character of fluorescent cells, and c) the calculation of fitness values among genotypes based on these data. The sole scientific issue associated with the first function is that data are pooled from multiple replicate experiments so that the replicate number is retained; this allows the examination of variability among replicates. The last point will be discussed in later chapters. Here I discuss the innovations required to enumerate fluorescent cells accurately.

4.5.1 Addressing overlap in emissions spectra

The use of fluorescent proteins excited by the same laser results in substantial overlap in their emissions spectra. While the use of wavelength selective filters will maximize the light being directed to optical detectors dedicated for a specific fluorescent protein, spillover between detectors cannot be eliminated using filters alone as the emissions spectra of fluorescent proteins is quite broad (see Figure 19).

The problem of spectral overlap is solved with compensation, or the process of removing the false positive spill-over signals of other fluorescent signals from the detector of interest [12, 284-285]. For example, I may dedicate an optical detector with a 510 ± 20 nm filter as registering fluorescence from EGFP; however, light from EYFP will also be captured by this detector. Therefore, the total signal registered on a EGFP detector reflects input from both EGFP and EYFP. Considering five detectors and a system of five fluorophores, I deconvolute these signals as:

$$\mathbf{R} = \mathbf{M}*\mathbf{A}$$

where **R** is the 1x5 matrix of raw signal detected, **A** is the 1x5 matrix of actual signal produced, and **M** is the 5x5 compensation matrix indicating the degree to which each fluorophore registers on each detector. Individual values are determined using control cultures bearing single fluorophores and noting the magnitude of signal on non-dedicated detectors. Compensation values are chosen so that signal on these detectors is less 10. I then solve for the actual signal produced by each fluorophore by simple linear algebra:

$$\mathbf{M}^{-1}*\mathbf{R} = \mathbf{M}^{-1}*\mathbf{M} * * = \mathbf{A}$$

The process of proper signal compensation for complex cultures comprises three steps. First, a compensation matrix is deduced as above for cells grown in liquid media; this is the least complex mixture of events and allows for initial compensation values to be determined with accuracy.

Cells grown overnight in liquid culture consistently yielded the strongest fluorescent signals relative to other growth conditions, but noise was consistently lower than that of cells grown on solid media for several days. Thus, a compensation matrix was produced that reflected these experimental conditions. Compensation values were deduced by examining cells bearing a

single fluorescent protein, and removing the contribution of this protein from optical detectors dedicated to other fluorophores (Table 17).

Table 17. Compensation matrix for single fluorescent protein-tagged cells grown in liquid culture

	Signal Compensation				
Detector	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	100.00	38.8	0.00	0.00	0.00
EYFP	48.45	100.00	6.42	0.00	0.00
ECFP	1.16	0.00	100.00	0.00	0.00
mKalama1	0.00	0.00	0.00	100.00	0.00
DsRed-Express2	8.94	23.85	0.00	0.00	100.00

Second, this matrix is altered for use with cells grown for 72-96 hours on solid media, either in the presence (Table 18) or absence of predators (Table 19). Not surprisingly, fluorescence from these cells is weaker than that from cells grown overnight on liquid media. Thus, it is necessary to accordingly adjust compensation values for these experiments. The initial compensation matrix deduced above provides an excellent starting point; deducing a compensation matrix de novo using plate-generated data is difficult as the number of events attributable to noise is far higher, obfuscating genuine data points.

Table 18. Compensation matrix for single fluorescent protein-tagged cells grown on solid media in the absence of protozoan predators

	Signal Compensation				
Detector	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	100.00	43.11	0.00	0.00	0.00
EYFP	44.26	100.00	0.00	0.00	4.30
ECFP	3.61	0.03	100.00	71.13	1.10
mKalama1	3.76	0.00	0.00	100.00	0.00
DsRed-Express2	12.85	38.16	0.00	0.00	100.00

Table 19. Compensation matrix for single fluorescent protein-tagged cells grown on solid media in the presence of protozoan predators

	Signal Compensation				
Detector	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	100.00	45.31	0.00	0.00	0.59
EYFP	36.75	100.00	0.00	0.00	2.83
ECFP	11.45	9.27	100.00	81.67	1.21
mKalama1	11.45	6.02	0.00	100.00	1.60
DsRed-Express2	15.03	42.21	0.00	0.00	100.00

Lastly, the compensation matrix must be adjusted for use in doubly-labeled cells. After proper compensation, one would expect the strength of a GFP-labeled cell to be identical among population doubly-labeled with GFP and another fluorophore. Difference in the relative GFP intensity among these classes would reflect imbalances in the compensation matrix. It was at this step that a weakness in traditional flow cytometry methods was uncovered. Despite my best efforts, the compensation matrix deduced using the Summit program failed to resolve the individual signals in doubly-labeled cells. That is, raw signals on all five fluorescent detectors could not be deconvoluted reliably into only two signals from the known fluorescent proteins. Rather, multiple signals were predicted, which did not reflect the composition of my cells.

This weakness is attributable to two sources. First, the compensation matrix plays the central role in separating raw signal captured by each detector into predicted actual originating from fluorescent proteins; even subtle imbalance in this matrix will lead to a failure in predicting genuine fluorescence. Second, there is no correction for the autofluorescence of the sample. Genuine signal is predicted solely from the raw signal and the compensation matrix; thus the relative contribution of autofluorescence varies between weak and strong raw signals. This variable contribution would require either separate compensation matrices for weak and strong signals, or correction for autofluorescence before matrix deconvolution.

I took a more rigorous approach to the construction of the compensation matrix. First I examined fluorescent signal from unlabeled cells to estimate autofluorescence under each experimental growth condition. Next, rather than proceeding individually through each detector on Summit and eliminated perceived spillover, I used Ferdinand to calculate spillover directly after having eliminated the contribution of baseline autofluorescence (Table 20). This resulted in compensation matrices that were subtly different from those calculated using Summit (Table 21).

These small differences translated into vast improvements in the prediction of genuine signal; in general, fluorescent events had two major fluorescent contributors, and the fluorescence from other sources was minimized.

Table 20. Compensation matrix for double-fluorescent protein tagged cells grown on solid NM-C media obtained using Ferdinand

	Signal Compensation				
Detector	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	100.00	35.49	2.22	0.20	0.24
EYFP	47.52	100.00	0.99	0.01	4.45
ECFP	0.19	0.23	100.00	35.08	0.16
mKalama1	1.51	0.77	4.70	100.00	0.00
DsRed-Express2	15.14	43.81	1.92	0.61	100.00

Table 21. Compensation matrix for double-fluorescent protein tagged cells grown on solid NM-C media obtained using Summit 4.3

	Signal Compensation				
Detector	EGFP	EYFP	ECFP	mKalama1	DsRed-Express2
EGFP	100.00	42.00	0.00	0.00	0.00
EYFP	41.00	100.00	0.00	0.00	0.00
ECFP	0.00	0.00	100.00	40.5	0.00
mKalama1	0.00	0.00	0.00	100.00	0.00
DsRed-Express2	8.50	51.20	0.00	0.00	100.00

4.5.2 Sorting fluorescent cells into classes: The failure of threshold methods

Despite the care taken to increase signal strength, to isolate each fluorophore's emissions onto dedicated optical detectors, to minimize spillover and to deconvolute magnitude of emitted light from the values recorded on detectors, the data provided from flow cytometers are not easily translated into unambiguous sorting of events into classes of fluorescent cells. First a great number of events originate from debris, including dead cells, amoebae and fragments of agar. These events are problematic in that they can show autofluorescence, or may contain bound fluorescent proteins (*e.g.,* a protein aggregate). Others represent clumps of bacterial cells preventing unambiguous classification. As described above, many of these events can be eliminated from consideration by using spectrophotometric properties of the events, such as pulse width (PW), side scatter (SSC) or the amount of DNA present (measured by binding the SYTO™ 62 APC fluorescent stain).

When all of these events are removed, one can then consider the strength of fluorescent signal present on each dedicated detector after signals have been compensated for spillover. If one follows conventional practice, then events registering above that of autofluorescence on fluorescent scales (where 10^4 units is the maximum possible for an optical detector) are scored as positive events [316]; in this case, the autofluorescence of *Salmonella* consistently registered below 10 units on all detectors. This approach is satisfactory when one is solely interested in presence/absence measures but is severely problematic when careful measurements of the relative abundance of different fluorophores are being collected. As illustrated in Figure 17 located at the beginning of this chapter, two critical problems are illustrated in Figure 21; here, a mixture of five cell types – each labeled with a single, distinct fluorophore – is being analyzed. First, it is clear that a simple threshold does not separate clear positives from clear negatives;

values of compensated signal strength show a continuous range of values (Figure 21); for some fluorophores (*e.g.,* ECFP in Figure 21E), signal is much closer to the arbitrary threshold of 10; my experimental design must accommodate situations in which fluorescent signals are weaker. Thus, I must have a valid way of discriminating signal within a range of values that are very likely to represent positive events, while in others it does not. In this example, standard threshold analyses failed to accurately classify collected data. Here, only single color-tagged cells were analyzed; threshold methods identified double and triple-color events that did not exist in this data set (Table 22).

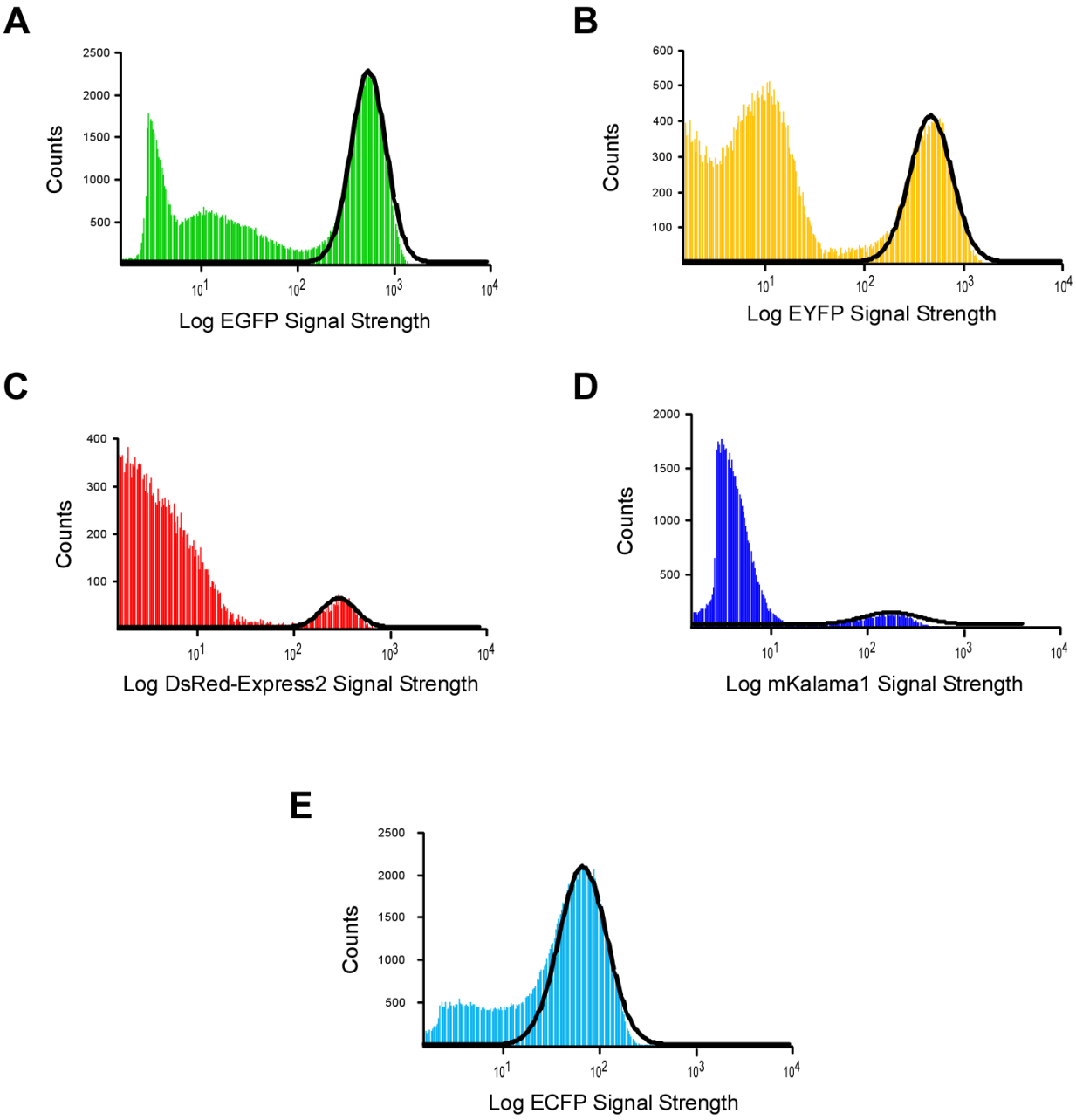

Figure 23. Histograms of compensated fluorescent signal from a mixture of five *Salmonella* cells each tagged with a unique fluorescent protein

A. EGFP. **B.** EYFP. **C.** DsRed-Express2. **D.** mKalama1. **E.** ECFP. Black curves indicate Gaussian distributions fit to peaks using Ferdinand.

Table 22. Simple threshold methods of data analysis fail to resolve single-fluorescent tag data

G = EGFP; Y = EYFP; R = DsRed-Express2; K = mKalama1; C = ECFP. Dashed lines indicate the absence of event counts classified into a given fluorescent category.

Category	Total	Replicate					
		1	2	3	4	5	6
Unclassified	77652	16196	12464	14340	14795	11848	8009
G----	81050	12678	18025	11891	9454	11951	17051
-Y---	14935	2810	2575	2340	2269	2145	2796
--R--	2418	386	332	365	351	361	623
---K-	4116	452	480	655	587	697	1245
----C	90403	13897	15515	14647	11795	14281	20268
GY---	6650	748	1021	1116	844	1250	1671
G-R--	280	60	70	35	19	46	50
G--K-	479	68	72	85	51	73	130
G---C	3199	647	820	455	341	414	522
-YR--	629	106	102	108	93	86	134
-Y-K-	36	8	6	5	6	2	9
-Y--C	283	64	57	44	26	39	53
--RK-	14	3	3	1	2	3	2
--R-C	172	43	33	23	15	21	37
---KC	2484	247	267	342	269	399	960
GYR--	57	17	14	6	5	4	11
GY-K-	81	3	9	21	7	15	26
GY--C	14	1	3	5	2	1	2
G-RK-	3	0	1	1	1	0	0
G-R-C	5	2	1	1	0	1	0
G--KC	63	5	10	9	6	12	21
-YRK-	0	0	0	0	0	0	0
-YR-C	3	0	0	0	1	0	2
-Y-KC	7	1	0	1	1	3	1
--RKC	4	0	0	2	0	1	1
GYRK-	0	0	0	0	0	0	0
GYR-C	0	0	0	0	0	0	0
GY-KC	0	0	0	0	0	0	0
G-RKC	0	0	0	0	0	0	0
-YRKC	0	0	0	0	0	0	0
GYRKC	0	0	0	0	0	0	0

4.5.3 Z-gating: a robust approach to classification of fluorescent events

To address these issues, I took advantage of the distribution of signal strength supplied from a fluorescent signal source. Inspection of Figure 23 shows that variation in genuine signal strength follows a Gaussian distribution; weak signal events represent autofluorescent particles or signal from other fluorophores that has remained detectable after compensation. To estimate the relative abundance of different fluorophores, I estimated the mean and variance of these distributions (Table 23); this was accomplished by minimizing the sum of squared deviations for a Gaussian distribution, assessing fit across a standard window. For example, in Figure 23A, I optimized the curve fit between μ-0.9σ and μ+0.9σ. Because the mean and deviation of the curves can be derived using Ferdinand, Z-scoring thresholds can be appropriately set to permit rigorous examination of data, addressing the issues of data classification using bivariate gating discussed earlier in Chapter 4.1.3 (see Figure 17 for an example).

Table 23. Mean and deviations of curves fit to Gaussian distributions of fluorescent signals from Figure 23

Peak	Mean	Deviation
EGFP	2.6959	0.1917
EYFP	2.6125	0.2230
DsRed-Express2	2.4369	0.1862
mKalama1	2.4128	0.3394
ECFP	1.7359	0.2749

Establishment of curve fits to Gaussian distributions of fluorescent class events were used to establish Z-scoring thresholds used to classify genuine fluorescent events from noise. To identify appropriate minimum and maximum Z-scoring thresholds for data classification, I altered both the minimum (under constant maximum Z-score) and the maximum (under constant minimum Z-score) Z-score thresholds used to seed the automated curve-fitting function in Ferdinand and examined the *p* values obtained using the χ^2 goodness-of-fit test along the Z-score explorations for a sample data set containing cells labeled with five different fluorescent proteins (Figure 24). The χ^2 goodness-of-fit test was used to determine the statistical significance between the observed number of data points and that predicted by the Gaussian distribution for each fluorescent class. Ideally, properly-fit curves having appropriate Z-scoring thresholds will have χ^2 *p* values indicating an insignificant difference between the observed number of data points and that predicted by the Gaussian distribution determined using Ferdinand. The fit between the Gaussian distribution and the underlying data was evaluated only between the minimum (Figure 24AB) and maximum Z-values (Figure 24C).

As shown in Figure 24A, minimum Z-score thresholds below -1.0 yielded curves having unacceptably low *p* values; goodness-of-fit χ^2 *p* values greater than 10^{-3} were considered appropriate due to the large sample size of data classes. Figure 24B depicts a zoomed-in inset of Figure 24A showing the region bounded by minimum Z-scores of -3.0 to -1.0 with *p* values ranging from 0.0 to 0.1. Here, the curve fit to the ECFP fluorescent class distribution did not have an insignificant *p* value until reaching a minimum Z-score threshold of -1.0. Additionally, I established an appropriate maximum Z-score threshold of 1.5 for curves fit to all fluorescent classes (Figure 24C). Particularly for curves fit to classes EGFP and ECFP, *p* values dropped into the significant range between maximum Z-score thresholds of +1.5 and +2.0, although other

curves remained more stable for maximum Z-scores up to +2.5. Thus, I chose Z-scoring thresholds of -1.0 to +1.5 to robustly classify events into proper fluorescent classes.

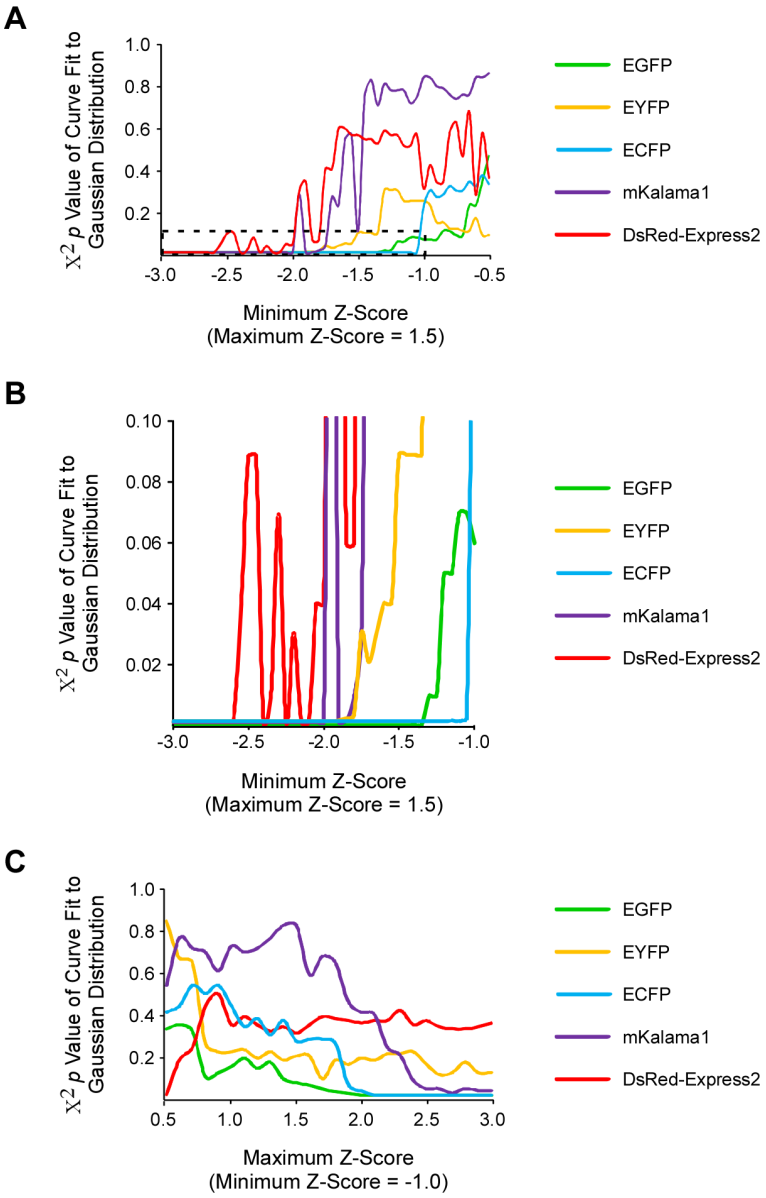

Figure 24. Variation of minimum and maximum Z-scores affects the goodness-of-fit of curves fit to Gaussian distributions of five fluorescent signal classes as measured by χ2 *p* values on a sample data set

Goodness-of fit χ^2 *p* values were obtained by altering both the minimum (under constant maximum Z-score) and the maximum (under constant minimum Z-score) Z-score thresholds used to seed the automated curve-fitting function in Ferdinand. **A**. χ^2 *p* values for curves varying the minimum Z-score threshold under constant maximum Z-score of 1.5 for a sample data set. Dashed line indicates area of zoom for Figure 24B. **B**. Inset of Figure 24A as indicated by dashed line. χ^2 *p* values for curves varying the minimum Z-score threshold under constant maximum Z-score of 1.5 for a sample data set. **C**. χ^2 *p* values for curves varying the maximum Z-score threshold under constant minimum Z-score of -1.0 for a sample data set.

Relative abundance of fluorescence classes as assessed using Z-gating thresholds remained stable (Figure 25AB). Here, ratios of each fluorescent class were determined varying the maximum Z-score while holding the minimum Z-score constant (Figure 25A) and varying the minimum Z-score while holding the maximum Z-score constant (Figure 25B). In both cases, changes in Z-gating thresholds did not substantially change the relative abundance of fluorescent classes. Only when Z-threshold lie at the extreme tails of the Gaussian distribution (less than -3.0) does noise begin affect event classification. However, the DsRed-Express2 and mKalama1 fluorescent classes were found in relatively low abundance in the data presented in Figure 25. For fluorescent classes containing low ratios of counts, it is extremely critical to choose proper Z-scoring thresholds that provide the most robust classification of data to identify the genuine hierarchy of fluorescent class events. Improper classification of noise events into these fluorescent classes may disproportionately impact classes with low counts in experiments. Thus, the minimum and maximum Z-score thresholds as determined in Figure 24 provide the best parameters for rigorous data classification.

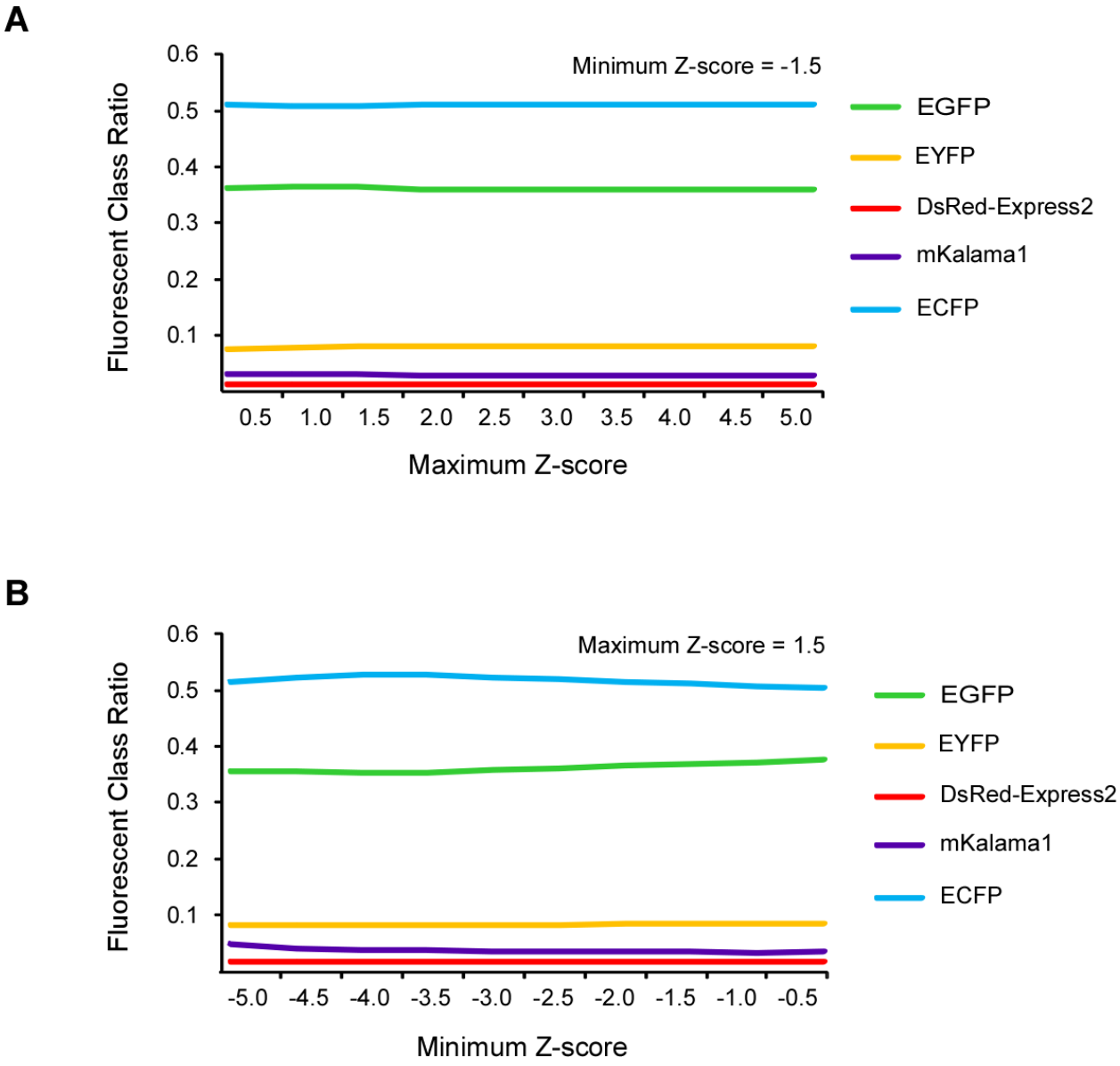

Figure 25. Z-scoring values do not substantially impact measurement of fluorescent class ratios for a sample five-color mixture of cells

A. Fluorescent class ratio as a function of maximum Z-score when minimum Z-score is held constant to a value of - 1.5. **B**. Fluorescent class ratio as a function of minimum Z-score when maximum Z-score is held constant to a value of +1.5.

After optimal fits were acquired, I determined the Z-value for each event's signal on each detector and used this information to classify events on a sample five-fluorescent class data set (Table 24). Thus the problem of an arbitrary threshold has been eliminated by expressing the signal on each detector as a function of the population of positive events themselves. For cells bearing single fluorophores, it is classified by the strongest Z-value, assuming it exceeds a specified threshold. For cells bearing two fluorophores, it is classified by the two strongest Z-scores. A single-color example is shown in Table 24; here, threshold methods failed to classify events that are easily assigned classes by the Z-gating approach (Table 22).

Table 24. Z-scoring methods resolve classification ambiguities encountered using traditional threshold scoring

The Z-scoring approach was used to score classes of data represented in Figure 23; miscalls of double and triple-color events were sorted into appropriate single color classes.

Class	Total	Replicate					
		1	2	3	4	5	6
Unclassified	136045	30319	25420	23245	23632	19715	13714
G----	54936	5054	10435	8297	5912	9082	16156
-Y---	12266	2040	2151	1876	1773	1763	2663
--R--	2009	255	270	279	271	307	627
---K-	4517	367	439	627	510	751	1823
----C	75264	10407	13165	12174	8842	12035	18641

While this approach is robust, it depends on proper fitting of the distribution of signal strength to a Gaussian distribution. For this to be accomplished, the data must satisfy several criteria. First, a large number of true positive events must be collected. Second, compensation must be accurate so that spillover does not occlude the distribution of true signal. Third, compensation matrices for doubly-labeled cells must be balanced so that the distribution of specific signals from different doubly-tagged cells is congruent. That is, compensation of EGFP signal from EGFP+ECFP and EGFP+EYFP doubly labeled cells must be balanced so that the mean and variance of these distributions is identical; otherwise, an accurate fit to the Gaussian distribution cannot be accomplished.

4.5.4 Testing the Z-gating approach

I tested this approach on mixtures of singly-labeled cells. Series of mixes were created which lacked particular fluorophores. These mixtures were analyzed and the events were sorted into fluorescence classes by traditional threshold means and by the Z-gating approach. As seen in Table 25, the Z-gating approach yielded very clean data, with very low numbers events being improperly placed into fluorescent classes which were not included in the experiment.

Table 25. Z-scoring methods appropriately classify events in single-color dropout mixes

Single fluorescent tagged *Salmonella* cells were grown in overnight liquid cultures and mixed in approximately equal ratios to create a mixture of all five colors and mixtures in which one color was missing. Classifications were determined using the Z-scoring approach in Ferdinand.

	Strain Mix Counts					
	All colors	**No EGFP**	**No EYFP**	**No ECFP**	**No mKalama1**	**No DsRed-Express2**
EGFP	38950	1	12891	31028	18707	32771
EYFP	7821	11612	2	6498	3492	6809
ECFP	7574	9572	2475	0	2385	5578
mKalama1	33289	41584	14561	23004	2	24143
DsRed-Express2	16724	22686	8307	12766	6284	0

5.0 O-ANTIGEN CHAIN LENGTH INFLUENCES *SALMONELLA* FITNESS AGAINST PROTOZOAN PREDATION

As discussed above, previous work established that amoebae could discriminate among different serovars of *Salmonella* [387]; predators are differentially distributed in the environment [386-387]; and predators in any given environment shared feeding preferences [386]. In addition, genetic variability at the *Salmonella rfb* locus could not be explained by conventional immune selection models [43]. These data establish the plausibility of amoebae predators as mediators of genetic diversity at the *Salmonella rfb* locus, but they fall short of attributing discrimination of *Salmonella* by amoebae to the O-antigen. In this chapter, I employ the techniques detailed in Chapter 4 to directly test if fitness against protozoan predation is influenced by O-antigen identity.

Previously, significant limitations in genetic manipulation techniques prevented the exploration of the full scope of antigenic diversity within *S. enterica* Subspecies I. Thus, fitness against protozoan predation was only assessed for three *rfb* near-isogenic strains, effectively representing two different O-antigen serotypes, against two genetically distinct amoebae [387]. Although this work showed fitness differences for these near-isogenic strains against predation [387], the small size of the data set only enables the conclusion that prey discrimination can be affected by the O-antigen. Moreover, these experiments did not control for the presence or absence of the H-antigen. Because the H-antigen is a major *Salmonella* cell surface antigen, I

cannot discount the possibility that it may influence fitness against predation. Based on this information alone, I was unable to definitively conclude that the O-antigen is the major *Salmonella* cell surface antigen impacting fitness against protozoan predation.

The crude nature of assays used previously to determine fitness against predation was also a significant limitation to a more comprehensive test of the protozoan predation hypothesis. Relative fitness hierarchies of *Salmonella* strains against predation had to be inferred from a large series of pairwise competition tests, as the methodology for assessing fitness only permitted simultaneous competition of two strains [387]. A more rigorous relative fitness hierarchy of strains can be established by performing simultaneous competitions of all strains in a given set in the presence and absence of amoebae predators. This approach requires gathering very large sets of data, given the breadth of antigenic diversity within *Salmonella*; line tests as performed previously and adapted for data collection as described in Chapter 3 are impractical for this purpose. Moreover, the need to compete strains at the same time requires a much more sophisticated approach to the tagging and discrimination of strains.

I addressed the technical limitations of previous work by developing a genetic system for the manipulation of natural isolates of *Salmonella* described in Chapter 2, using this approach to construct the necessary suite of strains to establish that the O-antigen is a major influence on *Salmonella* fitness against protozoan predation as discussed in Chapter 3. To validate the hypothesis that the *Salmonella rfb* locus is under diversifying selection as mediated by the differential feeding preferences of grazing amoebae predators, I need to demonstrate that fitness changes on a two-fold gradient based on the identity of the O-antigen. However, using line tests to measure fitness are insufficient to address this issue; thus, I developed a new fitness assay

using flow cytometry as described in Chapter 4 to yield the more powerful data sets required to test the protozoan predation hypothesis.

Here, I combine the technical advances described in Chapters 2 and 4 to further investigate the hypothesis that protozoan predation is driving diversifying selection on the *Salmonella rfb* locus. If this is true, then the O-antigen should be the major antigen contributing to *Salmonella* fitness against predation. Consequently, if the O-antigen is the major surface antigen affecting *Salmonella* fitness against predation, then strains that only vary in terms of O-antigen chain length should be discriminated by protozoan predators. Thus, I used several of the *rfb* near-isogenic strains constructed in Chapter 3 to derive a collection of strains that only vary in terms of O-antigen chain length. Wild-type strains have a bimodal distribution of O-antigen chain lengths of 16-35 repeats and 100+ repeats [231-232]. Strains mutated at *fepE* display O-antigen chain length classes of 16-35 repeats [231], while those mutated at *wzzB* display O-antigen chain length classes of 1-16 repeats and 100+ repeats (potential bias for long O-antigens is due to variable expression of *fepE*) [17, 231]. Strains lacking functional forms of both the *fepE* and *wzzB* genes have O-antigens of 1-16 repeats [17, 231]. Mutation of the *rfc* gene results in the production of O-antigens having only one polysaccharide monomer [17]. Additional discussion on O-antigen biosynthesis and chain length is found in Chapters 1 and 2 and illustrated in Figure 9.

If the O-antigen is a major contributor to fitness against predation, then the amount of O-antigen present on the outside surface of a cell must influence fitness against predation. Thus, I should observe differences in fitness against predation among strains that only vary in O-antigen chain length.

5.1 MATERIALS AND METHODS

5.1.1 Media and growth conditions

Bacterial strains were routinely propagated overnight at 37°C in LB broth and on agar media. Media was supplemented when necessary with ampicillin (200 μg/mL); chloramphenicol at 20 μg/mL and at 10 μg/mL when used with ampicillin; kanamycin at 20 μg/mL and at 10 μg/mL when used with ampicillin; and tetracycline at 10 μg/mL and at 5 μg/mL when used with ampicillin. Amoebae used were isolated by Wildschutte and Lawrence [386] and were propagated according to the protocols discussed in Chapter 3.2.2. Predation competition experiments were conducted using NM-C media, described in Chapter 4.4.2.

5.1.2 Strains

Directed gene knockouts of genes involved in O-antigen polymerization (*rfc*) and regulation (*wzzB* and *fepE*) were performed as described in Chapter 2.2.5. Bacteriophage P1 was used to transduce these constructs into *galE* mutants of *rfb* isogenic strains (Chapters 2.2.3 and 3.2.3) (Table 26). Phage P1 was used to repair the *galE* mutation as described in Chapter 2.2.3. Strains were then made electrocompetent as described in Chapter 2.2.5 and transformed with plasmids pEGFP, pEYFP, pECFP, pDsRed-Express2, and pKAB9-mKalama1 (Chapter 4.2.1; Table 26).

Table 26. List of strains used to assess the contribution of O-antigen chain length to *Salmonella* fitness against predation

Table continued on next page.

Parent *rfb* Near-Isogenic Strain	O-serotype	O-antigen Chain Length Genotype	Predicted O-antigen Chain Length	Competition Series 1		Competition Series 2	
				Fluorescent Tag	Strain	Fluorescent Tag	Strain
KAB081 (SARB2)	(3,10)	Wild-type	16-35, 100+	EYFP	KAB800	ECFP	KAB801
		wzzB-8771::*aph*	1-16, 100+	mKalama1	KAB802	EYFP	KAB803
		fepE-2910::*tet*	16-35	ECFP	KAB804	DsRed-Express2	KAB805
		fepE-2910::*tet wzzB*-8771::*aph*	1-16	DsRed-Express2	KAB806	EGFP	KAB807
		rfc-1727::*cat*	1	EGFP	KAB808	mKalama1	KAB809
KAB082 (SARB3)	(1,4,12)	Wild-type	16-35, 100+	EYFP	KAB810	ECFP	KAB811
		wzzB-8771::*aph*	1-16, 100+	mKalama1	KAB812	EYFP	KAB813
		fepE-2910::*tet*	16-35	ECFP	KAB814	DsRed-Express2	KAB815
		fepE-2910::*tet wzzB*-8771::*aph*	1-16	DsRed-Express2	KAB816	EGFP	KAB817
		rfc-1727::*cat*	1	EGFP	KAB818	mKalama1	KAB819

Table 26 continued. List of strains used to assess the contribution of O-antigen chain length to *Salmonella* fitness against predation. Table continued on next page.

Parent *rfb* Near-Isogenic Strain	O-serotype	O-antigen Chain Length Genotype	Predicted O-antigen Chain Length	Competition Series 1		Competition Series 2	
				Fluorescent Tag	Strain	Fluorescent Tag	Strain
KAB084 (SARB20)	(8,20)	Wild-type	16-35, 100+	EYFP	KAB820	ECFP	KAB821
		wzzB-8771::*aph*	1-16, 100+	mKalama1	KAB822	EYFP	KAB823
		fepE-2910::*tet*	16-35	ECFP	KAB824	DsRed-Express2	KAB825
		fepE-2910::*tet wzzB*-8771::*aph*	1-16	DsRed-Express2	KAB826	EGFP	KAB827
		rfc-1727::*cat*	1	EGFP	KAB828	mKalama1	KAB829
KAB086 (SARB36)	(6,8)	Wild-type	16-35, 100+	EYFP	KAB830	ECFP	KAB831
		wzzB-8771::*aph*	1-16, 100+	mKalama1	KAB832	DsRed-Express2	KAB833
		fepE-2910::*tet*	16-35	ECFP	KAB834	EYFP	KAB835
		fepE-2910::*tet wzzB*-8771::*aph*	1-16	DsRed-Express2	KAB836	EGFP	KAB837
		rfc-1727::*cat*	1	EGFP	KAB838	mKalama1	KAB839

Table 26 continued. List of strains used to assess the contribution of O-antigen chain length to *Salmonella* fitness against predation.

Parent *rfb* Near-Isogenic Strain	O-serotype	O-antigen Chain Length Genotype	Predicted O-antigen Chain Length	Competition Series 1		Competition Series 2	
				Fluorescent Tag	Strain	Fluorescent Tag	Strain
KAB087 (SARB52)	(1,9,12)	Wild-type	16-35, 100+	EYFP	KAB840	ECFP	KAB841
		wzzB-8771::*aph*	1-16, 100+	mKalama1	KAB842	EYFP	KAB843
		fepE-2910::*tet*	16-35	ECFP	KAB844	DsRed-Express2	KAB845
		fepE-2910::*tet* *wzzB*-8771::*aph*	1-16	DsRed-Express2	KAB846	EGFP	KAB847
		rfc-1727::*cat*	1	EGFP	KAB848	mKalama1	KAB849

5.1.3 Predation competition tests to measure fitness

Two sets of reciprocally tagged strains were grown overnight in liquid LB supplemented with ampicillin at 200 μg/mL. Cells were pelleted with centrifugation and resuspended in PBS; following resuspension, cells were mixed in approximately equal ratios based on OD_{600} measurements. OD_{600} measurements of final competition mixes ranged from 0.900 to 1.200; a 100 μL aliquot of the competition mixture was spread onto NM-C plates supplemented with ampicillin at 200 μg/mL. For PREDATOR plates, the centers of fourteen plates were seeded with 10^4 acid-base washed amoebae cysts; at least eight NO PREDATOR plates were set up for each experiment. Plates were incubated at 33°C for approximately 3.5 days to permit the feeding front of predation to reach the outer edges of the PREDATOR plates.

A total of six NO PREDATOR and six to ten PREDATOR plates from each experiment were chosen for analysis. Plates were eluted with 5 mL of PBS with 0.02% Tween20 added to reduce cell clumping. To diminish the number of amoebae present in samples, the eluate was filtered through two 5 μM pore size overlayed CellMicroSieve™ biologically inert nylon mesh filters (Bio-Design Inc. of New York). All samples were diluted in PBS + 0.02% Tween20 to an approximate concentration of 10^7 cells/mL.

5.1.4 Flow cytometry

Prior to analysis, SYTO™ 62 nucleic acid dye (Invitrogen, Carlsbad, CA) was added to a final concentration of 50 nM to aid in the separation of *Salmonella* cells from debris and amoebae. Events were acquired on a Beckman-Coulter CyAn ADP (Beckman-Coulter) flow cytometer at a

rate of <1000 events per second. Filters and detector settings are listed in Chapter 4.X. Commercially available software (Summit 4.3, Dako Colorado Inc.) was used for the operation of the cytometer. Data analysis was performed using Ferdinand (developed in Chapter 4).

5.1.5 Calculation of fitness values

Events from replicate plates were pooled for analyses; plates which provided data that were significantly different from other plates' data were discarded. Signal strength was compensated as described (Chapter 4) and Gaussian curves were fit to the distribution of compensated signal strength of each individual fluorophore (Chapter 4). Events were sorted into fluorescent cell classes by the Z-scoring algorithm, typically selecting events lying within 2 standard deviations of the mean compensated value for that fluorophore. The frequency of each cell class was calculated as the ratio of the observed number of cells in that class to the total number of cells in each fluorescent class; cells with signal strength outside the Z-scoring range were deemed unscored and ignored.

As discussed above, strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates):

$$w_S = \mathrm{P}_{\mathrm{observed}}/\mathrm{P}_{\mathrm{expected}}$$

The fitness of any bacterial strain against predation (w_S) is a function of both its genotype (w_G) and its fluorophore (w_F). The genotype affects fitness since different genotypes present different antigens; the fluorophore affects fitness since the longevity of the protein expression differs among fluorescent proteins; these differences affect one's ability to detect them via flow cytometry. Since these factors are independent, I may express strain fitness as:

$$w_S = w_G * w_F$$

Using logarithmic transformation, I can reduce a set of strain fitness values to a system of linear equations of the form:

$$\mathrm{Ln}(w_S) = \mathrm{Ln}(w_G) + \mathrm{Ln}(w_F)$$

For clarity, I expressed log-transformed fitness values as w':

$$w_S' = w_G' + w_F'$$

To calculate both w_G and w_F, I performed two sets of fitness calculations; the reciprocally-tagged set of strains paired each genotype with a different fluorescent protein. Thus, I can eliminate fluorophore fitness from this system of equation by subtraction, as:

$$w_{S1}' - w_{S2}' = (w_{G1}' + w_{F1}') - (w_{G2}' + w_{F1}') = w_{G1}' - w_{G2}'$$

where G1 and G2 represent the two different genotypes associated with fluorophore F1 in the two sets of strains. Thus the genotype fitness values can be expressed in terms of the observed strain fitness values as:

$$\Delta\mathbf{W} = \mathbf{M}*\mathbf{W_G}$$

where $\Delta\mathbf{W}$ is the matrix of differences in strain fitness, $\mathbf{W_G}$ is the matrix of genotype fitness values, and M is the matrix of coefficients noting the genotypes. For example:

$$\begin{bmatrix} w_{S1}' - w_{S6}' \\ w_{S2}' - w_{S7}' \\ w_{S3}' - w_{S8}' \\ w_{S4}' - w_{S9}' \\ w_{S6}' - w_{S10}' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ -1 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_{G1}' \\ w_{G2}' \\ w_{G3}' \\ w_{G4}' \\ w_{G6}' \end{bmatrix}$$

Thus, like the deconvolution of actual signal during compensation, I could solve for genotype fitness values (w_G') by matrix inversion:

$$\mathbf{M}^{-1}*\Delta\mathbf{W} = \mathbf{M}^{-1}\mathbf{M}*\mathbf{W_G} = \mathbf{W_G}$$

Two modifications made to this system of linear equation to allow deconvolution. First, the coefficient matrix **M** cannot be inverted since the array of linear equations does not have a unique solution. To circumvent this problem, I arbitrarily assign a fitness value of 1.0 (w_G'=0) to one of the genotypes, thus creating a unique solution. Because fitness values are ultimately normalized, the results are identical regardless of which w_G' value is initially set to zero. Second, even with this substitution, the equations often do not converge in N-dimensional space due to measurement error. That is, the measured strain fitness is better described as:

$$w_S' = w' * w_F' + \varepsilon$$

where ε is measurement error. I assume equal error across all measurements and deconvolute genotype fitness values accordingly.

To accommodate measurement error, this system of equations is solved by stepwise substitution. I may then solve for fluorophore fitness in each set of strains using their respective strain fitness (w_S') and genotype fitness (w_G') values. Final fluorophore fitness values are determined as the arithmetic mean of the two values. Final fitness values are determined by the reverse of the log transformation.

Note that this method of calculating fitness differs from that used in other population genetic studies employing the simultaneous competition of strains in the same environment, such as a chemostat [72-73, 79, 344]. In these experiments, relative fitness of a genotype or phenotype is measured as a selection coefficient (s) reflecting the degree to which the fitness of one genotype/phenotype is less than that of a favored genotype/phenotype, where s ranges in value from 0 to 1, with s = 0 representing a relatively neutral genotype/phenotype ; s = -1 representing a beneficial genotype/phenotype and s = 1 representing complete lethality. Critically, s reflects a growth rate difference, as strains are competed; for example, the fitness of a mutation may be

evaluated by competing mutant and wild-type strains. The nature of these experiments prevents the direct measurement of fitness as *s*, as strains cannot be simultaneously competed in the presence and absence of protozoan predators. Thus, relative fitness must be calculated as differential susceptibility to predation.

This approach to measuring fitness permits the deconvolution of genotype fitness from fluorescent tag fitness, although I cannot solve for any epistatic interaction between genotype and fluorescent tag. However, I do not have any biological evidence to suggest that epistatic interactions exist in my experimental design; there is no reason to suggest that the only factor that differs among strains, O-antigen chain length, will influence expression and folding of fluorescent proteins. Additionally, I must solve for fluorescent tag fitness in every experiment through the use of simultaneously performed reciprocal tag experiments rather than perform data corrections based on universal fluorescent fitness determinations from completion tests on strains that only vary in terms of fluorescent tag. In Chapter 4, I determined that environmental conditions such as cell density and nutrient concentration influence the signal from fluorescent proteins. Because I expect these and other environmental conditions to vary between experiments, I do not believe that applying a single fluorescent tag fitness correction is a fair analysis for the data collected here.

5.1.6 Technical acknowledgements

I designed all primers used to create and verify directed gene replacements of all strains listed in this chapter as well as engineer the *rfb* near-isogenic strain collection. Under my supervision, undergraduate researchers Ben Cross and Mark Brown constructed strains KAB800-849.

Ferdinand flow cytometry data analysis software was developed by Dr. Jeffrey Lawrence at the University of Pittsburgh.

5.2 THE ROLE OF O-ANTIGEN CHAIN LENGTH IN SUSCEPTIBILITY TO PROTOZOAN PREDATION

To test the hypothesis that amoebae recognize bacterial prey through interaction with their O-antigens, I engineered a series of *rfb* near-isogenic strains of *Salmonella* that only varied in terms of O-antigen chain length (Table 26). As discussed in Chapter 2, wild-type strains have bimodal classes of O-antigen chain lengths, and mutant derivatives typically have much shorter chains. I predict that altering the chain length of the O-antigen will influence *Salmonella* fitness against predation. Critically, the nature of this influence should depend on the identity of the O-antigen monomers and the predator being examined. Several models can be used to explain the relationship between protozoan feeding preference and O-antigen identity.

First, feeding preference may simply depend on how well the predator is able to bind the O-antigen, with most preferred strains representing O-antigens that are tightly bound by predators and least preferred strains representing more weakly bound O-antigens. In this case, shortening the length of the O-antigen should provide less substrate for amoebae binding, and thus I would expect universally shortening the amount of O-antigen present on the outside of the cell independent of antigen identity should increase fitness against predation. Second, the O-antigen may be used by predators to actively indicate a food source; in this case, most preferred strains display O-antigens that predators recognize as "food" while least preferred strains display O-antigens that predators are less able to recognize as "food." If this model is true, then I would

expect that shortening O-antigen chain length of highly preferred strains would increase fitness. Shortening the O-antigen chain length of lesser preferred strains may slightly improve or have no significant impact on fitness against predation depending on the nature of the relationship between amoebae and the O-antigen. Third, certain O-antigens may be actively avoided by predators; in this case, least preferred strains may possess O-antigens that amoebae avoid or recognize as "non-food." If this model is true, then I would expect that shortening the O-antigen chain length of least preferred strains would actually decrease fitness against predation, as the cell would essentially lose the benefit of displaying an O-antigen that confers protection against predation. Conversely, any fitness effect of shortening the chain length of most preferred strains should be more complex depending on the nature of the predator-O-antigen relationship. It is important to note here that I did not expect one simple model to completely explain the relationship between amoebae and the *Salmonella* O-antigen. These experiments are intended to further test the hypothesis that selective pressure from protozoan predation drives diversifying selection at the *Salmonella rfb* locus as well as to further investigate the mechanism of how protozoan predators interact with the *Salmonella* O-antigen.

To test these predictions, the series of *rfb* near-isogenic strains of *Salmonella* that only varied in terms of O-antigen chain length was challenged with the amoebae *Naegleria gruberi* NL, *Acanthamoeba* sp. R2-1, and *Tetramitus* sp. BD1-1. Competition experiments were performed by mixing fluorescently-tagged derivatives of the series of chain-length mutants and growing them in the presence or absence of predators. Relative cell abundance was measured by flow cytometry as described in Chapter 4.

The results of one experiment are shown in Figure 26. Here, mutant derivatives of *Salmonella enterica* Typhimurium LT2 which a) bear the O-antigen region from SARB3 (1,4,12)

and b) have a series of mutations altering the length of this antigen are competed in the presence and absence of the predator *Naegleria gruberi* NL. Relative fitness of the strains is calculated by changes in relative abundance between no-predator and predator plates. The relative fitness among the strains varied significantly; when normalizing to a fitness of 1.0, the least-preferred strain was consumed approximately three-fold faster than the most-preferred strain, yielding a normalized strain fitness (w_S) of 0.328. Additionally, error for fitness measurements were considerably low, with error rates typically less than 5.0% (Table 27); indicating the robustness of fitness.

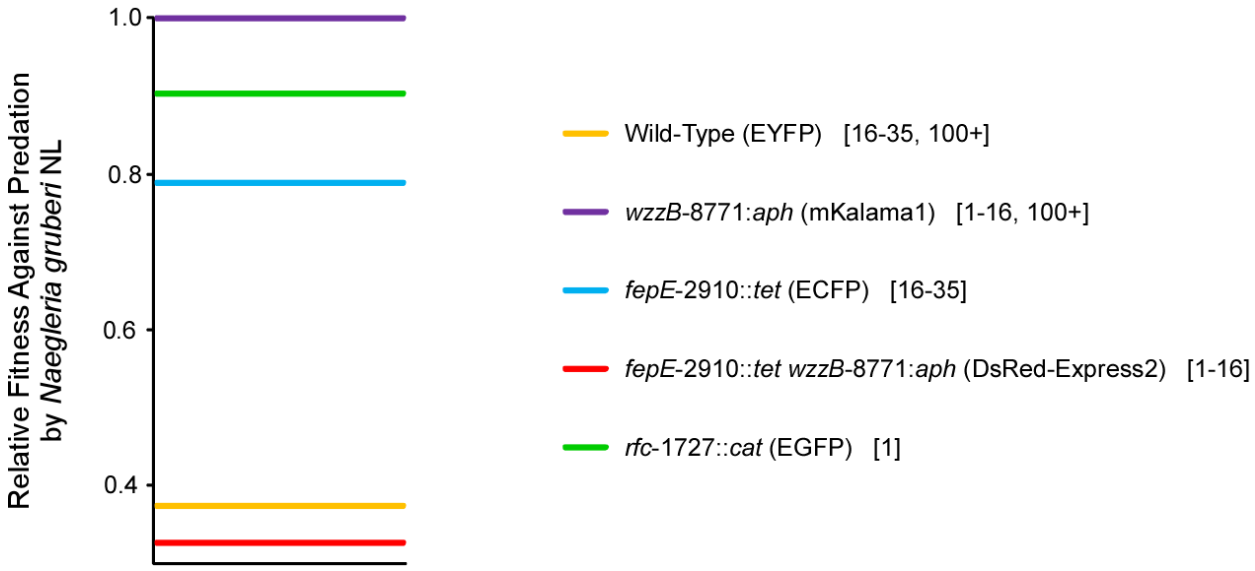

Figure 26. Relative strain fitness of O-antigen chain length mutant derivatives of the *rfb* near-isogenic strain KAB082 against predation by the amoeba *Naegleria gruberi* NL

O-antigen chain length derivatives of the *rfb* near-isogenic strain KAB082 containing the *rfb* region from SARB3 [O-serotype (1,4,12)] were fluorescently tagged and competed in the presence and absence of predation by the laboratory strain of the amoeba *Naegleria gruberi* NL. Fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). The strain having the highest raw fitness was assigned a value of 1; the fitness values of all other strains were normalized to this value. Fluorescent tag identity is shown in parentheses. Predicted O-antigen chain length is depicted in brackets.

Table 27. Raw strain fitness values with error for fluorescently-tagged O-antigen chain-length derivatives of the *rfb* near-isogenic strain KAB082 (SARB3) against predation by *Naegleria gruberi* NL

O-antigen chain length derivatives of the *rfb* near-isogenic strain KAB082 containing the *rfb* region from SARB3 [O-serotype (1,4,12)] were fluorescently tagged and competed in the presence and absence of predation by the laboratory strain of the amoeba *Naegleria gruberi* NL. Fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Raw strain fitness and error were determined using bootstrapping with 1000 iterations.

Strain	Genotype	Tag	Predicted O-antigen Chain Length	Raw Strain Fitness	Error	Percent Error
KAB800	Wild type	EYFP	16-35, 100+	0.6813	0.0162	2.38
KAB802	*wzzB*-8771::*aph*	mKalama1	1-16, 100+	1.8173	0.0537	2.95
KAB804	*fepE*-2910::*tet*	ECFP	16-35	1.4339	0.0391	2.72
KAB806	*fepE*-2910::*tet* *wzzB*-8771::*aph*	DsRed-Express2	1-16	0.5966	0.0145	2.43
KAB808	*rfc*-1727::*cat*	EGFP	1	1.6433	0.0457	2.78

5.2.1 Experimental variability did not explain raw strain fitness differences

Fitness estimates require precise estimates of cell numbers, and flow cytometry offers only indirect measures of relative cell abundance. To address possible experimental sources of error, I assessed the robustness of my measures.

5.2.1.1 Gaussian distributions were appropriately fit

The distribution of signal strength was fit to a Gaussian distribution; this allowed for the separation of signal from noise and enabled precise counting of positive events. To determine if these fits were robust, I measured the sum of squared deviations (SSD) between the observed number of data points and that predicted by the Gaussian distribution for each fluorescent class. For example, Figure 27 depicts several improperly-fit curves along with the properly-fit curve (depicted in black) to the Gaussian distribution for EGFP fluorescent signal from a sample set of plates. For the majority of curves, the SSD values were less than 1.0, indicating optimal fits of the curves to actual data (Table 28). Additionally, the χ^2 goodness-of-fit was used to determine the statistical significance between the observed number of data points and that predicted by the Gaussian distribution for each fluorescent class (Table 28). Here, I consider any *p* value at the 10^{-3} level and above to indicate properly fit curves due to the large sample size of some fluorescent classes. No significant skew or kurtosis was observed within the Z-scoring range (apparent skew is attributable to residual spillover and autofluorescence). Thus, differences in strain fitness cannot be explained by improper classification resulting from poor curve fitting.

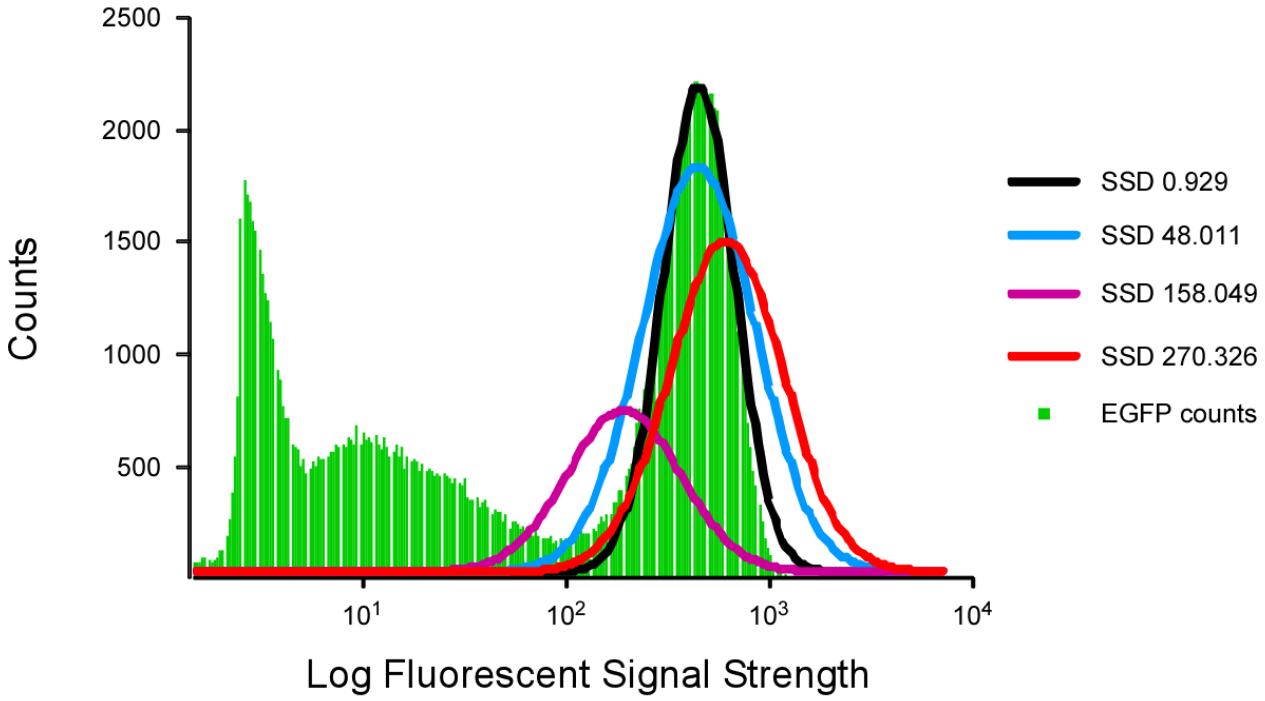

Figure 27. Sum of squared deviations for curves fit to Gaussian distribution of EGFP signal from a sample experiment set

A series of Gaussian curves were properly fit (shown in black) and improperly fit (blue, purple, and red curves) to a distribution of EGFP signal in Ferdinand. Improperly fit curves have sum of squared deviations (SSD) values higher than 1.0.

Table 28. Measuring robustness of curves fit to Gaussian distributions for sample data sets with and without predators

The sum of squared deviations (SSD) was measured between the observed number of data points in a given Ferdinand file and that predicted by the Gaussian distribution fit to each fluorescent class. In most cases, curves were fit to obtain SSD values less than 1.0. The corresponding χ^2 values are also reported; in almost all cases, data were not significantly different from normal.

Gaussian Distribution	Experimental Condition			
	No Predator		Predator	
	χ^2	P value	χ^2	P value
EGFP	43.479	0.003	26.292	0.29
EYFP	81.743	0.003	24.683	0.59
CFP	17.969	0.30	5.212	0.92
mKalama1	51.026	0.23	13.404	0.94
DsRed-Express2	62.444	0.06	27.978	0.79

5.2.1.2 Z-scores used to assess fitness were appropriately established using fit of Gaussian distributions

Because the curves fit to the Gaussian distribution of fluorescent signal classes were robust, I used curve-fitting properties to determine the Z-scoring thresholds that permitted separation of genuine fluorescent signal from noise. In this way, I was able to set scoring thresholds that enabled precise counting of positive events. Here, choice of appropriate Z-scoring thresholds in curve measurements are critical for excluding noise events, including those arising from autofluorescence or spill-over from other fluorescent signals, from events arising from the specific fluorophores that defined each Gaussian distribution. Fluorescent signal distributions typically varied in both signal strength and in proximity of genuine signal to noise events in both no predator and predator experimental conditions. Therefore, certain fluorescent signal classes are more susceptible to inappropriate inclusion of noise events than others, depending on the nature of the fluorophore and experimental condition. In these experiments, it is critical that noise events are not classified as genuine fluorescent signal in order to accurately determine ratios of fluorescent events on each plate within each data set, as inclusion of noise events would have resulted in improper fitness determinations.

To identify appropriate minimum and maximum Z-scoring thresholds for data classification, I altered both the minimum (under constant maximum Z-score) and the maximum (under constant minimum Z-score) Z-score thresholds used to seed the automated curve-fitting function in Ferdinand and examined the *p* values obtained using the χ^2 goodness-of-fit test along the Z-score explorations (Figure 28). Here, the fit between the Gaussian distribution and the underlying data was evaluated only between the minimum and maximum Z-values. Goodness-of-fit χ^2 *p* values greater than 10^{-3} were considered appropriate due to the large sample size of

most fluorescent classes. Significant *p* values indicated that the Gaussian distribution was not a good fit to the observed data. I expected to reject Gaussian distributions when the range over which they are being evaluated included significant numbers of noise events. As shown in Figure 28A, curves fit for ECFP below a minimum Z-score of -1.0 on no predator plates showed inappropriately low *p* values, which indicated curves fit below this value were significantly different than predicted. Therefore, accurate proportions of fluorescent events would include excessive noise events for ECFP if those events having Z-scores lower than -1.0 were included. It is not surprising that fits to the observed ECFP signal reached its noise fluorophore threshold first, as ECFP signal is the weakest among those examined in these experiments (see Chapter 4.3 and Figure 19 for additional discussion). Additionally, curves fit above a maximum Z-score threshold of 1.5 also yielded unacceptably low *p* values (Figure 28B); above that point, the EGFP curve failed to conform to the observed data. Similar results were observed using predator plates, with a minimum Z-score of -1.0 (Figure 28C) and maximum Z-score of +1.5 (Figure 28D) yielding *p* values above 10^{-3}. Although a general trend was observed towards increasing *p* values towards higher minimum Z-scores and lower maximum Z-scores, I chose Z-score cutoff values of -1.0 to +1.5 to define the region of all Gaussian distributions containing positive signal. Here, using inappropriately high minimum Z-scores and low maximum Z-scores would have resulted in decreasing overall numbers of events, thus increasing stochasticity in fluorescent class ratio determination. Thus, although this approach did not count all positive events of each fluorescent class, it defined a region of all Gaussian distributions which minimized the probability of including noise events from each fluorescent class while maximizing the numbers of positively-scored events to avoid excessive stochasticity.

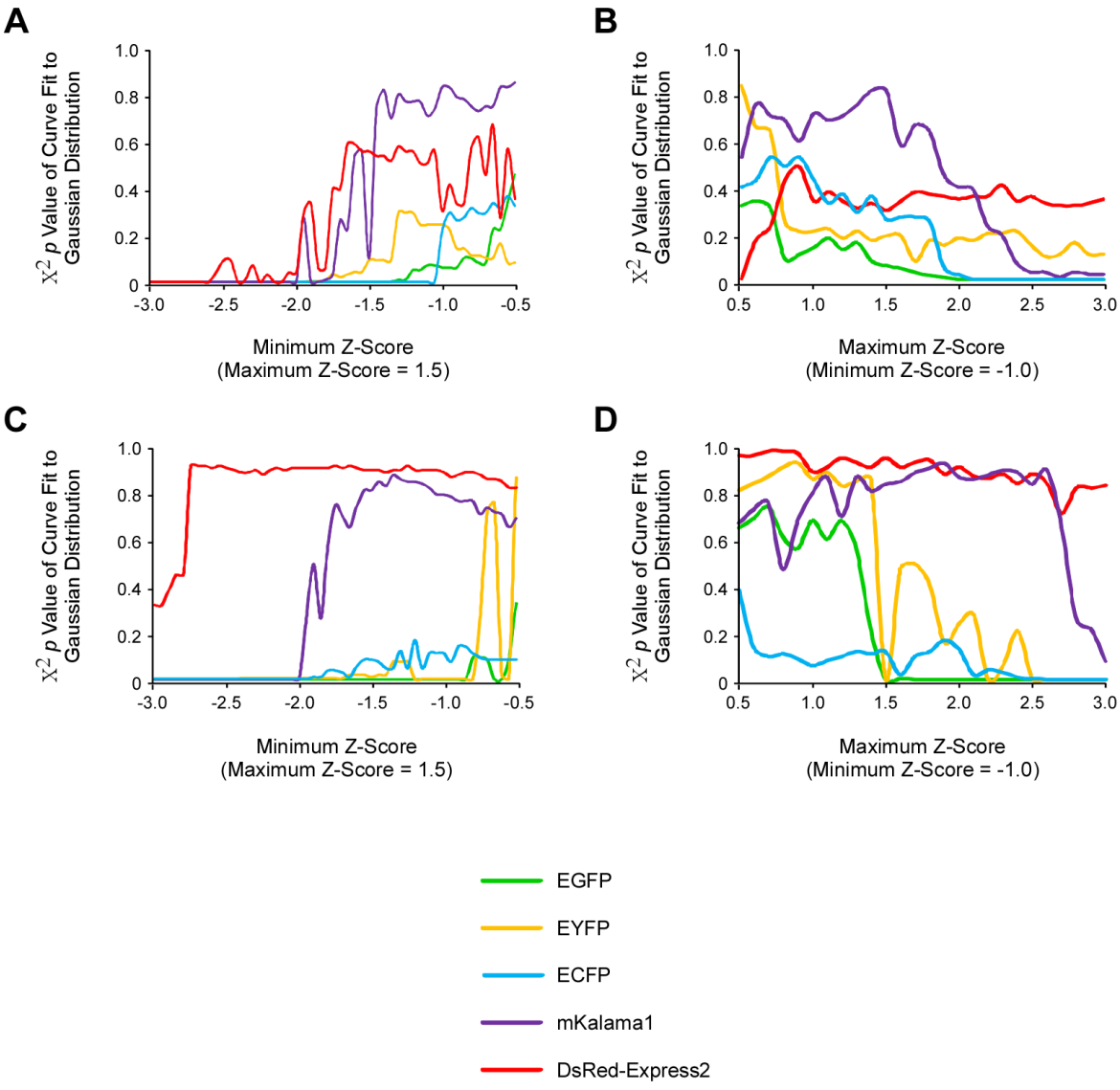

Figure 28. Optimal curve fitting is defined by setting appropriate Z-score thresholds

Variation of minimum and maximum Z-score thresholds affects the goodness-of-fit of curves fit to Gaussian distributions of five fluorescent signal classes as measured by χ^2 *p* values for sample no predator and predator data sets. Goodness-of fit χ^2 *p* values were obtained by altering both the minimum (under constant maximum Z-score) and the maximum (under constant minimum Z-score) Z-score thresholds used to seed the automated curve-fitting function in Ferdinand. **A**. χ^2 *p* values for curves varying the minimum Z-score threshold under constant maximum Z-score of 1.5 for a sample no predator plate data set. **B**. χ^2 *p* values for curves varying the maximum Z-score threshold under constant minimum Z-score of -1.0 for a sample no predator plate data set. **C**. χ^2 *p* values for curves varying the minimum Z-score threshold under constant maximum Z-score of 1.5 for a sample predator plate data set. **D**. χ^2 *p* values for curves varying the maximum Z-score threshold under constant minimum Z-score of -1.0 for a sample predator plate data set.

5.2.1.3 Within-experiment variability was low in no predator data sets

Experimental replication was achieved by collecting data from 6 – 10 replicate plates for each experimental class (plates with or without predators). To examine variability between replicates, I calculated the sum of squared deviations (SSD) of the observed abundance of each fluorescent class to that expected given the mean proportion of each fluorescent class for a given experiment. For competition tests performed in the absence of amoebae predators, SSD values were very low (Table 29), indicating a robust level of within-experiment replication for this set. Significance was assessed by χ^2 tests with N-1 degrees of freedom where N is the number of data points being fit.

Table 29. Goodness-of-fit tests on a sample set of replicate no predator plates

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. In the majority of cases, SSD values were less than 0.050, indicating low levels of within-experiment variability among plates without protozoan predators.

Replicate	1	2	3	4	5	6
1	0.000	0.010	0.001	0.002	0.002	0.006
2	---	0.000	0.006	0.005	0.003	0.001
3	---	---	0.000	0.001	0.001	0.002
4	---	---	---	0.000	0.001	0.002
5	---	---	---	---	0.000	0.001
6	---	---	---	---	---	0.000

5.2.1.4 Within-experiment variability in predator data sets is acceptable but varies among replicates

For competition tests performed in the presence of amoebae predators, SSD values were also acceptably low, albeit higher than those observed on completion plates lacking protozoan predators (Table 30). Plates having significantly aberrant SSD values from other plates in a given experiment were commonly observed for plates containing predators, as shown in Table 30, but were very rarely observed for plates lacking predators, as shown in Table 29. For no predator plates, within-experiment variability most likely arises from environmental conditions that differentially impact the growth of bacterial strains among plates and experimental handling and processing procedures. In this case, low within-experiment variability indicates that these effects are minimal for plates lacking predators. However, a higher degree of within-experiment variability among predator plates was frequently observed in these experiments. In this case, I attribute this observation to varying rates of predation among plates containing amoebae predators.

Table 30. Goodness-of-fit tests on a sample set of replicate predator plates

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. Shaded regions indicate unacceptably high SSD values; this replicate was eliminated from the data set. Generally, plates having a majority of SSD values above 0.05 were discarded from data sets; removal of these aberrant plates tended to improve overall SSD values for other plates as atypical data were removed from population summaries.

Replicate	1	2	3	4	5	6
1	0.000	0.003	0.009	0.028	0.020	0.184
2	---	0.000	0.003	0.042	0.015	0.218
3	---	---	0.000	0.051	0.015	0.244
4	---	---	---	0.000	0.091	0.073
5	---	---	---	---	0.000	0.314
6	---	---	---	---	---	0.000

It is impossible to ensure identical rates of predation on plates within an experiment, because any environmental condition that can locally vary among plates, such as humidity, nutrient availability and bacterial cell density, impacts the rate of predation. As amoebae predators consume the lawn of bacteria outward from the center of the plate, the ratio of cells present on the entire plate change. Thus, I expect the ratio of cells recovered from whole plates to differ as predation occurs over time, with plates containing more uneaten bacteria due to slower rates of predation reflect the "expected" ratio of cells rather than the "observed" ratio of cells. Because environmental conditions vary among plates, I do not expect the rate of predation to be identical for all plates in a given experiment.

Ideally, differences in the rates of predation among plates could be mitigated by incubating all plates in a given experiment for a sufficiently long enough time to ensure that each plate has been completely swept by the protozoan predators. Under the environmental conditions used in these experiments, I typically observed completion of predation to vary between 60 and 96 hours of incubation, with predation on most plates finishing between 72 and 84 hours of incubation. However, I encountered an experimental trade-off between permitting predators to fully sweep the lawn of bacteria and obtaining sufficiently strong fluorescent signal from tagged bacterial cells. As discussed in Chapter 4, a trade-off exists in growth conditions that permit maximum fluorescent signal and those that favor development of uniform bacterial cell lawns; thus I optimized growth conditions that best reconciled these two variables at the expense of each other. Not surprisingly, a similar trade-off exists between maximum fluorescent signal and incubation time.

Although large numbers of bacterial cells are present on plates post-exposure to predation, these cells cannot be counted unless they emit fluorescent signals within the range of

detection of the flow cytometer. As discussed in Chapter 4, levels of excitation and emission for each fluorescent protein used in the multicolor tagging scheme vary considerably; thus, it is critical that experimental conditions favor fluorescent detection in order to maximize the probability of detecting all fluorescent signals. In practice, the ability to detect fluorescent signals decayed over increasing incubation times; plates incubated for 96 hours tended to produce very few detectable fluorescent counts relative to plates incubated for less time. Therefore, I chose an incubation time of 72-84 hours as a compromise between being able to collect sufficiently bright enough fluorescent signals to enable maximal detection of bacterial cells while allowing predation to complete on the majority of plates in any given data set. Additionally, I chose to expand my sample of predator replicate plates relative to no predator replicate plates to allow for the collection of and subsequent removal of plates with highly aberrant SSD values.

Higher between-replicate variability is expected as the predator population on some plates may not have finished reducing the prey population to sufficiently low numbers to induce predator encystation; if true, the post-predation populations of prey cells would coexist with pre-predation populations, altering the relative abundance of the fluorescent classes. If protozoan predators have not completely swept a plate of bacterial cells, then I do not expect the ratios of fluorescent classes on that plate to be representative of a plate fully exposed to predation. Thus, these plates cannot be included in the final data analysis. When plates with such highly-aberrant sets of cell abundances were detected, they were discarded from the analysis. Although increasing the environmental complexity of plates via addition of amoebae increases the variability of within-experiment replicate samples, I find these SSD values indicate that total within-experiment variation alone cannot account for the differences in strain fitness I observed

for the O-antigen chain length mutant collection (see below). Additionally, effects of high within-experiment variability are further discussed in Chapter 5.3.

5.2.1.5 Fitness assessment using Z-scoring is robust

Relative cell abundances were determined using the Z-scoring approach (Chapter 4), which effectively eliminates both noise and variability introduced by the arbitrary placement of a signal threshold between fluorescent and non-fluorescent events. For any one experiment, variability in cell abundances as a function of Z-value threshold was minimal (Figure 25, Chapter 4). However, such modest variation may have impacts on relative fitness, and moreover, more stringent Z-scoring thresholds may be required in order to separate genuine counts of particular fluorescent signals having low signal from noise, spill-over, and autofluorescence. To determine if the inferred fitness hierarchy varies as a function of the Z-value thresholds used, I examined the effects of altering Z-score cutoff values in assessing fitness. As illustrated in Figure 29A, relative fitness among strains is not substantially impacted by alteration of the maximum Z-score cutoff values used to determine fitness when the minimum Z-score is held constant. However, alteration of the minimum Z-score threshold does significantly impact fitness hierarchies at scores of -5.0 to -1.5, as shown in Figure 29B.

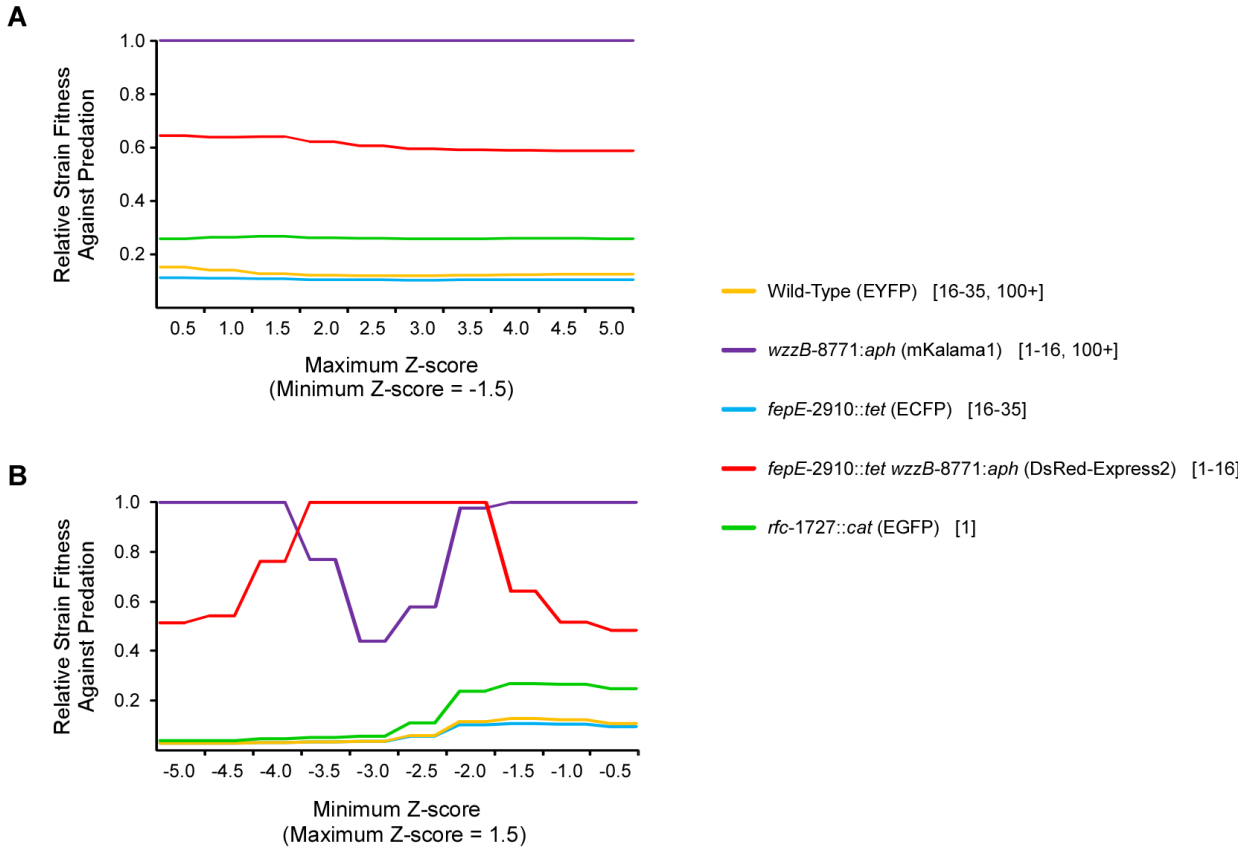

Figure 29. Properly chosen Z-scoring values do not substantially impact relative strain fitness calculations for a sample competition set

A. Relative strain fitness as a function of maximum Z-score when minimum Z-score is held constant to a value of - 1.5. **B**. Relative strain fitness as a function of minimum Z-score when maximum Z-score is held constant to a value of +1.5.

Most notably, the strains *wzzB*-8771::*aph* (mKalama1) and *fepE*-2910::*tet wzzB*-8771::*aph* (DsRed-Express2) were most susceptible to significant changes in fitness hierarchy between minimum Z-score thresholds of -5.0 to -1.5 when the maximum Z-score threshold was kept constant at 1.5. These strains were tagged with the two fluorescent proteins (mKalama1 and DsRed-Express2) having the lowest fluorescent signal strength relative to other fluorescent proteins, as shown in Chapter 4 (Figure 23). Additionally, I consistently observed large fractions of autofluorescence and spill-over signal from other detectors in the mKalama1 and DsRed-Express2 detectors, especially for no predator plate samples. For these two strains, using too low minimum Z-score thresholds inappropriately classify noise events lying outside of the Gaussian distribution as genuine signal, which significantly impacts fitness calculations. Thus, the minimum Z-score threshold must be selected with caution as to avoid assigning noise events to a particular fluorescent class.

To minimize the effect of noise such as autofluorescence and spill-over in analyzing fitness data, I chose to calculate fitness using Z-score thresholds of -1.0 and +1.5. While this range does not include the entirety of events within each fluorescent class, it strongly decreases the likelihood of assigning a noise event to a true fluorescent class. Setting appropriate Z-scoring thresholds, especially the minimum Z-score value, effectively minimizes the effect of noise on fitness data. Importantly, fitness hierarchies did not change along the Z-scoring threshold I chose to analyze competition data. Taken together, the robust curve-fitting, low within-experiment variability, and Z-scoring thresholds used to measure fitness indicate that my approach to fitness determination is sufficient to rigorously examine strain fitness against protozoan predation.

5.2.2 Fitness hierarchy among O-antigen chain length mutants varies among strains and predators

Predators may discriminate among O-antigen chain-length variants of *Salmonella* for mundane reasons. For example, the O-antigen may occlude the preferred receptor, and thus the different mutants conceal the preferred receptor to varying degrees. Alternatively, the different chain lengths may provide resistance to degradation within feeding vacuoles, leading to differential survivorship following expulsion of spent feeding vacuoles by the predator. In these cases, I would expect the fitness hierarchy seen in Figure 26 to be shared among strains varying in O-antigen identity. However, if O-antigen identity influences predator choice, then this hierarchy should differ between different combinations of predator and prey.

To test this hypothesis, the fitness hierarchy among O-antigen chain length mutant derivatives of the SARB3 *rfb* near-isogenic strain KAB082 against predation by the amoeba *N. gruberi* NL was compared to other *rfb* near-isogenic prey having different O-serotypes. Critically, these strains only differ from each other in terms of O-antigen identity and chain length; all other physiological attributes and cell surface structures are identical. Thus, if discrimination among a given collection of strains is observed, then I am able to conclude that any observed differences in fitness against predation can be attributed to the O-antigen. As observed for five different O-serotypes, O-antigen chain length mutants do not have identical fitness against predation by *N. gruberi* NL (Figure 30). Although overall fitness patterns are similar, the hierarchies of relative fitness against predation for all five O-serotype chain length variant strain sets are not completely identical. These data indicate that predator discrimination of O-antigen chain length variants of *Salmonella* is not simply due to differential digestion or occlusion of preferred receptors, rather, the differences in relative fitness here indicate that O-

antigen identity influences predator choice among strains differing only in terms of O-antigen chain length.

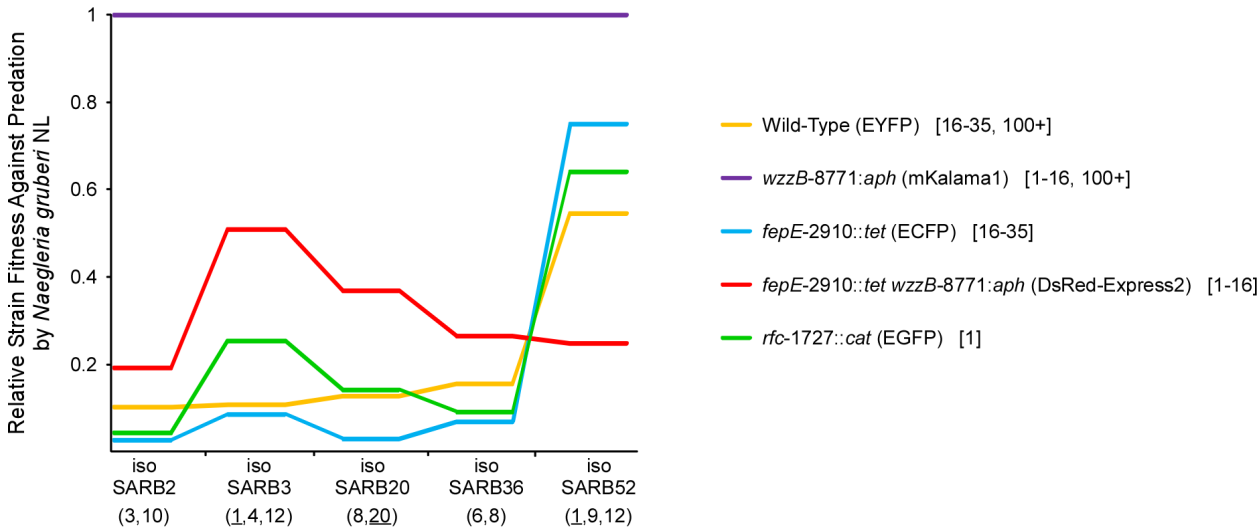

Figure 30. Relative strain fitness for O-antigen chain length mutant derivatives of five *rfb* near-isogenic strains against predation by *Naegleria gruberi* NL

O-antigen chain length mutants of *rfb* near-isogenic strains were competed in the presence and absence of predation by the laboratory strain of the amoeba predator *Naegleria gruberi* NL. For each experiment, the strain having the greatest raw fitness against predation was assigned a value of 1; the fitness values of all other strains in that experiment were normalized to this value. Raw strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). O-serotypes are listed in parentheses underneath strain designations along the X-axis. Fluorescent tag identity is shown in parentheses next to the strain legend. Predicted O-antigen chain length classes are depicted in brackets.

To determine if discrimination of *Salmonella* based on O-antigen chain length is a general property of amoebae and not just unique to the NL laboratory strain of *N. gruberi*, I competed the variant O-antigen chain length strain collection derivatives of *rfb* near-isogenic strains of SARB2 (KAB081), SARB3 (KAB082), and SARB20 (KAB084) against predation by two genetically unrelated amoebae *Tetramitus* sp. BD1-1 and *Acanthamoeba* sp. R2-1. If O-antigen identity truly influences *Salmonella* survival against protozoan predation, then I should observe differences in relative fitness against predation for different O-serotypes competed against different protozoan predators. If discrimination among strains is not a result of O-antigen identity but instead due to the possibilities discussed above, then I should not observe differences in strain relative fitness against predation by different amoebae.

Figure 31 depicts the relative fitness of all O-serotype chain length derivative strains against predation by *N. gruberi* NL, *Tetramitus* sp. BD1-1, and *Acanthamoeba* sp. R2-1; relative fitness of strains changes based on both the identity of the O-antigen and the identity of the protozoan predator. Measurement of raw strain fitness against predation is robust, as confirmed with very low deviations observed when events were randomly sampled from the population of experimental data with bootstrapping at 1000 iterations (Table 31).

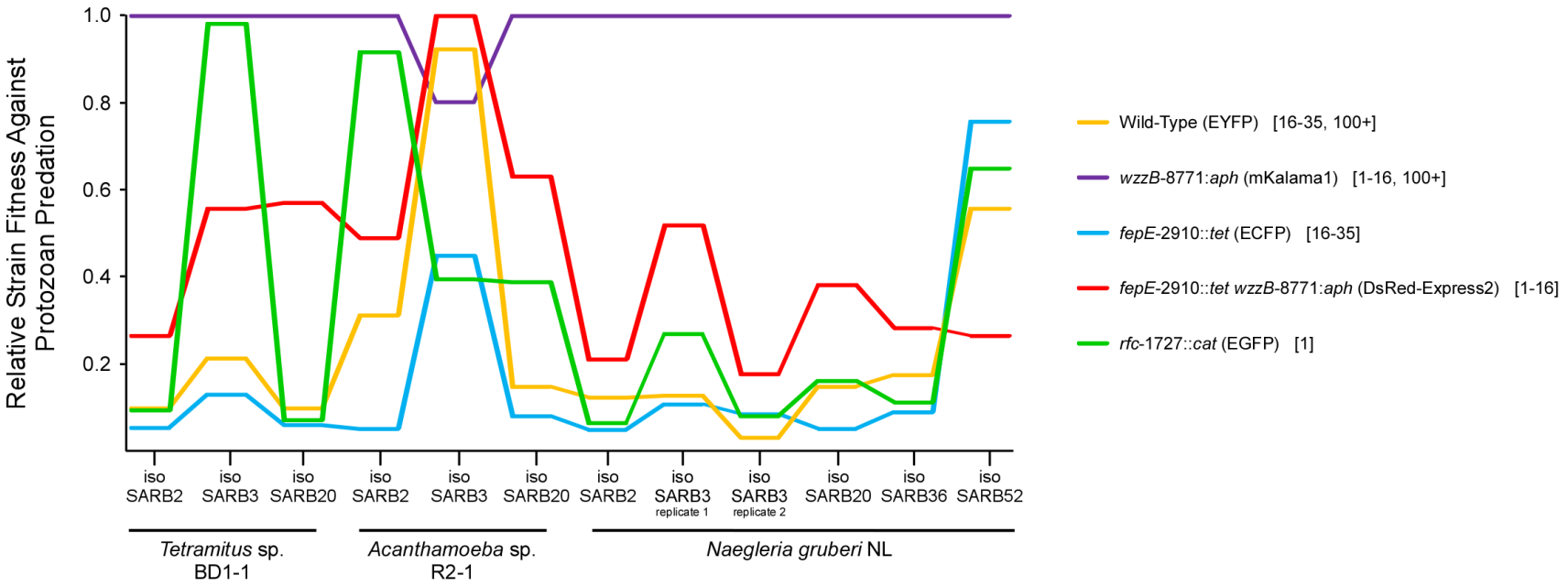

Figure 31. Relative strain fitness of O-antigen chain length mutant derivatives of *rfb* near-isogenic strains varies against protozoan predation

O-antigen chain length mutants of *rfb* near-isogenic strains were competed in the presence and absence of predation by three different protozoan predators. For each experiment, the strain having the greatest raw fitness against predation was assigned a value of 1; the fitness values of all other strains in that experiment were normalized to this value. Raw strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Fluorescent tag identity is shown in parentheses, and predicted O-antigen chain length classes are shown in brackets.

Table 31. Raw strain fitness with deviations for O-antigen chain length variant strain competed against protozoan predators

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Strain	Tag	iso SARB2 vs. *Tetramitus* sp. BD1-1		iso SARB3 vs. *Tetramitus* sp. BD1-1		iso SARB20 vs. *Tetramitus* sp. BD1-1		iso SARB2 vs. *Acanthamoeba* sp. R2-1	
		Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation
Wild-Type	EYFP	0.8692	0.0145	0.2856	0.0423	0.7486	0.0078	1.1362	0.044
wzzB-8771::*aph*	mKalama1	11.6389	0.4155	1.4987	0.0669	10.0324	0.1833	3.872	0.0746
fepE-2910::*tet*	ECFP	0.3298	0.0134	0.1602	0.0064	0.3474	0.0079	0.1015	0.0011
fepE-2910::*tet* *wzzB*-8771::*aph*	DsRed-Express2	2.8443	0.1765	0.8171	0.0326	5.602	0.1982	1.8404	0.0368
rfc-1727::*cat*	EGFP	0.818	0.0738	1.4724	0.0068	0.4665	0.0304	3.5392	0.0367

Table 31 continued. Raw strain fitness with deviations for O-antigen chain length variant strain competed against protozoan predators.

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Strain	Tag	iso SARB3 vs. *Acanthamoeba* sp. R2-1		iso SARB20 vs. *Acanthamoeba* sp. R2-1		iso SARB2 vs. *Naegleria gruberi* NL		iso SARB3 replicate 1 vs. *Naegleria gruberi* NL	
		Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation
Wild-Type	EYFP	1.8732	0.0166	0.5295	0.0044	1.836	0.0238	0.7812	0.0206
wzzB-8771::*aph*	mKalama1	1.6192	0.0263	4.274	0.0543	18.4366	1.0333	7.526	0.3812
fepE-2910::*tet*	ECFP	0.8818	0.0076	0.2383	0.0029	0.4406	0.0048	0.629	0.0038
fepE-2910::*tet* *wzzB*-8771::*aph*	DsRed-Express2	2.033	0.0511	2.6536	0.0844	3.4812	0.1457	3.8162	0.1447
rfc-1727::*cat*	EGFP	0.7708	0.0034	1.5896	0.0085	0.7136	0.0105	1.8855	0.0152

Table 31 continued. Raw strain fitness with deviations for O-antigen chain length variant strain competed against protozoan predators

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Strain	Tag	iso SARB3 replicate 2 vs. *Naegleria gruberi* NL		iso SARB20 vs. *Naegleria gruberi* NL		iso SARB36 vs. *Naegleria gruberi* NL		iso SARB52 vs. *Naegleria gruberi* NL	
		Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation
Wild-Type	EYFP	0.0683	0.0098	0.982	0.0077	1.0553	0.0071	0.9036	0.0078
wzzB-8771::*aph*	mKalama1	15.89	0.685	7.8169	0.1863	6.8882	0.1368	1.6572	0.0413
fepE-2910::*tet*	ECFP	0.9377	0.0051	0.204	0.0036	0.4523	0.0048	1.2428	0.1267
fepE-2910::*tet* *wzzB*-8771::*aph*	DsRed-Express2	2.452	0.0733	2.862	0.0811	1.8177	0.0751	0.4056	0.0213
rfc-1727::*cat*	EGFP	0.8874	0.0071	1.0796	0.0139	0.6024	0.0135	1.06	0.0014

Thus, predator choice of prey is influenced by O-antigen identity and is most likely a general property of amoebae. Although I demonstrated the role played by O-antigen identity in shaping *Salmonella* fitness against predation, these strain fitness values cannot be used to deduce the relationship between O-antigen chain length, identity, and fitness against protozoan predation. However, I cannot address the nature of the predator-prey interaction using these data alone because differences in strain fitness in these assays reflect two influences: differences in chromosomal genotype and differences in the fluorescent tag. I previously observed differences in my ability to detect the signals of certain fluorescent proteins, especially mKalama1 and DsRed-Express2, over time. Because flow cytometry only permits the enumeration of cells that fluoresce, I do not expect that the inferred fitness of strains will be independent of the fluorescent protein used to detect it. To examine different models of predator-prey interaction, I needed to deconvolute the fitness contributions of genotype and fluorophore, as discussed below.

5.2.3 Deconvolution of genotype and fluorophore fitness

The experimental design described above does not allow for separation of fitness contributions from genotype and fluorophore; there are only 5 experimental strain fitness (w_S) values calculated, and ten variables affecting them: five genotype fitness (w_G) and five fluorophore fitness (w_F) values. To separate the contributions of differences in genotype and fluorophore to strain fitness, I conducted reciprocal tag assays in which each genotypically-distinct strain was marked with two different fluorescent proteins; the two reciprocally-tagged sets of strains were simultaneously competed in the presence and absence of amoebae. These set of ten observed strain fitness values will allow for deconvolution of genotype fitness and fluorophore fitness.

When the hierarchies of strain fitness values inferred from each set of strains are compared (Figure 32), it is not surprising that relative fitness is not a simple function of either underlying chromosomal genotype or underlying fluorophore. That is, the most-fit strain in each set of strains does not necessarily share genotype or fluorophore, as its overall fitness in this assay depends upon both. This is not due to error, as deviations for raw strain fitness values in both tag sets (Tables 32 and 33) are very low. Thus, I proceeded to deconvolute strain fitness into its components of genotype fitness (w_G) and fluorescent tag fitness (w_F). Positing that these fitness contributions were independent, I log-transformed w_S so that log strain fitness (w_S') could be expressed as the sum of log genotype fitness (w_G') and log fluorophore fitness (w_F'). Final values were calculated as described in the Methods section.

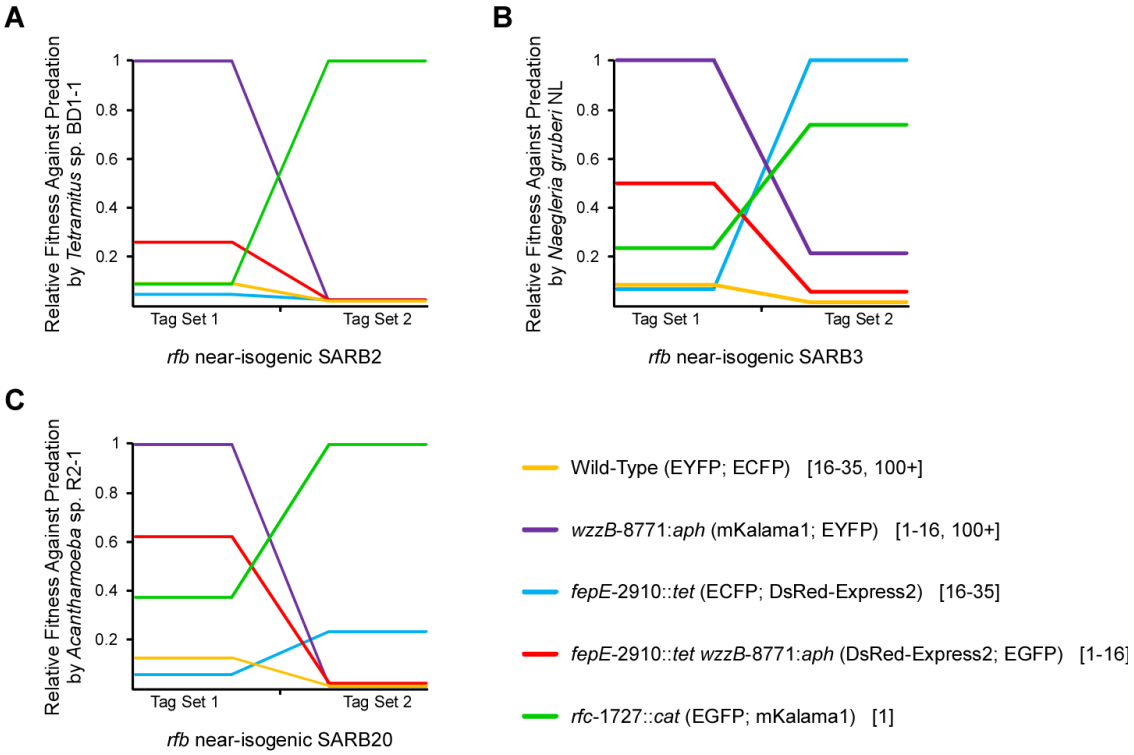

Figure 32. Comparison of relative strain fitness among reciprocal tag competition experiments illustrates a fluorescent tag component to strain fitness against predation

Because strain fitness values are not identical among reciprocally tagged experiments, strain fitness must be deconvoluted into component genotype and fluorescent tag fitness values. Raw strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Fluorescent tag identities are shown in parentheses, and predicted O-antigen chain length classes are listed in brackets. **A**. Relative strain fitness for reciprocally tagged O-antigen chain length mutant derivatives of *rfb* near-isogenic strain SARB2 vs. *Tetramitus* sp. BD1-1. **B**. Relative strain fitness for reciprocally tagged O-antigen chain length mutant derivatives of *rfb* near-isogenic strain SARB3 vs. *Naegleria gruberi* NL. **C**. Relative strain fitness for reciprocally tagged O-antigen chain length mutant derivatives of *rfb* near-isogenic strain SARB20 vs. *Acanthamoeba* sp. R2-1.

Table 32. Raw strain fitness with deviations for O-antigen chain length variant strains (tag set 1) against protozoan predators

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Strain	Tag	iso SARB2 vs. *Tetramitus* sp. BD1-1		iso SARB3 replicate 1 vs. *Naegleria gruberi* NL		iso SARB20 vs. *Acanthamoeba* sp. R2-1	
		Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
Wild-Type	EYFP	0.8692	0.0145	0.7812	0.0206	0.5295	0.0044
wzzB-8771::*aph*	mKalama1	11.6389	0.4155	7.526	0.3812	4.274	0.0543
fepE-2910::*tet*	ECFP	0.3289	0.0134	0.629	0.0038	0.2383	0.0029
fepE-2910::*tet* *wzzB*-8771::*aph*	DsRed-Express2	2.8472	0.1765	3.8162	0.1447	2.6536	0.0844
rfc-1727::*cat*	EGFP	0.8186	0.0738	1.8855	0.0152	1.5896	0.0085

Table 33. Raw strain fitness with deviations for O-antigen chain length variant strains (tag set 2) against protozoan predators

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Strain	Tag	iso SARB2 vs. *Tetramitus* sp. BD1-1		iso SARB3 replicate 1 vs. *Naegleria gruberi* NL		iso SARB20 vs. *Acanthamoeba* sp. R2-1	
		Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
Wild-Type	ECFP	0.1445	0.0074	0.3959	0.0046	0.2968	0.0034
wzzB-8771::*aph*	EYFP	0.8417	0.0242	2.7604	0.0567	0.3485	0.0109
fepE-2910::*tet*	DsRed-Express2	1.9907	0.1667	12.15	0.5241	10.6362	0.3803
fepE-2910::*tet* *wzzB*-8771::*aph*	EGFP	1.8224	0.0231	0.9069	0.0042	0.908	0.0029
rfc-1727::*cat*	mKalama1	370.4569	100.6973	9.0159	0.2978	46.1962	1.7967

To begin, I assessed the variability in fluorescent tag fitness values across experiments. I would predict that fitness differences should be modest, affected by differences in incubation temperature and the availability of oxygen and nutrients. As expected, these values were relatively similar across experiments, with EGFP, EYFP and ECFP showing similar values, and DsRed-Express2 and mKalama1 being more distinct (Figure 33). Additionally, deviations from deconvoluted raw fluorescent tag fitness are very low as determined by bootstrapping experimental data at 1000 iterations (Table 34). Detectable emission signal of fluorescent proteins requires proper protein structure; signal depends on chromophore structure [1, 60, 64, 120-122, 144, 221, 247, 342, 356, 372, 397]. Fluorescent protein stability can be highly susceptible to environmental conditions, such as the availability of molecular oxygen required for proper folding or photobleaching from excessive light exposure [1, 121, 356]. Additionally, the fluorescent proteins used to tag strains in these experiments also vary in quantum yield, or the ratio of emitted photons to absorbed photons; this property is often used to approximate the efficiency of a particular fluorescent protein [1, 66, 70, 122, 155, 246, 342]. Essentially, these fluorescent proteins also vary in the amount of emitted fluorescent signal in relation to the wavelength of light used to excite them. Simply put, a detection bias may exist in strains tagged with a certain fluorescent protein. Thus, it is not surprising that the inherent variability in the physical properties among different fluorescent proteins combined with my experiment-specific detection parameters resulted in differences in the ability to detect signals from certain fluorophores, resulting in a fluorescent tag fitness component to observed raw strain fitness.

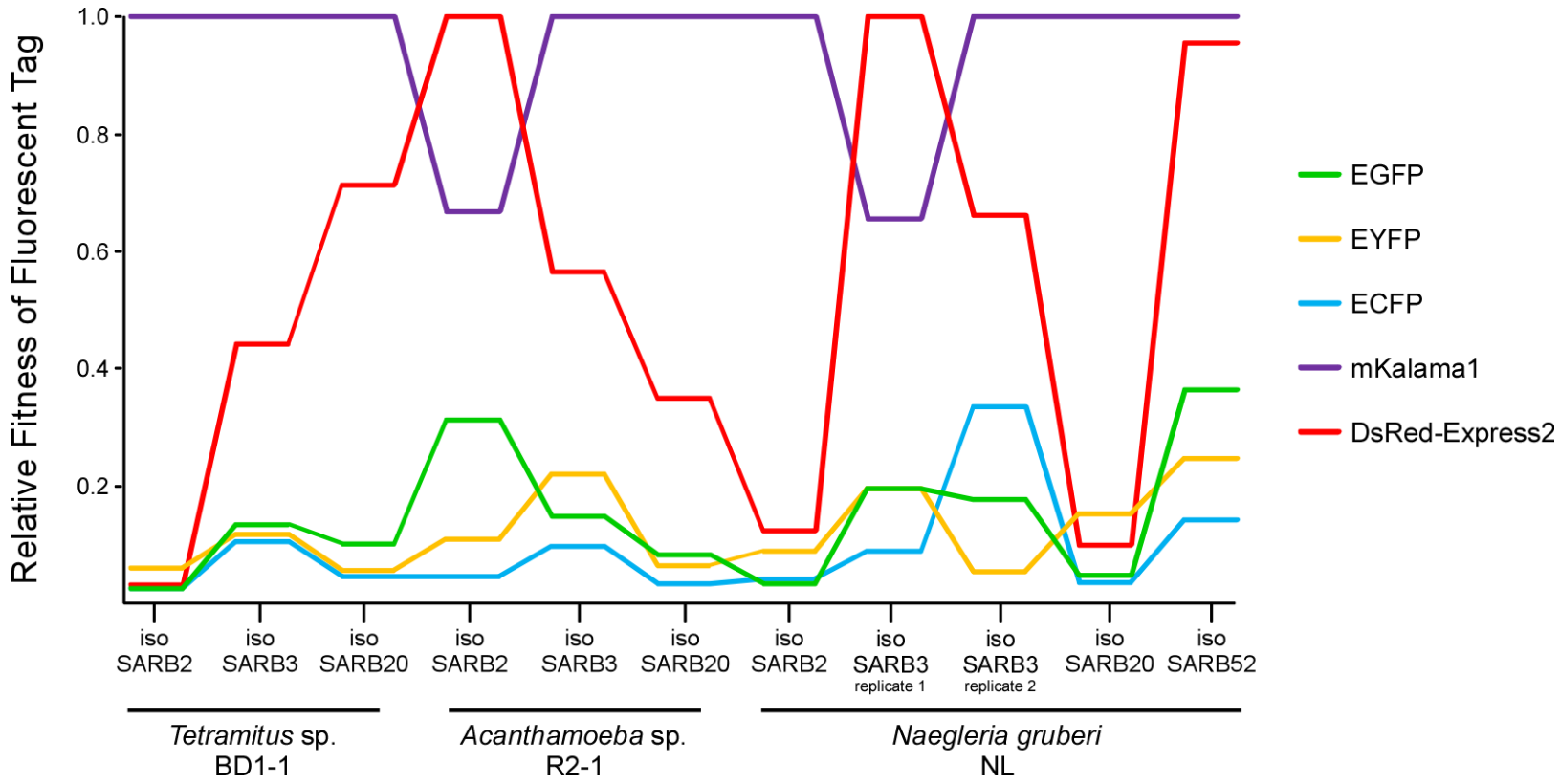

Figure 33. Relative fluorescent tag fitness values are very similar across predation competition experiments

O-antigen chain length mutants of *rfb* near-isogenic strains were competed in the presence and absence of predation by three different protozoan predators. Raw fluorescent tag fitness values were deconvoluted from raw strain fitness values, which were calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). For each experiment, the fluorescent tag having the greatest fitness contribution was assigned a value of 1; the fitness values of all other tags in that experiment were normalized to this value.

Table 34. Raw fitness values with deviation for fluorescent proteins used to tag O-antigen chain length derivative strains competed against protozoan predation

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Fluorescent Tag	iso SARB2 vs. *Tetramitus* sp. BD1-1		iso SARB3 vs. *Tetramitus* sp. BD1-1		iso SARB20 vs. *Tetramitus* sp. BD1-1		iso SARB2 vs. *Acanthamoeba* sp. R2-1	
	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation
EGFP	0.191	0.0283	0.3128	0.0114	0.9894	0.0423	2.019	0.0264
EYFP	1.8484	0.0445	0.2677	0.0204	0.4276	0.0377	0.6075	0.0118
ECFP	0.1671	0.0102	0.2324	0.0142	0.3038	0.0184	0.1724	0.003
mKalama1	47.0146	3.5468	2.7022	0.1438	12.1255	0.8943	4.4746	0.1128
DsRed-Express2	0.5483	0.0624	1.1634	0.0588	8.576	0.3798	6.7712	0.1165

Table 34 continued. Raw fitness values with deviation for fluorescent proteins used to tag O-antigen chain length derivative strains competed against protozoan predation

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Fluorescent Tag	iso SARB3 vs. *Acanthamoeba* sp. R2-1		iso SARB20 vs. *Acanthamoeba* sp. R2-1		iso SARB2 vs. *Naegleria gruberi* NL		iso SARB3 replicate 1 vs. *Naegleria gruberi* NL	
	Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
EGFP	0.6435	0.0091	0.8193	0.0188	0.361	0.0167	1.3337	0.032
EYFP	1.001	0.0121	0.5824	0.0131	2.0706	0.0489	1.3328	0.0367
ECFP	0.379	0.0063	0.1788	0.0041	0.6167	0.0138	0.5099	0.012
mKalama1	4.8973	0.0574	13.0408	0.2499	29.9177	1.3635	4.8139	0.1581
DsRed-Express2	2.7245	0.0587	4.3718	0.1146	3.1716	0.1392	7.435	0.2145

Table 34 continued. Raw fitness values with deviation for fluorescent proteins used to tag O-antigen chain length derivative strains competed against protozoan predation.

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Fluorescent Tag	iso SARB3 replicate 2 vs. *Naegleria gruberi* NL		iso SARB20 vs. *Naegleria gruberi* NL		iso SARB52 vs. *Naegleria gruberi* NL	
	Raw Fitness	Deviation	Raw Fitness	Deviation	Raw Fitness	Deviation
EGFP	0.7668	0.0324	0.4119	0.0208	0.8037	0.0112
EYFP	0.162	0.0148	2.0339	0.0329	0.5302	0.0238
ECFP	1.5488	0.0441	0.23	0.0062	0.284	0.0189
mKalama1	4.8127	0.3323	15.1397	0.5809	2.2972	0.0624
DsRed-Express2	3.1537	0.0859	1.2225	0.0625	2.1916	0.1049

In Chapter 4, I discussed optimization of experimental conditions that addressed significant trade-offs between obtaining strong fluorescent signal and a) among use of various filters and excitation lasers to allow the simultaneous use of multiple fluorescent proteins; b) growth conditions that favored maximal signal and supported the growth of both bacteria and amoebae; and c) incubating plates for a sufficiently long duration as to enable maximal exposure to protozoan predation. Thus, it is not surprising that the identity of a fluorescent tag influences its ability to be detected under these experimental conditions, given the varying intrinsic properties of the fluorescent proteins themselves, experimental conditions that compromise among many factors that influence signal strength, and environmental conditions that vary among experiments. Here, fluorescent tag fitness refers to the ability to detect a given fluorescent-tagged strain rather than the effect that a fluorescent tag has on a strain's susceptibility to predation. Critically, these experiments depend on the ability to detect fluorescently-labeled bacterial cells; I can only count cells that display detectable fluorescent signal based on the mechanical limitations of the flow cytometer and the stringent detection parameters used to enable the simultaneous use of five single fluorescent proteins. Thus, any factor that influences the ability to detect a fluorescent protein in a given experimental condition will influence the fitness of the tagged strain.

For this approach, it is more crucial for these fluorescent tag fitness values to be relatively constant among experiments to further support the validity of experimental design. Accordingly, the Pearson correlation coefficients between experiments are highly similar, with an average Pearson r = 0.8051 (Table 35). While the Pearson r reflects the similarity in magnitude of fluorescent tag fitness among experiments, this does not necessarily reflect the similarity in hierarchy of fluorescent tag fitness among experiments. To further examine this

issue, I analyzed the correlation of fluorescent tag fitness between experiments using the Spearman rank order correlation, which measures the similarity in rank of fitness values among experiments and is less sensitive to large outlier data points than the Pearson correlation. Similar to what I observed with the Pearson correlation coefficients, the Spearman rank order correlation coefficients between experiments are very high (Table 36), with the average Spearman r = 0.7673. Therefore, I conclude that both the experimental design and data analyses are robust, as such high correlation in final fitness values would not be possible by chance alone.

Table 35. Pearson correlation coefficients for pairwise comparisons of fluorescent tag fitness across predation competition experiments

Shaded regions indicate Pearson correlation coefficients equal to or greater than 0.5; darker regions indicate greater correlations. Average Pearson correlation coefficient for all pairwise comparisons is 0.8051.

Experiment		*Tetramitus* sp. BD1-1			*Acanthamoeba* sp. R2-1			*Naegleria gruberi* NL				
		iso SARB2	iso SARB2	iso SARB2	iso SARB2	iso SARB3	iso SARB20	iso SARB2	iso SARB3 (1)	iso SARB3 (2)	iso SARB20	iso SARB52
Tetramitus sp. BD1-1	iso SARB2	1	0.9289	0.7723	0.329	0.8769	0.9495	0.9968	0.3263	0.7955	0.9966	0.628
	iso SARB3	0.9289	1	0.9526	0.6497	0.987	0.998	0.9536	0.6497	0.9402	0.9332	0.8665
	iso SARB20	0.7723	0.9526	1	0.8438	0.9746	0.9323	0.8176	0.8481	0.9571	0.7836	0.971
Acanthamoeba sp. R2-1	iso SARB2	0.329	0.6497	0.8438	1	0.7236	0.6067	0.3951	0.9821	0.7357	0.3478	0.9405
	iso SARB3	0.8769	0.987	0.9746	0.7236	1	0.9791	0.9116	0.7376	0.9249	0.8928	0.9137
	iso SARB20	0.9495	0.998	0.9323	0.6067	0.9791	1	0.9696	0.604	0.9216	0.9529	0.8386
Naegleria gruberi NL	iso SARB2	0.9968	0.9536	0.8176	0.3951	0.9116	0.9696	1	0.3978	0.8311	0.9968	0.6822
	iso SARB3	0.3263	0.6497	0.8481	0.9821	0.7376	0.604	0.3978	1	0.7373	0.3544	0.9303
	iso SARB3	0.7955	0.9402	0.9571	0.7357	0.9249	0.9216	0.8311	0.7373	1	0.7869	0.8817
	iso SARB20	0.9966	0.9332	0.7836	0.3478	0.8928	0.9529	0.9968	0.3544	0.7869	1	0.6455
	iso SARB52	0.628	0.8665	0.971	0.9405	0.9137	0.8386	0.6822	0.9303	0.8817	0.6455	1

Table 36. Spearman rank order correlation coefficients for pairwise comparisons of fluorescent tag fitness across predation competition experiments

Shaded regions indicate Spearman rank order correlation coefficients equal to or greater than 0.5; darker regions indicate greater correlations. Average Spearman rank correlation coefficient for all pairwise comparisons is 0.7673.

Experiment		*Tetramitus* sp. BD1-1			*Acanthamoeba* sp. R2-1			*Naegleria gruberi* NL				
		iso SARB2	iso SARB3	iso SARB20	iso SARB2	iso SARB3	iso SARB20	iso SARB2	iso SARB3 (1)	iso SARB3 (2)	iso SARB20	iso SARB52
Tetramitus sp. BD1-1	iso SARB2	1	0.7	0.7	0.5	0.9	0.7	0.8	0.5	0.3	1	0.7
	iso SARB3	0.7	1	1	0.9	0.9	1	0.7	0.9	0.7	0.7	1
	iso SARB20	0.7	1	1	0.9	0.9	1	0.7	0.9	0.7	0.7	1
Acanthamoeba sp. R2-1	iso SARB2	0.5	0.9	0.9	1	0.8	0.9	0.6	1	0.6	0.5	0.9
	iso SARB3	0.9	0.9	0.9	0.8	1	0.9	0.9	0.8	0.6	0.9	0.9
	iso SARB20	0.7	1	1	0.9	0.9	1	0.7	0.9	0.7	0.7	1
Naegleria gruberi NL	iso SARB2	0.8	0.7	0.7	0.6	0.9	0.7	1	0.6	0.7	0.8	0.7
	iso SARB3	0.5	0.9	0.9	1	0.8	0.9	0.6	1	0.6	0.5	0.9
	iso SARB3	0.3	0.7	0.7	0.6	0.6	0.7	0.7	0.6	1	0.3	0.7
	iso SARB20	1	0.7	0.7	0.5	0.9	0.7	0.8	0.5	0.3	1	0.7
	iso SARB52	0.7	1	1	0.9	0.9	1	0.7	0.9	0.7	0.7	1

In contrast, genotype fitness varied substantially between experiments (Figure 34) with very low error (Table 37). Notably, the variation in genotype fitness was far greater than the variation in fluorophore fitness; whereas the average Pearson *r* for fluorescent tag fitness across all experiments was measured at 0.8051, the average Pearson *r* for genotype fitness across all experiments was measured at 0.0889 (Table 38). An identical trend was observed using the Spearman rank order correlation coefficient; the Spearman *r* for fluorescent tag fitness across all experiments was measured at 0.7673 while the Spearman *r* for genotype fitness across all experiments was 0.0764 (Table 39). While greatest fluorophore fitness was conferred consistently by either mKalama1 or DsRed-Express2, no genotype was consistently beneficial; each of the five genotypes conferred greatest fitness in at least one experiment (Figure 34). Because fluorescent tag fitness remains relatively constant between experiments independent of strain genotype and predator identity, I can conclude that the variability in genotype fitness does not reflect stochastic variation; rather, the differences I observed in genotype fitness reflect a complex interplay between the contributions of O-antigen identity and chain length to *Salmonella* fitness against predation. This supports the hypothesis that differences in relative strains fitness among different predator-prey combinations reflected differences in how the predator interacted with the prey, rather than differences in fluorophore survivability in the experiment. The robustness of these differences was supported by bootstrapping, where events were sampled from the total population with replacement; relative abundances and strain fitness values were recalculated 1000 times (Table 37).

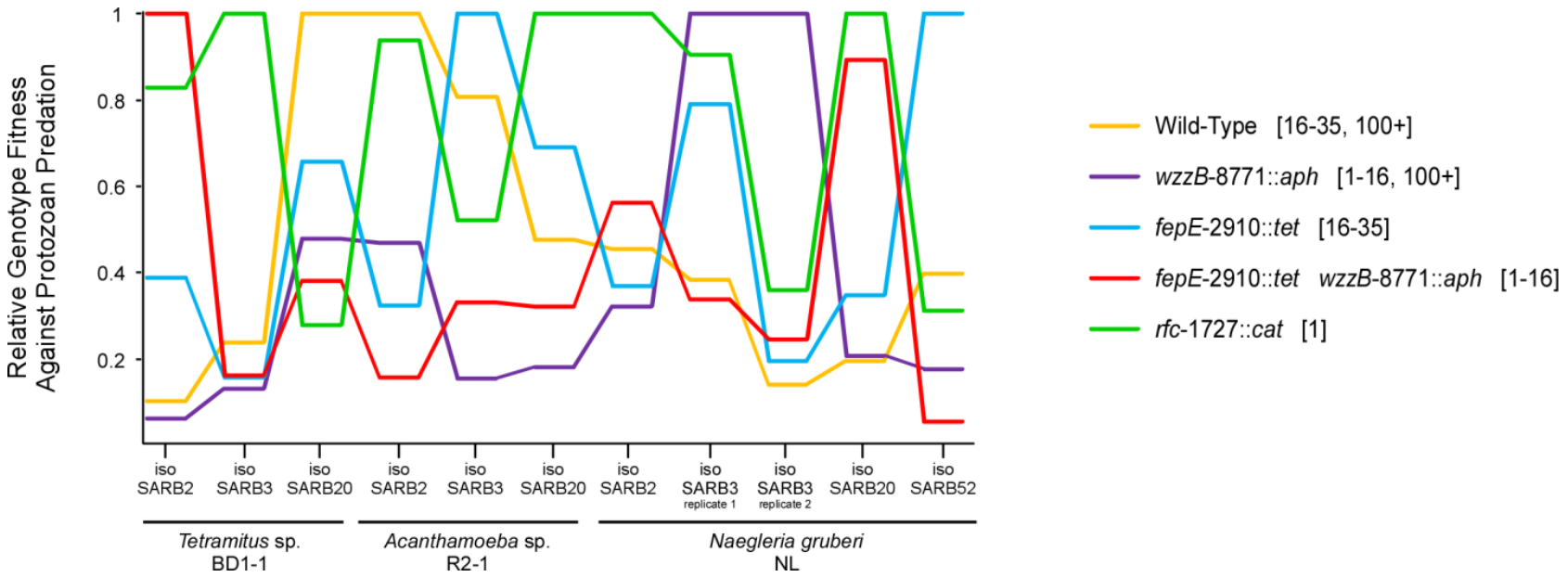

Figure 34. Relative genotype fitness of O-antigen chain length variant strains differs against protozoan predation

O-antigen chain length mutants of *rfb* near-isogenic strains were competed in the presence and absence of predation by three different protozoan predators. Raw genotype fitness values were deconvoluted from raw strain fitness values, which were calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). For each experiment, the genotype having the greatest fitness contribution was assigned a value of 1; the fitness values of all other tags in that experiment were normalized to this value.

Table 37. Raw genotype fitness with deviations for O-antigen chain length variant strains competed against protozoan predation

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Genotype	iso SARB2 vs. *Tetramitus* sp. BD1-1		iso SARB3 vs. *Tetramitus* sp. BD1-1		iso SARB20 vs. *Tetramitus* sp. BD1-1		iso SARB2 vs. *Acanthamoeba* sp. R2-1	
	Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
Wild-Type	0.6676	0.0225	1.2437	0.0862	2.2081	0.1591	2.8576	0.0492
wzzB-8771::*aph*	0.3515	0.0126	0.6464	0.0427	1.0435	0.0733	1.322	0.0282
fepE-2910::*tet*	2.8023	0.2626	0.8033	0.0391	1.4421	0.0742	0.8993	0.0163
fepE-2910::*tet* *wzzB*-8771::*aph*	7.3652	1.0747	0.8185	0.0325	0.8239	0.0324	0.4152	0.0052
rfc-1727::*cat*	6.0816	1.1184	5.486	0.2431	0.5947	0.0265	2.678	0.0461

Table 37 continued. Raw genotype fitness with deviations for O-antigen chain length variant strains competed against protozoan predation

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Table continued on next page.

Genotype	iso SARB3 vs. *Acanthamoeba* sp. R2-1		iso SARB20 vs. *Acanthamoeba* sp. R2-1		iso SARB2 vs. *Naegleria gruberi* NL		iso SARB3 replicate 1 vs. *Naegleria gruberi* NL	
	Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
Wild-Type	2.3069	0.0333	1.2845	0.0274	0.9869	0.0204	0.6813	0.0162
wzzB-8771::*aph*	0.4076	0.0035	0.4631	0.0072	0.6862	0.02	1.8173	0.0537
fepE-2910::*tet*	2.8684	0.0565	1.8827	0.049	0.7955	0.0239	1.4339	0.0391
fepE-2910::*tet* *wzzB*-8771::*aph*	0.9199	0.0151	0.8576	0.0206	1.2223	0.0572	0.5966	0.0145
rfc-1727::*cat*	1.4765	0.0197	2.7413	0.0653	2.2013	0.1065	1.6433	0.0457

Table 37 continued. Raw genotype fitness with deviations for O-antigen chain length variant strains competed against protozoan predation

Values were obtained by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Genotype	iso SARB3 replicate 2 vs. *Naegleria gruberi* NL		iso SARB20 vs. *Naegleria gruberi* NL		iso SARB52 vs. *Naegleria gruberi* NL	
	Raw Fitness	**Deviation**	**Raw Fitness**	**Deviation**	**Raw Fitness**	**Deviation**
Wild-Type	0.5479	0.0217	0.5739	0.0088	2.8503	0.1711
wzzB-8771::*aph*	4.269	0.3701	0.6138	0.0149	1.2066	0.0476
fepE-2910::*tet*	0.7828	0.0145	1.0547	0.0356	7.3815	0.4088
fepE-2910::*tet* *wzzB*-8771::*aph*	1.0053	0.0349	2.7829	0.1422	0.3096	0.0053
rfc-1727::*cat*	1.4964	0.0895	3.1158	0.1698	2.2059	0.0411

Table 38. Pearson correlation coefficients for pairwise comparisons of O-antigen chain length genotype fitness against protozoan predation

Darker red colors indicate greater negative Pearson correlation coefficients; darker blue colors indicate greater positive Pearson correlation coefficients. Average Pearson correlation coefficient = 0.0889.

Experiment	BD1-1 iso SARB2	NL iso SARB20	NL iso SARB2	BD1-1 iso SARB3	R2-1 iso SARB20	R2-1 iso SARB2	NL iso SARB3 (1)	NL iso SARB3 (2)	R2-1 iso SARB3	NL iso SARB52	BD1-1 iso SARB20
BD1-1 iso SARB2	1	0.9593	0.6732	0.4433	0.3811	-0.2904	-0.256	-0.3852	-0.1644	-0.2748	-0.6917
NL iso SARB20	0.9593	1	0.8373	0.6568	0.4928	-0.0764	-0.0911	-0.264	-0.2509	-0.3593	-0.7653
NL iso SARB 2	0.6732	0.8373	1	0.9505	0.7574	0.4612	0.0975	-0.2521	-0.0783	-0.2536	-0.5491
BD1-1 iso SARB3	0.4433	0.6568	0.9505	1	0.8272	0.6065	0.3374	-0.1171	-0.0084	-0.1083	-0.4729
R2-1 iso SARB20	0.3811	0.4928	0.7574	0.8272	1	0.478	0.253	-0.4761	0.5166	0.4345	-0.2054
R2-1 iso SARB2	-0.2904	-0.0764	0.4612	0.6065	0.478	1	0.0817	-0.1367	0.2313	-0.0293	0.3386
NL iso SARB3 (1)	-0.256	-0.0911	0.0975	0.3374	0.253	0.0817	1	0.6768	-0.2002	0.2116	-0.4278
NL iso SARB3 (2)	-0.3852	-0.264	-0.2521	-0.1171	-0.4761	-0.1367	0.6768	1	-0.746	-0.3904	-0.3611
R2-1 iso SARB3	-0.1644	-0.2509	-0.0783	-0.0084	0.5166	0.2313	-0.2002	-0.746	1	0.8674	0.6062
NL iso SARB52	-0.2748	-0.3593	-0.2536	-0.1083	0.4345	-0.0293	0.2116	-0.3904	0.8674	1	0.3913
BD1-1 iso SARB20	-0.6917	-0.7653	-0.5491	-0.4729	-0.2054	0.3386	-0.4278	-0.3611	0.6062	0.3913	1

Table 39. Spearman rank order correlation coefficients for pairwise comparisons of O-antigen chain length genotype fitness against protozoan predation

Darker red colors indicate greater negative Spearman rank order correlation coefficients; darker blue colors indicate greater positive Spearman rank order correlation coefficients. Average Spearman rank order correlation coefficient = 0.0764.

Experiment	BD1-1 iso SARB2	NL iso SARB20	NL iso SARB2	BD1-1 iso SARB3	R2-1 iso SARB20	R2-1 iso SARB2	NL iso SARB3 (1)	NL iso SARB3 (2)	R2-1 iso SARB3	NL iso SARB52	BD1-1 iso SARB20
BD1-1 iso SARB2	1	0.8	0.8	0.5	0.4	-0.5	-0.6	-0.1	0.1	-0.3	-0.6
NL iso SARB20	0.8	1	0.7	0.4	0.5	-0.4	0	0.4	-0.1	-0.3	-0.9
NL iso SARB 2	0.8	0.7	1	0.9	0.6	0.1	-0.4	-0.1	0.1	-0.2	-0.6
BD1-1 iso SARB3	0.5	0.4	0.9	1	0.7	0.5	-0.3	-0.3	0.3	0.1	-0.3
R2-1 iso SARB20	0.4	0.5	0.6	0.7	1	0.3	0	-0.3	0.7	0.6	-0.2
R2-1 iso SARB2	-0.5	-0.4	0.1	0.5	0.3	1	0.3	-0.2	0.2	0.4	0.3
NL iso SARB3 (1)	-0.6	0	-0.4	-0.3	0	0.3	1	0.7	-0.3	0.1	-0.2
NL iso SARB3 (2)	-0.1	0.4	-0.1	-0.3	-0.3	-0.2	0.7	1	-0.8	-0.6	-0.7
R2-1 iso SARB3	0.1	-0.1	0.1	0.3	0.7	0.2	-0.3	-0.8	1	0.9	0.5
NL iso SARB52	-0.3	-0.3	-0.2	0.1	0.6	0.4	0.1	-0.6	0.9	1	0.6
BD1-1 iso SARB20	-0.6	-0.9	-0.6	-0.3	-0.2	0.3	-0.2	-0.7	0.5	0.6	1

The Pearson correlation coefficients (Table 38) and Spearman rank order correlation coefficients (Table 39) for pairwise comparisons of genotype fitness between experiments listed present a continuum of fitness relationships. This is to be expected if different genotypes are preferred to different degrees by individual predators. At extreme ends of this spectrum, I see what appears to be two broad classes of predator-strain experiments. To examine this issue further, I calculated the distance values for pairwise genotype fitness values between experiments and used these to construct a distance tree using the neighbor-joining method [290]. As shown in Figure 35, genotypes can be sorted into several classes; fitness hierarchies within these clusters show patters that are generally similar to each other but differ from those in the other class. Here, clustering depends on both the identity of the O-antigen and of the predator; neither all experiments of any given *rfb* near-isogenic strain set nor all experiments of any given predator cluster together. Critically, replicate competition experiments of *rfb* near-isogenic O-antigen chain length variant sets SARB3 vs. *N. gruberi* NL cluster together, suggesting the repeatability of this experimental approach. These patterns suggest that predators are recognizing prey not only as a function of the length of their O-antigens but also as a function of their identities.

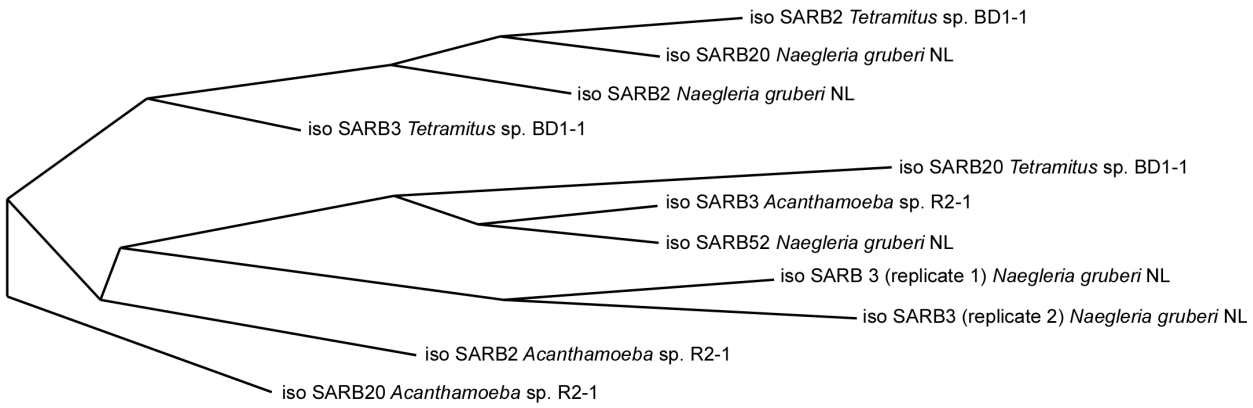

Figure 35. Neighbor-joining tree of distance values of pairwise comparisons of genotype fitness among predation experiments

5.3 REPRODUCIBILITY OF EXPERIMENTS

To examine data reproducibility, I performed three replicate experiments competing the O-antigen chain length derivative series from *rfb* near-isogenic SARB3 against predation by *N. gruberi* NL. These experiments were conducted at different times and utilized independent preparations of overnight bacterial cultures and growth media. Amoebae cysts used in these experiments were from the same cyst preparation but were stored for different amounts of time.

5.3.1 Fluorescent tag fitness is similar across replicate experiments

As expected, fluorescent tag fitness was very similar across experiments, with the average Pearson $r = 0.7082$ (Table 40) and Spearman rank order $r = 0.6667$ (Table 41). These values are very similar to the average fluorescent tag fitness across all experiments (Tables 35 and 36). Although environmental conditions that influence fluorescent tag fitness, such as humidity, temperature, and bacterial cell lawn density, vary between experiments, global fluorescent tag fitness did not appear to significantly change between experiments. Therefore, I again conclude that differences in fluorescent tag fitness were insufficient to explain the differences observed in fitness of strains in these predation competition tests. Additionally, the approach used to deconvolute fluorescent fitness from raw strain fitness is robust.

Table 40. Pearson correlation coefficients for pairwise comparisons of fluorescent tag fitness of three replicate experiments of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (SARB3) against predation by *Naegleria gruberi* NL

Darker blue colors indicate higher degrees of correlation. Average Pearson $r = 0.7082$.

Replicate	1	2	3
1	1	0.7373	0.5732
2	0.7373	1	0.8142
3	0.5732	0.8142	1

Table 41. Spearman rank order correlation coefficients for pairwise comparisons of fluorescent tag fitness in three replicate experiments of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (SARB3) against predation by *Naegleria gruberi* NL

Darker blue colors indicate higher degrees of correlation. Average Spearman $r = 0.6667$.

Replicate	1	2	3
1	1	0.6	0.8
2	0.6	1	0.6
3	0.8	0.6	1

5.3.2 Within-experiment variability can result in differences in genotype fitness among experiments

While fluorescent tag fitness did not significantly differ among replicate experiments, only the first two replicate experiments produced similar genotype fitness values and clustered with each other (Figures 34-35 and Tables 42-43). The third replicate experiment produced genotype fitness values that were different from the two other replicates. Genotype fitness between replicate experiments 1 and 2 were very similar (Pearson r = 0.6768 and Spearman r = 0.7). Correlations in genotype fitness were nearly identical to those observed for fluorescent tag fitness (Tables 40 and 41). Because both fluorescent tag and genotype fitness values did not differ substantially differ between experiments 1 and 2, I believe that the experimental techniques described in this chapter are robust.

While the fluorescent tag fitness of experimental replicate 3 are highly correlated to those of experimental replicates 1 and 2, the genotype fitness of this replicate differ greatly from both other replicates (Figure 36 and Tables 42-43). Replicate 3 showed overall negative correlation coefficients when compared to replicates 1 and 2, with Pearson r = -0.362 and Spearman r = -0.45, indicating that the genotype fitness hierarchies are substantially different between these two data sets.

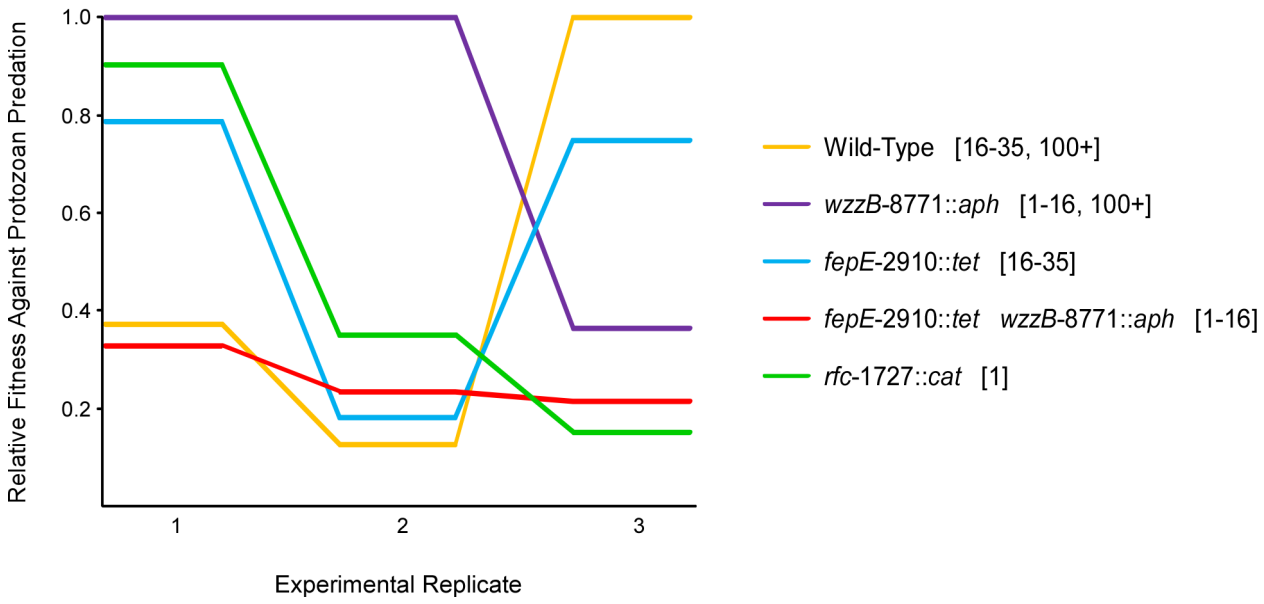

Figure 36. Relative genotype fitness of replicate competition experiments of O-antigen chain length derivatives of *rfb* near-isogenic SARB3 (KAB082) vs. *Naegleria gruberi* NL

Raw genotype fitness values were deconvoluted from raw strain fitness values, which were calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). For each experiment, the genotype having the greatest fitness contribution was assigned a value of 1; the fitness values of all other genotypes in that experiment were normalized to this value. For experiments 1 vs. 2, Pearson r = 0.6768 and Spearman r = 0.7. For experiments 1 and 2 vs. experiment 3, Pearson r = -0.362 and Spearman r = -0.45. Predicted O-antigen chain length classes are shown in brackets. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Table 42. Pearson correlation coefficients for pairwise comparisons of three replicate experiments of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 against predation by *Naegleria gruberi* NL

Darker red colors indicate greater negative Pearson correlation coefficients; darker blue colors indicate greater positive Pearson correlation coefficients.

Replicate	1	2	3
1	1	0.6768	-0.326
2	0.6768	1	-0.398
3	-0.326	-0.398	1

Table 43. Spearman rank order correlation coefficients for pairwise comparisons of three replicate experiments of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 against predation by *Naegleria gruberi* NL

Darker red colors indicate greater negative Spearman rank order correlation coefficients; darker blue colors indicate greater positive Spearman rank order correlation coefficients.

Replicate	1	2	3
1	1	0.7	-0.2
2	0.7	1	-0.7
3	-0.2	-0.7	1

As discussed above, problematic calculation of fluorescent tag fitness cannot explain the differences observed in genotype fitness in replicate experiment 3 as compared to other replicate experiments. Rather, the similarity in fluorescent tag fitness indicated that the data collection and fitness deconvolution were themselves robust; it is only the resulting genotypic fitness values that are questionable. These fitness values were supported by bootstrapping with random resampling of 1000 events within the data population. As discussed in Chapter 5.2.1.2, within-experiment variability tended to be relatively low among plates without predators. Not surprisingly, this is also true for replicate experiment 3. Goodness-of-fit tests reveal that within-experiment variability among no predator plates in competition tag set 1 (Table 44) and competition tag set 2 (Table 45) are acceptably low and comparable to other experiments. Thus, within-experiment variability in no predator plates also cannot explain the aberrant genotype fitness values observed for replicate experiment 3 compared to those obtained for replicate experiments 1 and 2.

Table 44. Goodness-of-fit tests on a set of replicate no predator plates for an aberrant competition experiment of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (iso SARB3)-Tag Set 1 vs. *Naegleria gruberi* NL

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. Similar to other experiments, no predator plates had acceptably low SSD values (< 0.050), indicating a low degree of within-experiment variation for this set of plates.

Replicate	1	2	3	4	5	6
1	0.000	0.003	0.004	0.004	0.043	0.020
2	---	0.000	0.000	0.000	0.025	0.009
3	---	---	0.000	0.001	0.021	0.009
4	---	---	---	0.000	0.024	0.007
5	---	---	---	---	0.000	0.013
6	---	---	---	---	---	0.000

Table 45. Goodness-of-fit tests on a set of replicate no predator plates for an aberrant competition experiment of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (iso SARB3)-Tag Set 2 vs. *Naegleria gruberi* NL

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. Similar to other experiments, no predator plates had acceptably low SSD values (< 0.050), indicating a low degree of within-experiment variation for this set of plates.

Replicate	1	2	3	4	5	6
1	0.000	0.000	0.003	0.007	0.005	0.002
2	---	0.000	0.003	0.009	0.007	0.004
3	---	---	0.000	0.004	0.002	0.001
4	---	---	---	0.000	0.001	0.004
5	---	---	---	---	0.000	0.001
6	---	---	---	---	---	0.000

Differences in genotype fitness would be expected if the predator plates from replicate experiment 3 were harvested before predation was completed for most of the plates in this experiment. In many experiments discussed in this chapter, it was common to remove a few replicate plates from the predator plate data set due to high SSD values between plates. Typically, removal of these aberrant plates resulted in overall improvement of SSD values between remaining plates, yielding acceptably low within-experiment variation in the remaining predator plate data set. As discussed in Chapter 5.2.1.3, I expected a few aberrant plates to exist in each predator data set, as the experimental design used here represents a compromise between obtaining maximal fluorescent signal from *Salmonella* cells and incubating plates for a long enough time to allow completion of predation on all plates in the experiment. Because rates of predation differed among plates due to local environmental variability, it is impossible to optimize experimental conditions that ensured completion of predation on plates without sacrificing fluorescent signal strength.

Consistent with this scenario, predator plates in both tag sets in experimental replicate 3 showed little consistency between replicates, as evidenced by high SSD values between plates (Tables 46 and 47). Unlike what I observed in the majority of other predator data sets, removal of aberrant plates did not improve the SSD values between remaining replicate plates for either competition tag set. Because within-experiment variability was acceptably low for the no predator plates in this experiment, the genotype fitness differences observed for replicate experiment 3 are most likely due to high within-experiment variability in predator plates. Although data was collected from this experiment in the same time window as other experiments, the predator plates in this experiment may have been harvested too early if the environmental conditions in this experiment resulted in slower rates of predation compared to

other experiments. Moreover, using cysts stored for longer periods of time may result in slower rates predator germination, as the amoebae cysts used to inoculate predator plates in this experiment were approximately three weeks older than the cysts used to inoculate prior replicate experiments. In these experiments, predators are seeded in the center of plates and move outward to consume lawns of *Salmonella* cells. If plates were incubated for too little time, the amoebae predators may not have completely migrated to all portions of the plates. These plates may have contained regions of *Salmonella* cells that were unexposed or underexposed to predation; the ratios of cells on these plates may have been more similar to those of plates without predators than those containing predators. While a time-course predation assay would reveal changes of ratios of strains over time as more cells are exposed to amoebae predators, such an approach is impossible using mixed competition tests on solid media. Thus, I relied on the SSD values between replicate plates to discard plates from experiments that most likely represented plates not completely exposed to predation.

Table 46. Goodness-of-fit tests on a set of replicate predator plates for an aberrant competition experiment of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (iso SARB3)-Tag Set 1 vs. *Naegleria gruberi* NL

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. Shaded regions indicate unacceptably high SSD values. Unlike other experiments, removal of aberrant plates did not improve overall SSD values of other plates.

Replicate	1	2	3	4	5	6	7	8	9	10
1	0.000	0.057	0.072	0.098	0.208	0.314	0.218	0.135	0.248	0.158
2	---	0.000	0.029	0.094	0.065	0.179	0.104	0.027	0.123	0.048
3	---	---	0.000	0.052	0.058	0.129	0.070	0.028	0.106	0.089
4	---	---	---	0.000	0.120	0.185	0.053	0.074	0.076	0.125
5	---	---	---	---	0.000	0.049	0.046	0.017	0.080	0.096
6	---	---	---	---	---	0.000	0.119	0.109	0.194	0.274
7	---	---	---	---	---	---	0.000	0.032	0.011	0.08
8	---	---	---	---	---	---	---	0.000	0.049	0.039
9	---	---	---	---	---	---	---	---	0.000	0.058
10	---	---	---	---	---	---	---	---	---	0.000

Table 47. Goodness-of-fit tests on a set of replicate predator plates for an aberrant competition experiment of O-antigen chain length derivatives of *rfb* near-isogenic strain KAB082 (iso SARB3)-Tag Set 2 vs. *Naegleria gruberi* NL

Experimental variability was assessed based on the sums of squared deviations (SSD) of the counts of fluorescent classes of each replicate plate from that expected from of the entire population of plates. Shaded regions indicate unacceptably high SSD values. Unlike other experiments, removal of aberrant plates did not improve overall SSD values of other plates.

Replicate	1	2	3	4	5	6	7	8	9	10
1	0.000	0.095	0.049	0.044	0.016	0.030	0.023	0.014	0.066	0.043
2	---	0.000	0.242	0.041	0.074	0.036	0.063	0.082	0.240	0.017
3	---	---	0.000	0.124	0.110	0.130	0.131	0.091	0.053	0.142
4	---	---	---	0.000	0.068	0.015	0.044	0.058	0.152	0.007
5	---	---	---	---	0.000	0.035	0.016	0.011	0.102	0.050
6	---	---	---	---	---	0.000	0.011	0.021	0.110	0.013
7	---	---	---	---	---	0.011	0.000	0.008	0.097	0.038
8	---	---	---	---	---	---	---	0.000	0.056	0.050
9	---	---	---	---	---	---	---	---	0.000	0.162
10	---	---	---	---	---	---	---	---	---	0.000

These data highlight a critically important point: assessing the variability between plates is critical in providing confidence of the collected data. It was only through the lack of plate replication that I was able to reject the suspect experiments from the final data set. The aberrant genotype fitness values observed in the third replicate data set only served to confirm its unsuitability.

5.4 *SALMONELLA* FITNESS AGAINST PREDATION IS A COMPLEX INTERPLAY BETWEEN O-ANTIGEN IDENTITY AND CHAIN LENGTH

The results presented in this chapter demonstrate that predators can discriminate among *Salmonella* prey that only vary in terms of O-antigen chain length, providing additional support for the hypothesis that protozoan predation may drive diversifying selection at the *Salmonella rfb* locus. Here, I demonstrated that protozoan predators are able to discriminate among prey that only vary in terms of the chain length of the *rfb* product, the O-antigen. Thus, the O-antigen must play a strong role in shaping predator choice among prey; if not, I would have observed little if any fitness differences against predation among the *rfb* near-isogenic strain collection of O-antigen chain length mutants. However, O-antigen chain length was not the sole factor affecting prey choice, as the results presented in Chapter 3 show that O-antigen identity also influenced *Salmonella* fitness against protozoan predation.

Interestingly, fitness against predation occurred along a gradient of chain length, O-antigen identity, and the identity of the protozoan predator. Comparing the Pearson and Spearman rank order correlation coefficients of pairwise comparisons of genotype fitness values between experiments yielded two general clusters of experiments varying both in *rfb* identity of

prey and protozoan predator (Tables 38 and 39). Additionally, a distance tree of genotype fitness values obtained using the neighbor-joining method [290] revealed that fitness is a complex interaction between prey O-antigen identity and chain length and the identity of the predator amoeba (Figure 35).

While shorter O-antigens confer higher fitness against predation in some predator-strain competition experiments, longer O-antigens confer higher fitness in others; this is at odds with the "strength of binding" model explaining the relationship between the O-antigen and protozoan predators briefly discussed earlier. Because I did not observe that universally shortening O-antigens increased fitness against predation, I conclude that feeding preference most likely did not simply depend on how well a predator is able to bind the O-antigen.

For a given set of chain length mutant strains, the relationship of O-antigen chain length to fitness against predation occurs along a gradient depending on the identity of the O-antigen and of the protozoan predator (Figure 37). Here, strains having the highest relative fitness against predation for predator-strain experiments 1-5 represent the shortest chain length derivative classes tested: *fepE*-2910::*tet wzzB*-8771::*aph* (1-16 repeats) for experiment 1 and *rfc*-1727::*cat* (1 repeat) for experiments 2-5. The strain with the highest fitness varied substantially among the remaining predator-strain experiments, with each genotype representing the most fit strain at least once. Although I demonstrated that the nature of the O-antigen (chain length and identity) impact fitness against predation, I do not expect fitness to change on a linear fashion depending on chain length; shorter is not necessarily better in all circumstances. The relationship between O-antigen chain length and fitness against predation may be complicated by the potential unmasking of alternative receptors used for prey recognition in strains with shorter O-antigen chain lengths.

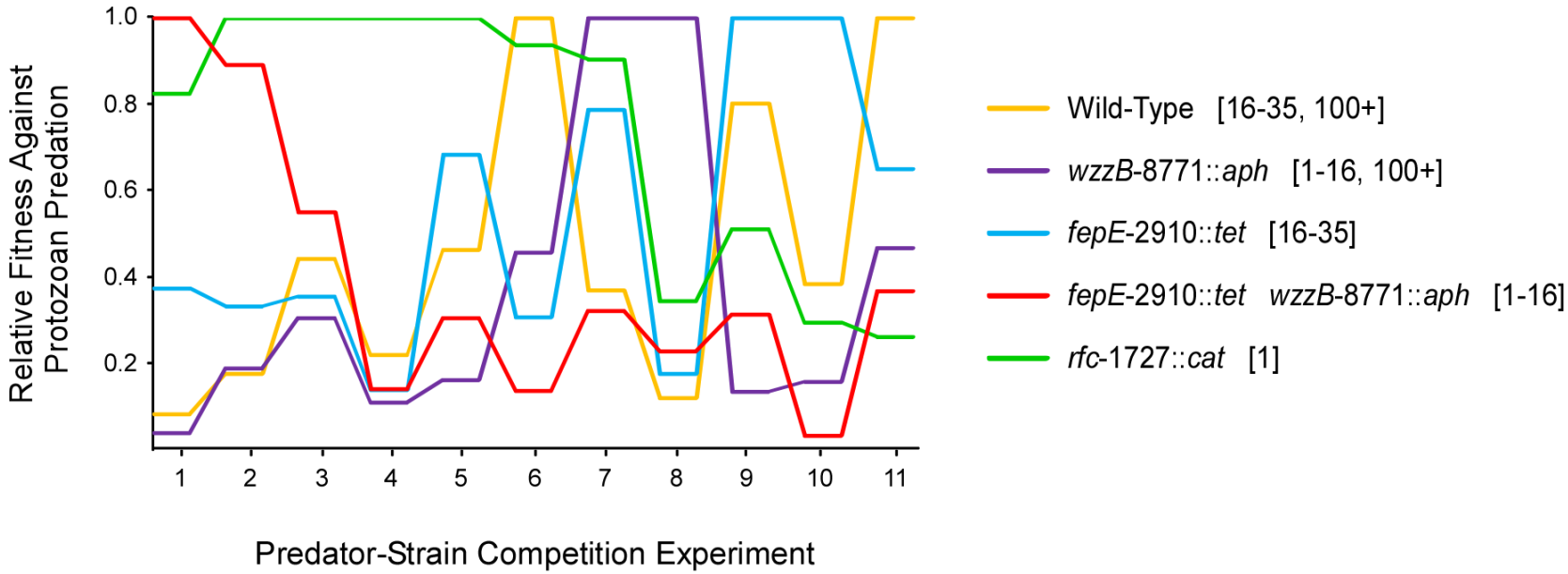

Figure 37. Relative genotype fitness rearranged according to distance values between experiments reveals fitness against predation occurs along a spectrum of O-antigen identity, chain length, and identity of protozoan predator

Raw genotype fitness values were deconvoluted from raw strain fitness values, which were calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). For each experiment, the genotype having the greatest fitness contribution was assigned a value of 1; the fitness values of all other genotypes in that experiment were normalized to this value. Predicted O-antigen chain length classes are shown in brackets. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations. **1** = iso SARB2 vs. *Tetramitus* sp. BD1-1; **2** = iso SARB20 series vs. *Naegleria gruberi* NL; **3** = iso SARB2 series vs. *N. gruberi* NL; **4** = iso SARB3 vs. *Tetramitus* sp. BD1-1; **5** = iso SARB20 series vs. *Acanthamoeba* sp. R2-1; **6** = iso SARB2 series vs. *Acanthamoeba* sp. R2-1; **7** = iso SARB3 series (replicate 1) vs. *Naegleria gruberi* NL; **8** = iso SARB3 series (replicate 2) vs. *N. gruberi* NL; **9** = iso SARB3 series vs. *Acanthamoeba* sp. R2-1; **10** = iso SARB52 vs. *N. gruberi* NL; **11** = iso SARB20 series vs. *Tetramitus* sp. BD1-1.

5.5 INSIGHT INTO THE MECHANISM OF PREY RECOGNITION

If protozoan predation drives diversifying selection at the *Salmonella rfb* locus, then the O-antigen must be mechanistically involved in prey choice. Although I demonstrated here and in Chapter 3 that the O-antigen is a major contributing factor to prey fitness against predation, the mechanism by which amoebae discriminate among *Salmonella* prey is unknown. The data presented in this chapter may provide some insight into the mechanism by which protozoan predators recognize prey.

5.5.1 Different strategies of prey recognition

Prey choice could simply be a function of how well an amoeba bind to a given O-antigen. In this model, strains that have O-antigens more strongly bound by amoebae will be more readily consumed than strains that have O-antigens more weakly bound by amoebae; hence strongly-bound strains will have lower relative fitness against predation than would more weakly-bound strains. If this is true, I would expect that shortening the chain length of strongly-bound O-antigens would have a dramatic impact on fitness against predation. In contrast, shortening the chain length of weakly-bound O-antigens would have little if any impact on fitness against predation; the removal of an already poorly recognized antigen should not impact fitness. In no case would shorter antigens lead to an increase in predation. However, the fitness hierarchies of O-antigen chain length mutant strains determined here reject this model (Figures 35 and 37). First, relative fitness among strains always differed, with a majority of experiments having two-

fold or greater fitness difference between the least and most fit strain (Figure 37). More importantly, there were many cases where shorter O-antigen chains led to increased predation.

These experiments suggest an alternative model of prey choice. Different amoebae isolated from the same environment share similar feeding preferences independent of predator relatedness [386], suggesting that environmental pressures on amoebae predators themselves may be influencing prey choice. This is not unexpected, as amoebae recognize and bind to the intestine surface; however, unlike binding to bacterial prey, amoebae do not attempt to phagocytose the intestinal epithelium. I posit that strong selective pressure should exist on native gut amoebae to avoid grazing upon host intestinal mucins, the bound polysaccharide that serves as a lubricant for the passage of material through the intestinal lumen [22, 154, 212, 252, 279]. If amoebae recognize intestinal mucins for adhesion to their environment, one would expect unrelated predators living in the same intestinal environment to share similar recognition properties. As a result, these amoebae may consume bacterial prey more slowly if those prey resemble the intestinal mucins through molecular mimicry. Here, the O-antigens that confer the highest fitness against predation are those that most resemble the mucins of their native intestinal environment. In this case, shorter O-antigens may not confer a benefit and may in fact result in reduced fitness against predation, as the cell loses its molecular camouflage. Shortening these O-antigens may expose alternative epitopes used by predators to recognize prey, resulting in less concealment of the cells in the environment. O-antigens that do not resemble host mucins may have an increased risk of predation simply because they do not blend in with the intestinal environment; here, shorter chain lengths may or may not confer a fitness advantage. Again, fitness against predation according to this model will occur along a gradient depending on how much O-antigen structures differ from host mucins.

The molecular mimicry model fits well with the hypothesis that *rfb* diversity in *Salmonella* is a result of diversifying selective pressure arising from grazing intestinal protozoa. For *Salmonella* to survive and replicate within a host intestine, it must first avoid being consumed by native gut amoebae. The population of native amoebae in a given host intestine are presumably adapted to the environmental conditions within that particular gut; it is reasonable to propose that the population of amoebae in a host intestine may generally avoid grazing upon the intestinal mucins present in that environment. Thus, the *Salmonella* serovars that have a high fitness in a given host may do so because their O-antigens conceal these cells against the background of the host intestinal mucins, making them less susceptible to predation than other serovars with different O-antigens. Because populations of amoebae are adapted for the conditions within each host's intestine, an O-antigen that provides a high degree of protection from predation in one gut environment may provide little or no protection against predation in another gut environment in which the nature of the intestinal mucins differ.

5.5.2 Potential trade-offs may exist between interactions with protozoan predators and the host immune system

Although my experiments on O-antigen chain length further point to the role of the O-antigen to fitness against protozoan predation, the question remains as to why *Salmonella* produce O-antigens of varying chain length. O-antigen chain length has been most extensively studied in *Salmonella* using serovars Typhimurium and Enteriditis; chain length is bimodal in wild type strains of these serovars [231-233]. *Salmonella* strains lacking appropriate O-antigen chain length are impaired in several critical steps in pathogenicity: complement resistance [107], cell invasion [233], and cell adhesion [241]. The presence of longer O-antigen chains, as mediated by

FepE and WzzB, are required for full virulence in the mouse model of infection, as cells mutated for both *fepE* and *wzzB* produce shorter, random-length O-antigen chains, have enhanced susceptibility to complement, and are more attenuated in the mouse model of infection [231]. Susceptibility to uptake by host phagocytic cells has also been associated with O-antigen chain length. In one study, *Salmonella Typhimurium* cells having longer O-antigens displayed reduced uptake by RAW264.7 macrophages [233], suggesting that longer O-antigens may help protect *Salmonella* from the actions of the host innate immune system.

This information led me to posit that a trade-off may occur between having protection against harsh environmental conditions and successfully invading cells (long O-antigens) and avoiding detection by protozoa/immune system molecules (short O-antigens to avoid many predators), with the chain-length distribution observed in *Salmonella* reflecting a balance between multiple selective pressures.

If a given serovar of *Salmonella* were always found in its preferred environment and if my hypothesis is true, then longer O-antigens would be favored for both predator evasion and successful interaction with the host immune system. Although longer antigens may protect *Salmonella* in some environments via molecular mimicry, *Salmonella* is not always found in an ideal environment. For example, serovar Dublin is more likely to cause disease in cattle than in chickens. However, it is unreasonable to propose that serovar Dublin only encounters the intestinal environment of cattle. When serovar Dublin passes through an intestinal tract of a different host, it may have reduced fitness against the native intestinal protozoa of this new host relative to those of cattle intestines. In this case, a stochastic distribution of varying O-antigen chain lengths may serve as a trade-off among selective pressures arising from the need to avoid protozoan predation, invade intestinal epithelial cells, and evade the host immune system

defenses during intracellular replication. While longer O-antigens better allow *Salmonella* cells to escape immune system defenses and survive harsh environmental conditions, they may also serve as targets for protozoan predators if the identity of that O-antigen confers a low fitness against predation.

6.0 O-ANTIGEN IDENTITY IS SUFFICIENT TO CONFER DIFFERENTIAL FITNESS OF *SALMONELLA* IN DIRECT COMPETITON AGAINST PROTOZOAN PREDATION

The results presented in Chapters 3 and 5 provided two lines of evidence that the O-antigen plays a major role in shaping *Salmonella* fitness against protozoan predation, a requisite to support the hypothesis explaining the maintenance of *rfb* diversity in *Salmonella*. In Chapter 3, I demonstrated that a) disruption of the O-antigen impacts fitness against predation; and b) protozoan predators discriminate among *Salmonella* strains that only vary at the O-antigen (*rfb* near-isogenic strains). Furthermore, I presented experiments in Chapter 5 which suggested that fitness against protozoan predation is a complex interaction between O-antigen chain length and identity. As discussed in Chapter 4, I developed significant technical improvements to competition tests using flow cytometry that permitted the direct competition of multiple strains in the presence and absence of predation, an approach that was used for all experiments in Chapter 5.

Critically, the experiments in Chapters 3 and 5 are not without limitations. First, the *rfb* near-isogenic strains were competed against predation in an indirect manner in Chapter 3. Second, while strains were competed directly in the presence and absence of amoebae in Chapter 5, neither these strains nor the strains used in Chapter 3 control for the variable presence of the H-antigen. Moreover, no previous experiment addressing *Salmonella* fitness against protozoan

predation utilized strains in which the presence or absence of the H-antigen was experimentally controlled [386-387]. Because the H-antigen is a major hypervariable cell surface antigen, its presence/absence must be controlled for in order to definitively establish the O-antigen as the major antigen shaping fitness against protozoan predation. To address these issues, I must a) directly compete a collection of *rfb* near-isogenic strains having wild-type O-antigen chain lengths against predation; and b) examine the fitness against predation of natural isolates of *Salmonella* in which the presence and absence of the O-antigen can be experimentally controlled. Here, I more rigorously examine the role of the O-antigen to *Salmonella* fitness against protozoan predation. Based on information presented in Chapters 3 and 5, I expect that O-antigen identity is sufficient to confer differential survivorship of *Salmonella* against protozoan predation irrespective of H-antigen identity.

6.1 MATERIALS AND METHODS

6.1.1 Media and growth conditions

All bacterial strains were routinely propagated as described in Chapter 2.2.1. When used alone, kanamycin was used at 20 μg/mL, tetracycline at 10 μg/mL, chloramphenicol at 20 μg/mL, spectinomycin at 250 μg/mL, hygromycin at 150 μg/mL, and ampicillin at 100 μg/mL. When used with multiple antibiotics, kanamycin was used at 10 μg/mL, tetracycline at 2.5 μg/mL, chloramphenicol at 10 μg/mL, spectinomycin at 100 μg/mL, hygromycin at 50 μg/mL, and ampicillin at 100 μg/mL. Non-Typhimurium strains used in these experiments were obtained from the *Salmonella* Reference Collection B (SARB) [35]; SARB strains 1, 3, 8, and 36 were

found to be defective for purine biosynthesis and require the addition of 0.008% guanosine to minimal media to support growth. Predation competition experiments were conducted using NM-C media, described in Chapter 4.4.2.

6.1.2 Strain construction of wild-type *rfb* near-isogenic strains for flow cytometry

Strains near-isogenic at *rfb* were constructed as described in Chapter 3.2.3 and were electroporated with single fluorescent protein tagging plasmids listed in Table 26 according to the methods described in Chapter 2.2.5. O-serotypes were verified by serotype agglutination. Strains are listed in Table 48.

Table 48. Wild-type *rfb* near-isogenic strains directly competed against protozoan predation to examine the role of O-antigen identity to fitness against predation using flow cytometry

Strain	SARB Parental Strain	O-serotype	Fluorescent Tag
KAB750	SARB2	(3,10)	pEGFP
KAB751			pDsRed-Express2
KAB752	SARB3	(1,4,12)	pEYFP
KAB753			pKAB9 (mKalama1)
KAB754	SARB20	(8,20)	pECFP
KAB755			pEYFP
KAB756	SARB36	(6,8)	pKAB9 (mKalama1)
KAB757			pEGFP
KAB758	SARB52	(1,9,12)	pDsRed-Express2
KAB759			pECFP

6.1.3 Construction of natural isolate *Salmonella* strains for flow cytometry

Bacteriophage P1 was used to engineer *galE*-6867::*aadA* derivatives of SARB strains that contained the markers *flhDC*-4820::*cat* and *fljBA*-4400::*tet*. The *flhDC*-4820::*cat* construct obliterates cellular control of the H-antigen through replacement of the flagellar master regulator [178]. To engineer strains for future experiments on the role of the H-antigen on fitness against predation, I introduced the *fljBA*-4400::*tet* construct into SARB recipient cells, removing flagellar phase variation in cells and forcing default flagellar production to the H1 antigen [3, 33, 115, 179, 394]. H-serotypic diversity is greatest for the *fliC*-encoded H1 antigen [213], and not all serovars display a serotype for the H2 antigen [106]. The Δ*rfbP-rfbB*-2773::*hph* construct was also introduced into strains in order to remove the presence of the O-antigen. The *galE*-6867::*aadA* marker was repaired using the strategy outlined in Chapter 2.4. Finally, each strain was electroporated with two different double fluorescent protein tagging vectors (created in Chapter 4.2.1). Derivatives of SARB1 containing the constructs *fljBA*-4400::*tet flhDC*-4820::*cat* and Δ*rfbP-rfbB*-2773::*hph* were created and tagged with appropriate vectors, but these strains were extremely sick in the presence of ampicillin and unable to grow on NM-C media; thus, these strains were eliminated from my experiments. Strains are listed in Table 49.

Table 49. Natural isolate *Salmonella* strains used to examine the role of the O-antigen to fitness against protozoan predation

Serotype key: underlining indicates phage conversion; brackets indicate variable expression.

SARB	Serotype		Genotype	Strain	Tagging Vector
	O-antigen	H1-antigen			
2	3,10	e,h	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB446	pKAB13
				KAB448	pKAB21
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB499	pKAB39
				KAB500	pKAB35
3	1,4,12	l,v	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB449	pKAB15
				KAB450	pKAB37
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB501	pKAB15
				KAB503	pKAB39
8	6,7	c	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB452	pKAB19
				KAB453	pKAB35
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB504	pKAB19
				KAB506	pKAB41
20	8,20	g,m,s	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB456	pKAB27
				KAB457	pKAB39
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB507	pKAB21
				KAB509	pKAB37
30	6,7	g,m,[p],s	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB459	pKAB21
				KAB460	pKAB11
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB512	pKAB11
				KAB511	pKAB21
36	6,8	e,h	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB461	pKAB35
				KAB462	pKAB19
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB513	pKAB35
				KAB515	pKAB13
54	11	r	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB464	pKAB37
				KAB465	pKAB15
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB516	pKAB37
				KAB518	pKAB19
55	1,4,[5],12	e,h	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB467	pKAB39
				KAB468	pKAB13
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB520	pKAB13
				KAB521	pKAB15
59	1,3,19	g,[s],t	*fljBA*-4400::*tet flhDC*-4820::*cat*	KAB471	pKAB11
				KAB472	pKAB27
			fljBA-4400::*tet flhDC*-4820::*cat* Δ*rfbP-rfbB*-2773::*hph*	KAB522	pKAB41
				KAB523	pKAB11

6.1.4 Competition experiments

Two sets of reciprocally tagged strains were grown overnight in liquid LB supplemented with ampicillin at 100 μg/mL. Cells were pelleted with centrifugation and resuspended in PBS; following resuspension, cells were mixed in approximately equal ratios based on OD_{600} measurements. OD_{600} measurements of final competition mixes ranged from 0.900 to 1.200; a 100 μL aliquot of the competition mixture was spread onto NM-C plates supplemented with ampicillin at 100 μg/mL. For PREDATOR plates, the centers of fourteen plates were seeded with 10^4 acid-base washed amoebae cysts; at least eight NO PREDATOR plates were set up for each experiment. Plates were incubated at 33°C for approximately 3.5 days to permit the feeding front of predation to reach the outer edges of the PREDATOR plates.

A total of six NO PREDATOR and six to ten PREDATOR plates from each experiment were chosen for analysis. Plates were eluted with 5 mL of PBS with 0.02% Tween20 added to reduce cell clumping. To diminish the number of amoebae present in samples, the eluate was filtered through two 5 μM pore size overlayed CellMicroSieve™ biologically inert nylon mesh filters (Bio-Design Inc. of New York). All samples were diluted in PBS + 0.02% Tween20 to an approximate concentration of 10^7 cells/mL.

6.1.5 Flow cytometry

Prior to analysis, SYTO™ 62 nucleic acid dye (Invitrogen, Carlsbad, CA) was added to a final concentration of 50 nM to aid in the separation of *Salmonella* cells from debris and amoebae. Events were acquired on a Beckman-Coulter CyAn ADP (Beckman-Coulter) flow cytometer at a rate of <1000 events per second. Filters and detector settings are listed in Chapter 4.

Commercially available software (Summit 4.3, Dako Colorado Inc.) was used for the operation of the cytometer. Data analysis was performed using Ferdinand (developed by Dr. Jeffery Lawrence as discussed in Chapter 4).

6.1.6 Technical acknowledgements

Strains listed in Table 48 were electroporated with fluorescent tagging plasmids under my supervision by undergraduate researchers Ben Cross and Mark Brown. I designed the primers used for directed gene replacement of constructs *fljBA*-4400::*tet* and *flhDC*-4820::*cat*; replacements were performed collaboratively by participants of the 2007 Gene Team at the University of Pittsburgh. Ferdinand was developed by Dr. Jeffrey Lawrence at the University of Pittsburgh.

6.2 *SALMONELLA* STRAINS THAT ONLY VARY AT THE O-ANTIGEN EXHIBIT DIFFERENTIAL FITNESS AGAINST PROTOZOAN PREDATION IN DIRECT COMPETITION

In Chapter 3.3 (Figure 15), I engineered a series of *rfb* near-isogenic strains of *Salmonella* that only varied at the O-antigen and competed these strains against three genetically distinct amoebae. While these experiments highlighted a relationship between O-antigen identity and fitness against predation, they relied on fitness assessment under conditions of indirect competition and measured using the crude method of line tests. Additionally, the work presented in Chapter 5 suggested that fitness against predation is a combinatorial interaction between O-

antigen chain length and identity, but I did not examine the fitness of the wild-type *rfb* near-isogenic strains against each other. Using the techniques developed in Chapters 4 and 5, I revisited the role of O-antigen identity in shaping fitness against protozoan predation as described below.

6.2.1 Predators directly discriminate among *rfb* near-isogenic strains

Although I previously demonstrated that amoebae can discriminate among strains that only vary at the O-antigen in Chapter 3.3, an argument could be made that these results were impacted by the intrinsic crude nature of line tests and the fact that strains were indirectly competed. To address these issues, I directly competed five *rfb* near-isogenic strains representing five distinct O-serotypes against predation (Table 48) against predation by three genetically distinct amoebae: *Naegleria gruberi* NL, *Acanthamoeba* sp. R2-1, and *Tetramitus* BD1-1. Fitness was assessed using the flow cytometry protocols discussed in Chapters 4 and 5. As shown in Figure 38, relative fitness of each *rfb* near-isogenic strain varies substantially among the three protozoan predators, with fitness differences between the most fit and least fit strains occurring along a range from at least a two-fold minimum difference and ten-fold maximum difference. Not surprisingly, fluorescent tag fitness among experiments displays a similar level of robustness as to the data presented in Chapter 5. Here, fluorescent tag fitness values are very similar among experiments, with the average Pearson $r = 0.9702$ (Table 50) and Spearman $r = 0.7333$ (Table 51). Genotype fitness values show greater differences than fluorescent tag fitness, with the average Pearson $r = 0.1540$ (Table 52) and Spearman $r = 0.2333$ (Table 53). All fitness values were supported by bootstrapping the entire data population at 1000 recalculations.

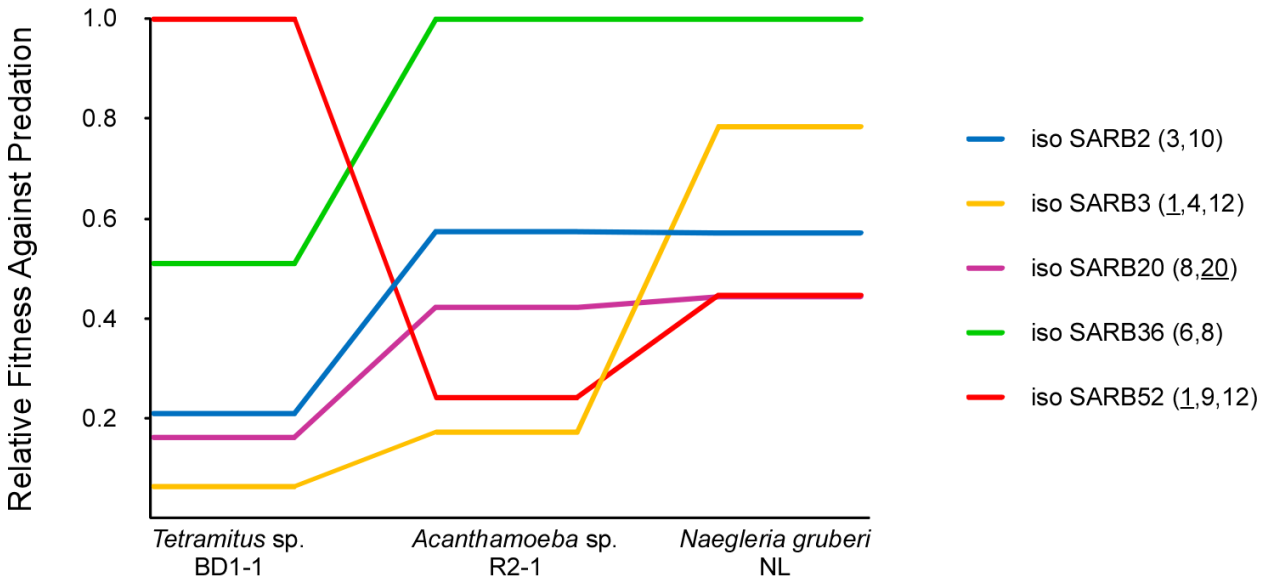

Figure 38. Protozoan predators discriminate among *rfb* near-isogenic *Salmonella* prey in direct competition

Raw genotype fitness values were deconvoluted from raw strain fitness values, which were calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). For each experiment, the genotype having the greatest fitness contribution was assigned a value of 1; the fitness values of all other genotypes in that experiment were normalized to this value. O-serotypes are shown in parentheses. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Table 50. Pearson correlation coefficients for pairwise comparisons of fluorescent tag fitness in competition experiments of five *rfb* near-isogenic strains against predation by three different amoebae

Darker blue colors indicate strength of correlation. Average Pearson $r = 0.9702$.

	***Tetramitus* sp. BD1-1**	***Acanthamoeba* sp. R2-1**	***Naegleria gruberi* NL**
***Tetramitus* sp. BD1-1**	1	0.9748	0.9893
***Acanthamoeba* sp. R2-1**	0.9748	1	0.9466
***Naegleria gruberi* NL**	0.9893	0.9466	1

Table 51. Spearman rank order correlation coefficients for pairwise comparisons of fluorescent tag fitness in competition experiments of five *rfb* near-isogenic strains against predation by three different amoebae

Darker blue colors indicate strength of correlation. Average Spearman $r = 0.7333$.

	***Tetramitus* sp. BD1-1**	***Acanthamoeba* sp. R2-1**	***Naegleria gruberi* NL**
***Tetramitus* sp. BD1-1**	1	0.9	0.6
***Acanthamoeba* sp. R2-1**	0.9	1	0.7
***Naegleria gruberi* NL**	0.6	0.7	1

Table 52. Pearson correlation coefficients for pairwise comparisons of genotype fitness in competition experiments of five *rfb* near-isogenic strains against predation by three different amoebae

Darker red colors indicate greater negative Pearson correlation coefficients; darker blue colors indicate greater positive Pearson correlation coefficients. Average Pearson $r = 0.1540$.

	***Tetramitus* sp. BD1-1**	***Acanthamoeba* sp. R2-1**	***Naegleria gruberi* NL**
***Tetramitus* sp. BD1-1**	1	0.0292	-0.1769
***Acanthamoeba* sp. R2-1**	0.0292	1	0.6098
***Naegleria gruberi* NL**	-0.1769	0.6098	1

Table 53. Spearman rank order correlation coefficients for pairwise comparisons of genotype fitness in competition experiments of five *rfb* near-isogenic strains against predation by three different amoebae

Darker blue colors indicate greater positive Spearman rank order correlation coefficients. Average Spearman $r = 0.2333$.

	***Tetramitus* sp. BD1-1**	***Acanthamoeba* sp. R2-1**	***Naegleria gruberi* NL**
***Tetramitus* sp. BD1-1**	1	0.4	0
***Acanthamoeba* sp. R2-1**	0.4	1	0.3
***Naegleria gruberi* NL**	0	0.3	1

If the O-antigen does not significantly contribute to fitness, I would expect that the fitness values of all *rfb* near-isogenic strains against all predators to have a very high degree of correlation, with correlation values of strain fitness between predators being very close to a value of 1. However, this is not the case as shown with both the Pearson correlation (Table 52) and the Spearman rank order correlation (Table 53). The information presented here allow me to reinforce the conclusions presented in Chapter 3 without most of the caveats used in previous fitness assays: because predators can discriminate among strains that only vary at the O-antigen and the fitness against predation for the *rfb* near-isogenic strains varied considerably among predators, the O-antigen must be a major determinant of protozoan feeding preference. Moreover, the experiments presented here address one of the major weaknesses of the *rfb* near-isogenic strain competitions presented in Chapter 3. Here, these strains were competed directly using a more sensitive method of fitness determination than line tests. Critically, these data indicate that amoebae can discriminate among strains that only vary at the O-antigen when strains were competed directly, further bolstering the arguments set forth in Chapter 3.3 through the use of more rigorous approaches to examine fitness against predation.

6.2.2 Relative fitness of *rfb* near-isogenic strains corroborated the role of O-antigen chain length in fitness against predation

The data discussed above support the observation I made in Chapter 5 indicating a link between O-antigen identity and chain length to fitness against protozoan predation and provide more robust evidence supporting the conclusion I drew in Chapter 3 that amoebae can discriminate among strains that only vary at the O-antigen by directly competing strains against protozoan predators. These same experiments can be used to examine the hypothesis that the O-antigen acts

as a molecular mimic of host intestinal mucins, potentially camouflaging *Salmonella* cells with particular O-antigens in the intestinal environment. For a given set of chain length mutant strains, the relationship of O-antigen chain length to fitness against predation occurred along a gradient depending on the identity of the O-antigen and of the protozoan predator (Figures 34 and 35). Although the results presented in Chapter 5 suggested that the nature of the O-antigen (chain length and identity) impact fitness against predation, I did not expect fitness to change on a linear fashion depending on chain length; a particular chain length class should not be necessarily more fit against predation in all experimental circumstances.

The relationship between O-antigen chain length and fitness against predation may be complicated by the potential unmasking of alternative receptors used for prey recognition in strains with shorter O-antigen chain lengths. Additionally, I did not expect *rfb* near-isogenic strains to display accurate phenocopies of parental O-antigens, as loci located outside of *rfb* may influence O-antigen identity (see Chapter 3 for further discussion). Table 54 presents two examples of the relationship between O-antigen chain length (from experiments discussed in Chapter 5) and identity (Chapter 6.2.1) for three different protozoan predators.

Table 54. Summary of O-antigen chain length fitness experiments and corresponding relative fitness of wild-type *rfb* near-isogenic strain against protozoan predation.

rfb Near-Isogenic Strain	Predator	Highest Relative Fitness in Chain Length Experiments: Genotype	Predicted O-antigen Chain Length	Relative Fitness When Competed Against Wild-Type *rfb* Near-Isogenic Strains
isoSARB2	*Tetramitus* sp. BD1-1	*fepE*-2910::*tet* *wzzB*-8771::*aph*	1-16	Moderate
	Acanthamoeba sp. R2-1	Wild-Type	16-35; 100+	Moderate
	Naegleria gruberi NL	*rfc*-2771::*cat*	1	Moderate
isoSARB3	*Tetramitus* sp. BD1-1	*rfc*-2771::*cat*	1	Low
	Acanthamoeba sp. R2-1	*fepE*-2910::*tet*	16-35	Low
	Naegleria gruberi NL	*wzzB*-8771::*aph*	16-35; bias towards 100+	High

Interestingly, the wild-type *rfb* near-isogenic strain derivative of SARB2 (strains KAB750-51) have a moderate relative fitness against predation against *Tetramitus* sp. BD1-1, *Acanthamoeba* sp. R2-1, and *N. gruberi* NL; however, both shorter and longer O-antigen chain length genotypes had the highest relative fitness against predation when competed against other O-antigen chain length derivatives. This further highlights the complex interaction between O-antigen chain length and identity in shaping *Salmonella* fitness against predation. Here, if fitness against predation was simply a function of general strength of binding of predators to O-antigens, I would have expected that universally shorter O-antigens would have higher fitness against predation regardless of the fitness of the wild-type parent strain, and this is not the case with derivatives of *rfb* near-isogenic strain SARB2.

The wild-type *rfb* near-isogenic strain derivative of SARB3 (KAB752-53) have a low relative fitness against predation by both *Tetramitus* sp. BD1-1 and *Acanthamoeba* R2-1; genotypes having shorter O-antigen chain lengths than wild-type had the highest relative fitness in the chain length experiments. However, the opposite was observed for predation by *N. gruberi* NL: the wild-type strain had a high fitness against predation, and the *wzzB*-8771::*aph* genotype was most fit relative to other chain length classes against predation by this amoeba. In this case, the *wzzB*-8771::*aph* strain may exhibit a bias towards the 100+ O-antigen chain length class [231-232]. Thus, the data presented here suggest that this O-antigen identity may be avoided by *N. gruberi* NL, as the wild-type strain derivative had a relatively high fitness against predation when competed against strains only differing at the O-antigen and a longer O-antigen chain length class had the highest fitness against predation relative to strains only differing in O-antigen chain length.

As illustrated in Chapter 5, differences in O-antigen chain length are sufficient to enable differential survivorship against protozoan predation. As shown in these experiments, differences in O-antigen identity are also sufficient to permit differential fitness against protozoan predation. Critically, both of these experiments were conducted using direct strain competition, eliminating the explanation that the fitness differences observed in Chapter 3.3 were due to other factors such as varying environmental conditions between replicate line test plates or factors impacting strain-to-strain competition. Taken together, these data would argue for a very strong role played by the O-antigen, and hence the *rfb* locus, in shaping fitness against protozoan predation, more confidently supporting the hypothesis that predation from grazing amoebae drives diversifying selection at the *Salmonella rfb* locus. Moreover, I provide further evidence here to support the hypothesis that the O-antigen serves as a type of molecular camouflage against host intestinal mucins; protozoan predators avoid grazing upon bacterial cells that display polysaccharide antigens that resemble the host environment.

6.3 PROTOZOAN PREDATORS DISCRIMINATE AMONG *SALMONELLA* PREY LACKING THE H-ANTIGEN

As discussed above, protozoan predators discriminate among strains that only vary at the *rfb* locus in direct competition, strongly suggesting that O-antigen identity is sufficient to permit differential susceptibility to protozoan predation in *Salmonella*. However, these experiments raise the fundamental question of how the O-antigen contributes to fitness against predation when it is presented in the context of the full spectrum of naturally-occurring surface antigen diversity of wild *Salmonella* isolates. Not only do natural *Salmonella* strains differ at the O-

antigen, they also differ at the H-antigen and potentially at hundreds of other cell surface antigens (see Chapter 1.5 for additional discussion). I first addressed this issue in Chapter 3.2 using the *rfb* near-exogenic strain collection, in which the *rfb* loci of strains from the SARB collection were replaced with that of the laboratory strain Typhimurium LT2, and indirectly competed these strains against protozoan predation. These results showed that disrupting the native *rfb* locus in the context of all other strain-specific *Salmonella* surface antigens led to dramatic changes in fitness against protozoan predation, highlighting the strong role played by the O-antigen in shaping differential survivorship against predation. However, this technical approach had two important caveats: a) indirect competition; and b) lack of experimental control of the H-antigen.

I address these caveats here using more sophisticated technical approaches in both assessment of fitness against predation and in strain construction. First, all competitions were performed directly using flow cytometry; double fluorescent-protein tagging vectors permitted the simultaneous competition of up to ten strains. Second, I introduced several key directed gene replacements into the SARB strains that permitted the experimental control of the H-antigen flagella. As discussed in Chapter 1.5.1, the H-antigen flagellae comprise the other hypervariable surface antigen in the Salmonellae. I did not control for the presence of H-antigen expression nor the random H-antigen phase variation known to occur in *Salmonella* in previous experiments, and it is possible that variability in environmental conditions among experiments could have fostered expression of H-antigens in some experiments but not in others, confounding results.

To deal with the H-antigen issue, I used the genetic techniques described in Chapter 2 to engineer a series of natural isolates of *Salmonella* that lack the ability to produce either phase of the H-antigen via introduction of the *flhDC*-4820::*cat* construct. The *flhDC* operon serves as the

flagellar master regulator, which is transcribed in response to certain environmental conditions and regulates the transcription of the flagellar biosynthesis operons; see Chapter 1.5.1 for a brief review on flagellar biosynthesis. Because these cells lacked the ability to transcribe flagellar biosynthesis genes, the H-antigen was absent from these cells under the environmental conditions of the experiments outlined in this section.

If the O-antigen strongly contributes to differential survivorship of *Salmonella* against grazing amoebae, then I expect natural strains lacking the H-antigen to have differential fitness against protozoan predation. To eliminate the H-antigen as a potential variable explaining discrimination of *Salmonella* prey by protozoan predators, I tested the fitness of nine different *Salmonella* strains lacking the ability to produce an H-antigen (Table 49) against predation by three different amoebae: *Naegleria gruberi* NL, *Acanthamoeba polyphaga* I, and *Hartmannella* sp. T3-1 using simultaneous competition assays and measured strain abundance in the presence and absence of predators using flow cytometry. If the O-antigen is a major contributing factor to prey recognition, then I should observe discrimination among O+ H- prey. In the unlikely case that the H-antigen is the primary surface antigen impacting fitness against predation, then I should observe a lack of discrimination among these genetically altered *Salmonella* prey, or the degree of discrimination will be far smaller than that observed for strains bearing H-antigens.

Fitness was examined in the same manner as experiments described in Chapters 5 and 6.1, except that positive events were defined as those having positive signal for two different fluorescent signals. Curves were fit to Gaussian distributions of fluorescent signals defined by the unique signal distributions of single double-labeled cells grown under identical experimental conditions. For example, to define events labeled with both EYFP and ECFP, curves were fit to the Gaussian distribution of events having positive EYFP signal that also had positive ECFP

signal and *vice versa*. Although I examined derivatives of SARB1 in prior work, derivatives of SARB1 were eliminated from the this strain set due to extremely low tolerance of ampicillin at concentrations required for maintenance of the tagging vector.

Here, I present strain fitness values rather than their deconvoluted components of genotype and fluorescent tag fitness values for several important reasons. First, the deconvolution process perpetuated artifactual error in cases when strain counts were low, which I observed in various experiments. Here, while the variety of genetic mutations in the natural isolate strains used here provided a much more rigorous level of experimental control of surface antigen presentation, they also resulted in some strain growth defects in the presence of antibiotic selection. I observed varying levels of sickness among strains grown under the required amount of antibiotic selection required for maintenance of the fluorescent tagging vectors, which led to underrepresentation of various strains in no predator and predator competition plates due to poor antibiotic tolerance. Low numbers of certain strains would perpetuate error in the deconvolution process, which could significantly impact genotype and fluorescent tag fitness values due to this artifact. Second, the goal of these experiments is to determine if predators discriminate among natural *Salmonella* prey lacking the H-antigen; I am only interested in determining if discrimination occurs, not the exact hierarchies of the strain fitness values. Third, I am less concerned about fluorescent tag fitness here than in previous work, as I already demonstrated in Chapters 5 and 6.2 that fluorescent tag fitness values are strongly consistent among experiments. For these reasons, the presentation of strain fitness against predation provides the most straightforward approach to examining the question if natural *Salmonella* strains lacking H-antigens have differential fitness against protozoan predation.

Figure 38 depicts the relative log strain fitness of the O+ H- strain collection against predation from three different amoebae. Consistent with all previous observations, all amoebae tested were able to discriminate among natural isolates of *Salmonella* lacking the H-antigen. If the O-antigen plays a stronger role in shaping fitness against predation than the H-antigen, then strains lacking the H-antigen should exhibit differential survivorship against protozoan predation. Conversely, if the H-antigen is the major cell surface antigen shaping fitness against predation, then its removal should result in reduced differential survivorship against protozoan predation. As shown in Figure 38, *N. gruberi* NL, *A. polyphaga* I, and *Hartmannella* sp. T3-1 all discriminated among strains of *Salmonella* lacking the H-antigen. Moreover, fitness differences were so extreme in some cases that I transformed these values to logarithmic values to simplify data presentation. Because amoebae discriminate among natural *Salmonella* prey lacking the H-antigen, I conclude that the O-antigen is a major surface antigen affecting fitness against predation.

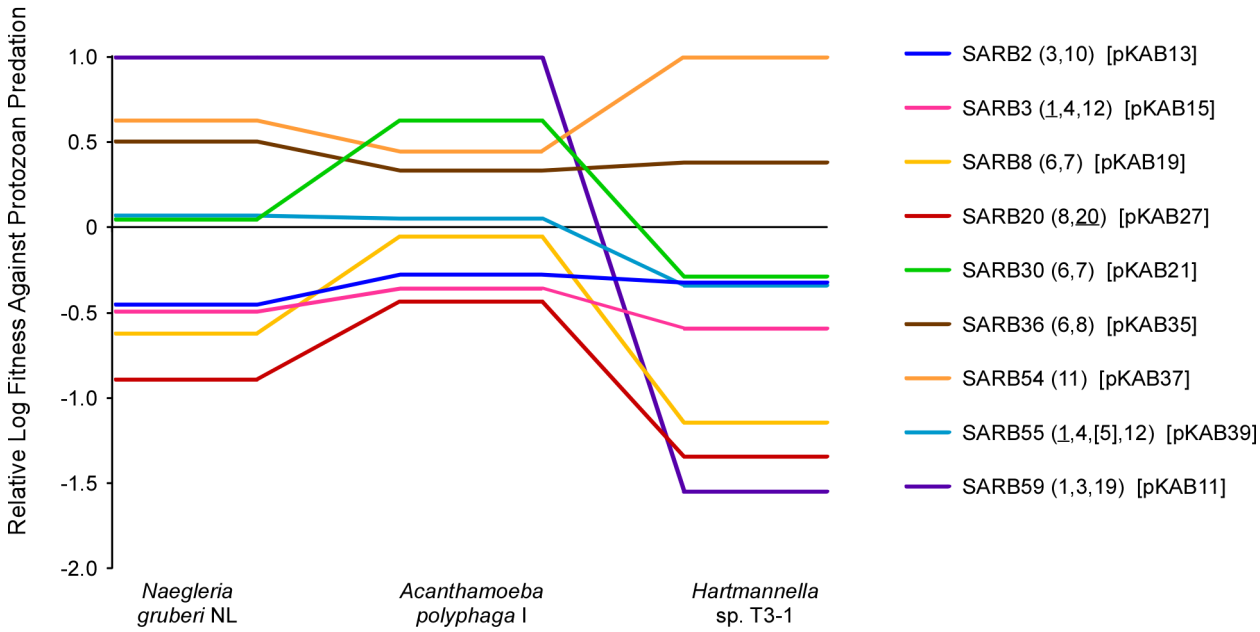

Figure 39. Natural isolates of *Salmonella* lacking the H-antigen exhibit differential fitness against protozoan predation

Strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Strain fitness values were log transformed for ease of presentation, as fitness differences were observed on the hundred- to thousand-fold scale. For each experiment, the strain having the greatest log fitness value was assigned a value of 1; all other strains were normalized to this value. O-serotypes are listed in parentheses; fluorescent tagging vectors are listed in brackets. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations. Descriptions of tagging vectors are listed in Table 11.

These experiments provide the most rigorous evidence to date that the O-antigen plays a very strong role in shaping *Salmonella* fitness against protozoan predation, supporting the hypothesis that *rfb* diversity in *Salmonella* is driven by diversifying selection from protozoan predators. Critically, this approach advances beyond other experiments presented previously by others [386-387] or in this dissertation because the variable presence of the H-antigen was directly addressed. Moreover, strains were competed directly, eliminating the possibilities that environmental conditions or intrastrain competition elements contributed to fitness in a greater manner than did the identity of the O-antigen.

6.4 PROTOZOAN PREDATION IS A LIKELY DRIVER OF DIVERSIFYING SELECTION AT THE *SALMONELLA RFB* LOCUS

If the O-antigen is a strong contributor to *Salmonella* fitness against predation, then its disruption should strongly impact the fitness against predation of any given strain. This issue was tested in Chapter 3.2, albeit using the crude approaches of line tests to measure fitness and the *rfb* near-exogenic strains, which consisted of a collection of natural isolates of *Salmonella* containing the *rfb* locus derived from serovar Typhimurium LT2. Using these data, I concluded that the O-antigen must play a major role in shaping fitness against protozoan predation because the fitness of wild-type natural *Salmonella* isolates differed widely from that of the *rfb* near-exogenic counterparts against predation by three different amoebae. However, this approach again suffers the caveats that a) competition was performed indirectly; b) the H-antigen was not experimentally controlled; and c) differences in the O-antigen among the *rfb* near-exogenic

strains series may exist due to inaccurate phenocopying of the LT2-derived *rfb* locus in non-Typhimurium genetic backgrounds.

Here, I directly address the fundamental weaknesses of experiments using the *rfb* near-exogenic strain collection in three major ways. First, strains were competed directly using flow cytometry to enumerate strains competed in the presence and absence of protozoan predators as discussed in Chapter 6.2. Second, I engineered natural isolate derivatives of *Salmonella* lacking the genes required to manufacture the H-antigen, thus placing the absence of this antigen under strict experimental control. Third, rather than creating a series of strains all having the same *rfb* locus, I removed the O-antigen entirely from a collection of SARB strains also lacking the H-antigen; phenocopying of an antigen using genes from another strain was not an issue in these new strains lacking both the major *Salmonella* cell surface antigens. In this case, I removed the major surface antigen conferring fitness against protozoan predation, permitting exposure of a wide variety of other surface antigens that vary among strains. Here, it is possible that removal of these antigens will unmask alternative receptors recognized by protozoan predators. Thus, I do not expect that fitness will remain stable among all strains lacking an O-antigen, although the degree to which strains are discriminated is unclear.

Using this more robust approach, I competed nine different *Salmonella* strains lacking both the O- and H-antigens against predation by three genetically distinct amoebae: *N. gruberi* NL, *A. polyphaga* I, and *Hartmannella* sp. T3-1 and compared the measured strain fitness values to those of the O+ H- strains (Figure 40). Again, fitness values were log transformed and normalized to aid in data presentation of widely differing fitness values within the same experiment. All predators discriminated among the O- H- strain series, and fitness hierarchies changed between the O+ H- and O- H- strain series to different degrees, allowing me to deduce

that the O-antigen strongly shapes fitness against protozoan predation. These data provide more robust support for the hypothesis that *Salmonella rfb* diversity is driven by diversifying selection from protozoan predation using an approach free of the major caveats of the data discussed in Chapter 3.

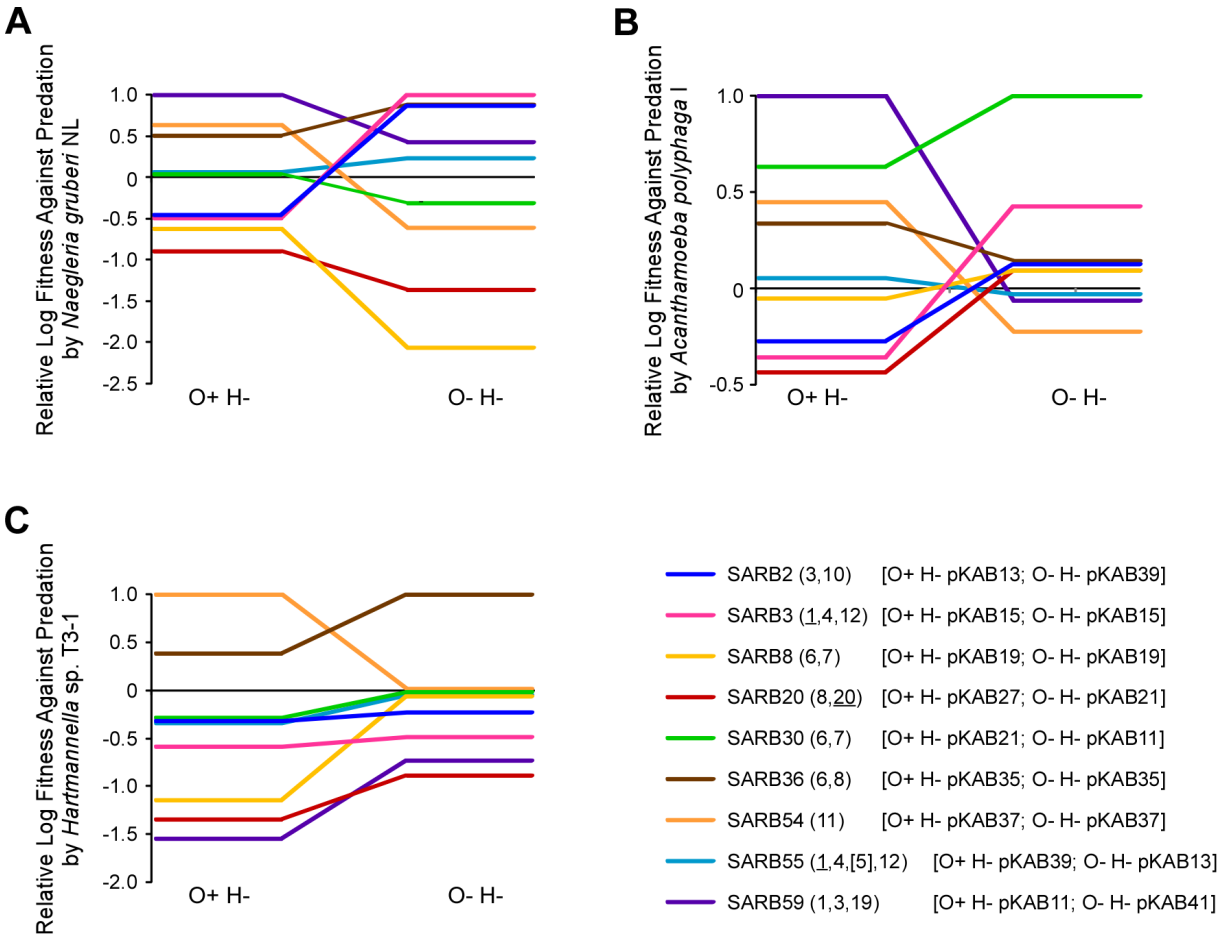

Figure 40. *Salmonella* strains lacking the O- and H-antigens have differential survival against protozoan predation.

Strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Strain fitness values were log transformed for ease of presentation, as fitness differences were observed on the hundred- to thousand-fold scale. For each experiment, the strain having the greatest log fitness value was assigned a value of 1; all other strains were normalized to this value. O-serotypes are listed in parentheses; fluorescent tagging vectors are listed in brackets. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

Interestingly, the O- H- strain series showed a similar overall degree of difference between the most fit and least fit strains for all amoebae tested (Figure 40). Here, it appears that protozoan predators are able to discriminate among *Salmonella* prey lacking both the major cell surface antigens. Two models can be used to explain this observation. First, removal of the O-antigen may result in the unmasking of alternate receptors used by amoebae to recognize food, and these receptors are not normally used in food recognition due to steric masking by the large O-antigen. Second, these receptors could be used to varying degrees to recognize prey in the presence of the O-antigen, but these contributions are extremely minor in light of the very strong role played by the O-antigen in shaping fitness against predation. Here, removal of the O-antigen may simply highlight the contributions of other cell surface antigens to prey choice. Because *Salmonella* strains potentially differ at hundreds of antigenically distinct outer membrane proteins (see Chapter 1.5 for further discussion), it is not surprising that strains lacking both the major cell surface antigens did not impact the ability of amoebae to differentially discriminate among *Salmonella* prey. However, in this case, the mechanism by which amoebae recognize *Salmonella* prey lacking the major cell surface antigens is unclear.

To examine the significance of difference between fitness hierarchies of O+ H- and O- H- *Salmonella* prey, I plotted relative log strain fitness of O+ H- strains vs. relative log strain fitness of O- H- strains and determined the correlation values (R^2) values of fitness hierarchies (Figure 41). For all three amoebae tested, prey is selected entirely differently when the O-antigen is absent, and this difference in prey selection occurs along a gradient of magnitude. Both *N. gruberi* NL ($R^2 = 0.1443$) and *A .polyphaga* I ($R^2 = 0.0011$) select prey entirely differently when the O-antigen is absent. However, *Hartmannella* sp. T3-1 displays a much lower degree of difference of prey selection in the absence of the O-antigen ($R^2 = 0.4887$). These data would

argue that different amoebae may use the O-antigen to varying degrees in prey selection, a trend that I also observed for the data presented in Chapter 3.2.

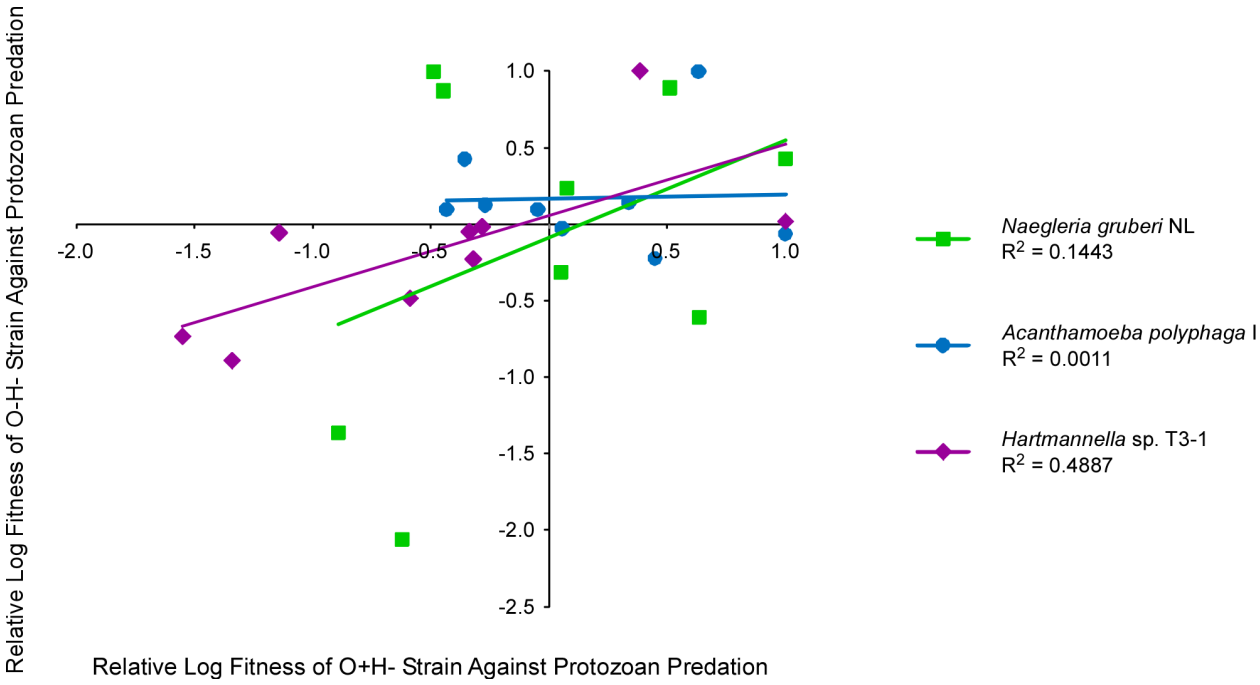

Figure 41. Removal of the O-antigen dramatically impacts fitness against protozoan predation of natural *Salmonella* strains in which the presence of the H-antigen is experimentally controlled.

Strain fitness was calculated as the deviation in observed cell frequency (that on predator plates) from the expected frequency (no predator plates). Strain fitness values were log transformed for ease of presentation, as fitness differences were observed on the hundred- to thousand-fold scale. For each experiment, the strain having the greatest log fitness value was assigned a value of 1; all other strains were normalized to this value. Fitness values were supported by bootstrapping events sampled from the total population with replacement with 1000 recalculations.

While the R^2 value for fitness of O+ H- vs. O- H- strains appears relatively low for *Hartmannella* sp. T3-1 when compared to those of *N. gruberi* NL and *A. polyphaga* I, this value still indicates a major contribution of the O-antigen to fitness against predation. The O-antigen is likely responsible for roughly 50% of the observed fitness against predation for serovars competed against this particular amoeba in light of the fact hundreds of other outer membrane surface antigens present in *Salmonella* may be used by *Hartmannella* sp. T3-1 in prey selection. Both the data presented here and in Chapter 3.2 suggest that the O-antigen is the major surface antigen imparting fitness against protozoan predation, although the degree of this contribution is dependent on the genotype of the protozoan predator. Although fitness is a complex interaction between *Salmonella* prey and amoebae predators, the O-antigen is the major element defining this predator-prey relationship.

6.5 THE O-ANTIGEN SHAPES THE PHYSIOLOGICAL BASIS FOR PREDATOR ESCAPE IN *SALMONELLA IN VITRO* AND FUTURE DIRECTIONS

Here, I demonstrated with a high degree of rigor three key elements that support the hypothesis that the O-antigen plays a strong role in the physiological basis of predator escape. First, O-antigen identity is sufficient to cause differential survivorship against protozoan predation. Second, O-antigen identity in the absence of any contribution from the H-antigen impacts *Salmonella* fitness against protozoan predation. Third, removal of the O-antigen dramatically impacts *Salmonella* fitness against predation. Given the results presented in this chapter and in all previous chapters in this dissertation, I conclude with confidence that the O-antigen is the major cell surface antigen contributing to *Salmonella* fitness against protozoan predation *in vitro*.

Additionally, the experiments presented here and in Chapter 5 shed light upon the mechanistic basis for prey recognition by protozoan predators, with the fitness of any given serovar being dependent upon O-antigen identity and the degree to which the particular predator uses the O-antigen in prey recognition.

Although absence of the H-antigen is not sufficient to eliminate discrimination of natural isolates of *Salmonella* by protozoan predators, the possibility remains that the H-antigen may play some role in shaping fitness against predation, albeit a potentially weaker role than that played by the O-antigen. Additionally, these experiments cannot rule out any synergistic effect between the O- and H-antigens in mediating predation escape in *Salmonella*. To address these issues, I designed a series of double-fluorescent protein tagging vectors containing *flhDC* under an arabinose-inducible promoter derived from pBAD24 [164]. Introduction of these vectors into the O+ and O- strains listed in Table 49 will yield a collection of strains in which the presence of the hypervariable *fliC*-encoded H1-antigen is experimentally controlled. Further competition assays using the H+ collection of strains will yield additional information on the role of the H-antigen alone and any synergistic relationships of the O- and H-antigens to predation escape in *Salmonella*.

These results provide compelling evidence that predation escape in *Salmonella* is mediated by the O-antigen, but all results presented up to this point were obtained from *in vitro* experiments. The next step in elucidating the physiological basis of *Salmonella* predator escape was to move beyond laboratory-controlled *in vitro* experiments using cultivable protozoa to determine if the presence of amoebae predators in can mediate *Salmonella* survival in natural intestinal environments. If protozoan predation truly drives diversifying selection at the

Salmonella rfb locus, then discrimination of prey based solely on the O-antigen must occur *in vivo*. I explored this issue further in Chapter 7.

7.0 DIFFERENTIAL SURVIVORSHIP OF *SALMONELLA* WITHIN ENTERIC ENVIRONMENTS

I demonstrated in Chapters 3, 5, and 6 that protozoan predators discriminate among *Salmonella* based on a complex interaction between O-antigen identity and chain length, suggesting that the O-antigen plays a major role in shaping *Salmonella* fitness against protozoan predation. Collectively, these data provide further support for the hypothesis that protozoan predators are driving diversifying selection at the *Salmonella rfb* locus. However, these experiments were all performed *in vitro*; for this hypothesis to be true, predators must discriminate among *Salmonella* prey based on the O-antigen *in vivo*. Regardless of how compelling laboratory competition experiments are, they still examine predator-prey interactions in an isolated, artificial environment. Thus, the next step in establishing protozoan predation as a selective force driving diversifying selection at the *Salmonella rfb* locus is to move beyond *in vitro* assays and compete bacteria within their native, enteric environment. Results from Wildschutte and Lawrence [386] established that predators isolated from the same environment share feeding preferences, suggesting that antigenically distinct *Salmonella* may experience differential survivorship when facing multiple predators. However, these experiments examined only cultivated predators.

Here, I discuss a preliminary *in vivo* competition experiment and future *in vivo* and *ex vivo* experiments designed to test my hypothesis. I chose the goldfish *Cassius auratus auratus* as a model *in vivo* system for two main reasons: a) I was confident that goldfish intestines are hardy

reservoirs of protozoa, as were successful at isolating amoebae from goldfish in many previous experiments [386] and b) goldfish are easily reared in the research laboratory setting and have regular eating habits, allowing ready manipulation of their enteric flora and fauna.

7.1 MATERIALS AND METHODS

7.1.1 Media and growth conditions

Bacteria were routinely propagated according to the protocols described in Chapter 2.2.1. Zeomycin was used at a final concentration of 100 μg/mL. Gentamycin was used at a final concentration of 50 μg/mL. No carbon source E media (or NCE media, derivative of E salts [371] lacking carbon source) was prepared with 0.2% galactose supplemented with 0.1 mM IPTG and 40 μg/mL XGAL.

7.1.2 Strain construction

Using the protocol described in Chapter 2.2.3, bacteriophage P1 was used to transduce the following markers into *galE*-6866::*aph* strains of SARB20 and 30: Δ*orgC*-*prgH*-5801::EYFP-*zeo*; *phs*-209::Tn*10d*Gn and *pdu*-219::Tn*10d*Cm. The construct *pdu*-214::Tn*10d*Tc was transduced into the SARB30 derivative strain KAB146 to serve as a tagging marker to permit culture-based differentiation from the SARB20 derivative KAB144.

7.1.3 Care of goldfish

Sixteen goldfish (*Cassius auratus auratus*) were obtained from Carolina Biological Supply; fish were divided into two groups of eight and added to one of two 15 gallon tanks. Two liters of water were exchanged between tanks daily; 10% of water was changed daily per tank. Fish were offered NutraFin® Max Goldfish Medium-Size Pellet Food (Rolf C. Hagen, Inc.) twice daily. Tank temperature was gradually raised to 30°C in both tanks in 0.5°C increments every 48 hours. Fish were acclimated to laboratory conditions approximately one month prior to experimentation. A protocol for proper care and use of goldfish was filed and approved with the University of Pittsburgh, Division of Laboratory Animal Resources (protocols 1005246, 0705497, and 0705498).

7.1.4 *In vivo* competition assay

Strains KAB144 and KAB146 were grown to approximately equal OD_{600} values in 5 mL LB broth supplemented with gentamycin and zeomycin. Cells were concentrated with centrifugation and rinsed twice with 0.9% NaCl; final volume for each culture was 1 mL. Cells were mixed together and added to a petri dish containing 2.0 g of NutraFin® Max Goldfish Medium-Size Pellet Food (Rolf C. Hagen, Inc.); pellets were allowed to completely soak up bacteria. Eight hours prior to inoculation of fish with bacteria, one tank (NO PREDATOR) was treated with 500 mg of metronidazole to reduce the number of protozoa. This tank was dosed every eight hours with 500 mg of metronidazole throughout the entire experiment. The other tank (PREDATOR) was untreated with antiprotozoal medication. Approximate strain concentration in inoculum was determined by serial dilution and replica printing for antibiotic resistance phenotypes.

In twelve hour intervals until 48 hours post inoculation, two fish per tank were removed and placed in separate beakers of water. Fish were euthanized with the addition of approximately 0.5 g of the aquatic anesthetic tricaine (ethyl 3-aminobenzoate methanesulfonate salt); to ensure complete euthanization, fish were dissected approximately five minutes after cessation of breathing and swimming behavior. Intestines were removed from fish by making a small cut ~0.5 cm behind the gill on the ventral side and cutting along the ventral side to the anus. Care was taken to avoid rupture of the intestines during dissection.

Following dissection, the intestines of each fish were placed in separate sterile petri dishes to permit sectioning of intestines and removal of intestinal contents, which were resuspended in 1 mL of 0.9% NaCl and transferred to sterile microcentrifuge tubes. To separate bacteria from larger cells and debris, intestinal contents were spun for 20 minutes at 2000 rpm at room temperature; supernatant was serially diluted in 0.9% NaCl to 10^{-12}. To enumerate experimentally introduced *Salmonella*, dilutions 10^{0} to 10^{-6} were plated to LB Zeo_{100} Gn_{50}; to enumerate all bacteria, dilutions 10^{0} to 10^{-12} were plated to LB. All plates were incubated overnight at 30°C.

Following incubation, colony forming units on all plates were counted. LB Zeo_{100} Gn_{50} plates were replica printed to NCE galactose XGAL IPTG, LB Zeo_{100} Gn_{50}, LB Cm_{20}, LB Tc_{20}, and LB. Replica print plates were incubated overnight at 37°C and colony forming units were enumerated. The expected phenotypes for experimentally introduced *Salmonella* are: galactose^{+}, β-galactosidase^{-}, zeomycin^{R}, gentamycin^{R}, chloramphenicol^{R}, and depending on tag of strain, tetracycline^{R}. For further validation, subsets of colony forming units scored positive for experimentally introduced *Salmonella* were subjected to colony PCR verification for the presence of the Δ*orgC-prgH*-5801::EYFP-*zeo* construct.

7.2 DIFFERENTIAL SURVIVORSHIP AMONG NATURAL ISOLATES OF *SALMONELLA IN VIVO*

To investigate the feasibility of using goldfish as a model host for competition experiments, I constructed two differentially-tagged *Salmonella* SARB strains using bacteriophage P1 transduction. Here, I do not want experimentally introduced strains to invade and replicate within host cells; I was only investigating the potential for these strains to exhibit differential survival in fish intestines in the presence and absence of native host protozoa. Thus, both strains were rendered nonpathogenic by the addition of Δ*orgC-prgH*-5801::EYFP-*zeo*, a directed gene replacement of *Salmonella* pathogenicity island I with the zeomycin resistance cassette and the EYFP gene. To ensure discrimination of *Salmonella* among bacteria native to the goldfish intestine, I added the markers *phs*-209::Tn*10d*Gn and *pdu*-219::Tn*10d*Cm, which confer resistance to both gentamycin and chloramphenicol to experimental strains KAB144 (SARB20 derivative) and KAB146 (SARB30 derivative). Finally, KAB146 contains the marker *pdu*-214::Tn*10d*Tc, tagging this strain with tetracycline, enabling its discrimination from KAB144.

While these strains harbored a fluorescent tag, I did not enumerate strains using flow cytometry. Because the *Salmonella* introduced represent a minor constituent of the enteric environment, strains introduced experimentally must be recovered by affinity separation using an antibody recognizing the enteric core antigen. However, the effects of O-antigen identity in modulating the efficacy of this antibody are still being explored. To remove this variable, I chose to select for strains based on a combination of antibiotic tags and validate their identity using biotypic markers.

These strains were competed in goldfish intestines in the presence and absence of protozoa via treatment with the antiprotozoal drug metronidazole (Figure 36). The percentage of

the SARB30 derivative strain KAB146 of total recovered experimental strains did not significantly differ between that of the initial inoculum and in fish treated with metronidazole (NO PREDATOR), suggesting no net change between the introduced and recovered strains between these two groups. However, I did observe a significant ($P < 0.05$, *t*-test) decrease in the percentage of KAB146 recovered from fish untreated with metronidazole (PREDATOR group); while *Salmonella* bacteria recovered from the inoculum and NO PREDATOR groups consisted of 55-60% KAB146, total *Salmonella* recovered from the PREDATOR group contained approximately 35% KAB146. This suggests that the presence of protozoa mediated differential survivorship of KAB146 (SARB30) relative to KAB144 (SARB20) in this experiment. Thus, I conclude that protozoan predators may influence *Salmonella* survival *in vivo*.

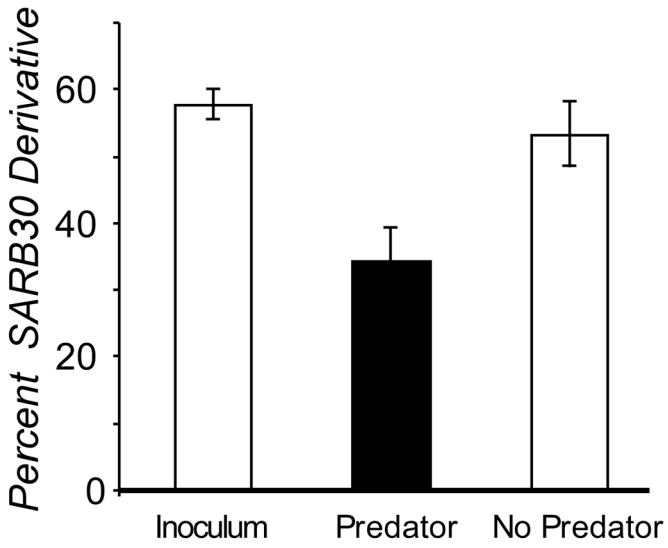

Figure 42. Predator-mediated survival of *Salmonella* within fish

Derivatives of two antigenically-distinct serovars of *Salmonella*, SARB20 (8,20:g,m,s:--) and SARB30 (6,7:g,m,[p],s:[1,2,7]) were introduced into goldfish. The proportion of the SARB30 derivative was assessed in the food inoculums and in the intestinal contents of metronidazole-treated and untreated goldfish after 24-48 hours incubation. Antigenic serotypes are listed as (O:H1:H2).

7.3 FUTURE *IN VIVO* AND *EX VIVO* EXPERIMENTS

The work presented here point to the feasibility of performing *in vivo Salmonella* competition tests to further test the link between protozoan predation and diversity at the *rfb* locus. Although I demonstrated here that protozoan predators can discriminate among natural isolates of *Salmonella in vivo*, these experiments need to be much more carefully controlled. The strains used here are natural isolates of *Salmonella*; they vary at many other loci that can affect *in vivo* survival. A more thoroughly controlled experiment should compete strains that vary only at the O-antigen; thus the *rfb* near-isogenic strains constructed in Chapter 3 should be used in order to assess the contribution to *Salmonella* fitness against protozoan predation *in vivo*. If differential survival is mediated by the O-antigen, then I should observe discrimination of the *rfb* near-isogenic strain collection in *in vivo* experiments. Additionally, the full spectrum of O-antigenic diversity must be examined in future *in vivo* experiments. The approach of antibiotic cassette tagging of strains to discriminate among strains as used here is too crude to examine more than a few strains. Therefore, the flow cytometry technique I developed in Chapter 4 can be adapted for use in *in vivo* competition experiments of up to ten different strains simultaneously. In this way, I can rapidly assess the contribution of the O-antigen to survival against predation in a natural environment with data of greater statistical power.

Although I chose goldfish due to their ease in experimental manipulation, they are not natural hosts of *Salmonella.* Use of hosts in future experiments should be selected with care, as animals must be untreated with any type of antibiotic medication, as these treatments disrupt the native flora of the host.

To overcome these issues, I propose conducting *ex vivo* experiments using the intestines of animals already euthanized for other purposes, including agriculture, research, or personal game hunting. Because no husbandry of animals is required, I will not require any special animal housing for *ex vivo* experiments; intestines will be harvested as soon as possible following euthanization, kept at constant appropriate temperature, and sectioned off for use in competition experiments. *Salmonella* cells will be prepared and injected into sealed sections of intestine, allowing the conduction of several experiments using the same intestinal material. Bacterial survivorship will be assessed using flow cytometry. Animals obtained through personal game hunting are essentially wild-caught; these animals should have no prior exposure to antibiotic agents. Using this approach, I can test a large number of variables in animal hosts that may influence *Salmonella* survival against protozoan predation under many conditions that may affect the population of predators in host intestines, including diet, geographic location, developmental stage, and season of capture.

The work presented in this dissertation addresses many of the challenges in examining differential survival of *Salmonella* as mediated by protozoan predators in natural intestinal environments. In Chapter 3, I developed the series of *rfb* near-isogenic strains that permit assessment of role of only the O-antigen to fitness against protozoan predation in natural environments. In Chapter 4, I discuss the implementation of multicolor flow cytometry for use in assessing the fitness of many *Salmonella* strains simultaneously, permitting the high-throughput collection of data representative of the full spectrum antigenic diversity within *Salmonella* subspecies I. Here, I demonstrate the feasibility of *in vivo* experiments and address the potential for protozoan predators to discriminate among antigenically distinct serovars *in vivo*. Additionally, this system makes possible carefully controlled *in vivo* and *ex vivo* experiments

designed to more rigorously test my hypothesis that protozoan predation drives diversifying selection at the *Salmonella rfb* locus. Importantly, these experiments were previously impossible without the technical advances developed in this dissertation.

8.0 THE *SALMONELLA RFB* LOCUS AS A CASE STUDY FOR FRAGMENTED SPECIATION IN BACTERIA

In this dissertation, I elucidated the physiological basis for predator escape in *Salmonella enterica*, where predator recognition of prey is strongly influenced by the identity of the O-antigen as well as its chain length. As a result, the differential distribution of predators within the environment [386] provides sufficiently strong selection to maintain diversifying selection on the *rfb* genes. This prediction was validated upon competition of antigenically-distinct strains of *Salmonella* within an enteric environment.

However, the term "diversifying selection" is somewhat misleading in this case. It is typically applied to cases where variant alleles are maintained within a species because distinct alleles are favored under different environments (Chapter 1). This would be true if one considered *Salmonella enterica* to be a single, cohesive species. This view has been challenged [275-276], and evidence has been offered that the speciation process is more complex in bacteria, where gene exchange – the hallmark process conferring genotypic similarity among conspecific organisms by purging variant alleles – operates independently at different genes. From this perspective, this work uses the *rfb* locus as a case study of the fragmented speciation model [275]. Unlike eukaryotic speciation in which the entire genome of a given organism is reproductively isolated from those of other species, bacterial speciation occurs at a locus-by-locus level within the genome. Genetic information is transferred among bacteria in small

fragments via horizontal transfer; common methods by which this is achieved include conjugation, transformation, and transduction. According to the fragmented speciation model, bacteria can readily exchange genetic information at some loci but not at others; recombination is blocked at loci that contribute to ecological differentiation between donor and recipient cells but can freely occur at loci uninvolved in these processes.

The work of Retchless and Lawrence [275] identified the *rfb* locus as one of the first loci separating *Salmonella* and *Escherichia* from the last common ancestor (Figure 37). Here, the relative divergence between orthologous genes is plotted as a function of their location along the inferred ancestral chromosome. Regions that are more divergent than average ceased to experience diversity-purging recombination between lineages at a point early in their separation. The *rfb* locus is found in one of the most divergent regions of the genome, suggesting that it was among the first loci to become genetically isolated in these nascent species (Figure 37, highlighted in blue). Loci specific to *Salmonella* pathogenicity, such as SPI1 and SPI2 [30, 167-168, 245, 322], were acquired later during the separation of *Salmonella* and *Escherichia* (Figure 37, highlighted in red), suggesting that pathogenicity itself was not the ecological change motivating speciation. As expected, excess polymorphism and variation is observed at the region including and flanking the *rfb* locus of *Salmonella* and *Escherichia* [37, 40-41, 200, 219, 239, 366, 374, 391-392] when compared with genes located elsewhere in the *Salmonella* and *Escherichia* genomes. My own, more extensive analysis shows that variation within species is also much higher than expected in the *rfb* region (Figures 5 and 6). Thus, variability near the *rfb* loci appears to be unaffected by diversity-purging selective sweeps, presumably because recombinants with variant *rfb* alleles would have decreased fitness in the environment to which the original *rfb* allele was adapted.

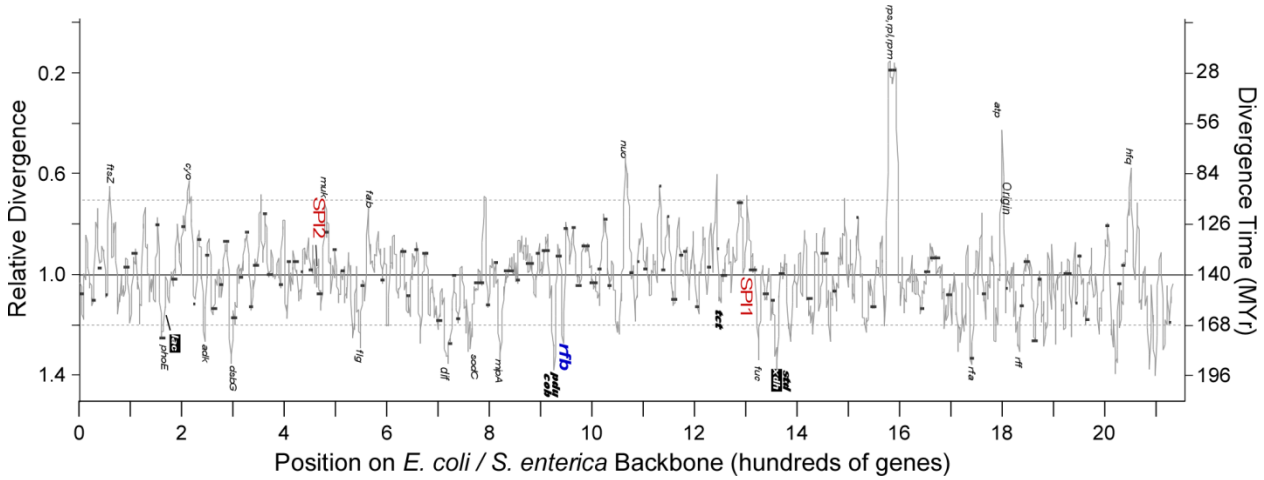

Figure 43. Divergence of chromosomal regions between *Salmonella enterica* and *Escherichia coli*

Relative divergence time for orthologous genes was plotted against chromosomal position in the *E. coli* K12-MG1655 genome averaged across a seven-gene window. Genes shared between *S. enterica* and *E. coli* are shown in italics. Figure from Retchless and Lawrence [275]; available online. Reprinted with permission from AAAS.

Because *Salmonella* is a pathogen, diversity at the *rfb* locus has been casually attributed to selective pressure arising from host immune systems. Selective pressure from host immune systems does significantly shape O-antigen diversity in organisms that have the capacity to alter O-antigen identity on a generational time scale, such as *Neisseria* and *Haemophilus*; see Chapter 1 for a complete discussion. However, expression of the *Salmonella* O-antigen is stable and is used as a serological characteristic by which strains are identified for epidemiologic purposes [166, 264]. Additionally, *Salmonella* interacts with the host immune system at a fundamentally different way than pathogens that contact the adaptive immune system directly, like *Neisseria* and *Haemophilus* [7-8, 19, 21, 24-25, 36, 46, 65, 87, 89, 131, 134, 139, 166, 223, 257, 291, 363, 368-369, 374]. Thus, I do not believe that the biology of *Salmonella* reflects that of an organism in which the *rfb* locus experiences selective pressure directly arising from the host immune system.

What ecological selective pressures could be acting on the *rfb* locus, contributing to *Salmonella*'s separation from *Escherichia* and hypervariability of this locus within the Salmonellae? This question was first addressed by the work of Wildschutte *et al* [387] and Wildschutte and Lawrence [386]. First, protozoan predators can discriminate among natural isolates of *Salmonella* [387]. Second, different environments contain different amoebae [386-387]. Third, predators isolated from the same environment can present a uniform selective pressure upon *Salmonella* [386]. While this work suggests that selective pressure from grazing intestinal protozoa fulfills the requirements for a hypothesis explaining the maintenance of diversifying selection at the *Salmonella rfb* locus, it stopped short of definitively establishing a relationship between the O-antigen and *Salmonella* fitness against predation.

This dissertation provides a more definitive link between protozoan predation and fragmented speciation at the *rfb* locus by implicating the O-antigen in the physiological mechanism of predator escape in *Salmonella*. Moreover, predators can discriminate among natural isolates of *Salmonella in vivo*, further highlighting the likelihood that protozoan predation in host intestines drives diversifying selection at *rfb*. Because the O-antigen is involved in prey choice, selective pressure from protozoan predation can serve as the ecological selective pressure blocking recombination at *rfb*, thus leading to the differentiation of *Salmonella* from *Escherichia* and eventually, all *Salmonella* serovars from each other. According to this model, recombination at the *rfb* locus is highly disfavored due to the important ecological role played by the *rfb* gene product, the O-antigen, in shaping *Salmonella* fitness against protozoan predation. For example, transfer of the *rfb* genes from serovar Dublin into serovar Gallinarium will most likely not result in enhanced fitness against predation by the resident amoebae in the intestines of chickens; it is much more likely that the resulting O-antigen from such a transfer will make this strain more susceptible to predation in this environment. If O-antigens do act as a type of mimicry of host intestinal mucins as discussed in Chapter 5, effectively imparting molecular camouflage upon serovars bearing particular O-antigens, then alteration of this structure should be highly disfavored among strains.

From an ecological perspective, in order for a *Salmonella* serovar to succeed in a given host intestinal environment, it must a) possess an O-antigen that maximizes the ability to escape predation from native gut amoebae; b) successfully evade host immune system defenses in order to invade and replicate within host cells; and c) survive in transient environments encountered between host to host transmission. Because *Salmonella* must first escape protozoan predation before having the opportunity to interact with the host immune system, I can consider the *rfb*

locus as one of the founding loci that began the fragmented speciation process. Once a strain acquires a *rfb* locus that confers an O-antigen permitting maximal predator escape in a given environment, then recombination is disfavored at this locus. Thus, selection on O-antigen identity provides a form of "post-mating" genetic isolation, whereby recombinants are disfavored and removed from the population. Next, it may go on to acquire other specific loci that enable adaptation to a particular environment via horizontal gene transfer, such as pathogenicity islands; thus recombination is further blocked at these loci. The lack of recombination allows for the accumulation of mutations that differentiate the two lineages; these sequence differences provide a robust "pre-mating" barrier to recombination as mismatch correction systems prevent the initiation of recombination events. As this process continues over time, speciation occurs in a fragmented manner throughout the genome as recombination becomes more and more disfavored at loci that confer specific adaptation to particular ecological niches.

In addition to presenting the *rfb* locus as a case study for the fragmented speciation model, this dissertation produced two crucial technological advances for the study of microbial population genetics: genetic manipulation of non-Typhimurium *Salmonella* and multicolor flow cytometry for assessment of microbes in complex environments. These developments make possible future experiments to more rigorously examine the capacity for protozoan predators to discriminate among *Salmonella* strains *in vivo* based solely on the identity of the O-antigen. These approaches also enable other microbial population genetic studies in a wide variety of organisms in complex environments.

Although my research provides a critical link between O-antigen diversity in *Salmonella* and fitness against protozoan grazing, the lifestyle of *Salmonella* constrains the general application of this information to microbes as a whole. *Salmonella* is primarily a mammalian

intestinal pathogen [131, 264], and as such, its lifestyle is mainly constrained to intimate relationships with individual hosts and their immune systems. Therefore, experiments using *Salmonella* as a model organism cannot entirely eliminate the confounding variable of pathogenicity from the environmental factors shaping O-antigen diversity. The observation that amoebae are able to discriminate among *Salmonella* based on surface antigens could be due to direct selective pressure from grazing by intestinal protozoa as I hypothesized, an indirect consequence of selection arising from the host immune system, or a more complicated combination of these direct and indirect sources of selection.

Therefore, new model organisms, preferably non-pathogenic and/or non-host-associated, need to be studied in order to link microbial antigenic diversity with the ubiquitous presence of grazing protozoa in nature. The approaches developed in this dissertation permit a more rigorous, and previously unattainable, examination of how changes at both the environmental and genetic levels impact microbial survival in natural settings, potentially explaining the observation that particular bacterial species occupy certain environmental niches but not others. Additionally, my work provides a technological framework that enhances the capacity for identifying the physiological and ecological attributes that drive fragmented speciation in bacteria.

BIBLIOGRAPHY

1. Ai, H.W., N.C. Shaner, Z. Cheng, R.Y. Tsien, and R.E. Campbell. (2007). Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. Biochemistry. **46**(20): p. 5904-10.
2. Aldridge, P., J. Gnerer, J.E. Karlinsey, and K.T. Hughes. (2006). Transcriptional and translational control of the *Salmonella fliC* gene. J Bacteriol. **188**(12): p. 4487-96.
3. Aldridge, P.D., C. Wu, J. Gnerer, J.E. Karlinsey, K.T. Hughes, and M.S. Sachs. (2006). Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica.* Proc Natl Acad Sci USA. **103**(30): p. 11340-5.
4. Almeida, M.A., M.A. Cunha, and F. Alcantara. (2001). Loss of estuarine bacteria by viral infection and predation in microcosm conditions. Microb Ecol. **42**(4): p. 562-571.
5. Althouse, C., S. Patterson, P. Fedorka-Cray, and R.E. Isaacson. (2003). Type 1 fimbriae of *Salmonella enterica* serovar Typhimurium bind to enterocytes and contribute to colonization of swine *in vivo.* Infect Immun. **71**(11): p. 6446-52.
6. Andersen-Nissen, E., K.D. Smith, K.L. Strobe, S.L. Barrett, B.T. Cookson, S.M. Logan, and A. Aderem. (2005). Evasion of Toll-like receptor 5 by flagellated bacteria. Proc Natl Acad Sci USA. **102**(26): p. 9247-52.
7. Andrews, T.D. and T. Gojobori. (2004). Strong positive selection and recombination drive the antigenic variation of the PilE protein of the human pathogen *Neisseria meningitidis*. Genetics. **166**(1): p. 25-32.
8. Apicella, M.A., M. Shero, G.A. Jarvis, J.M. Griffiss, R.E. Mandrell, and H. Schneider. (1987). Phenotypic variation in epitope expression of the *Neisseria gonorrhoeae* lipooligosaccharide. Infect Immun. **55**(8): p. 1755-61.
9. Ausubel, F.M., *Current protocols in molecular biology*. 2001, J. Wiley: New York.
10. Ayo, B., E. Santamaria, A. Latatu, I. Artolozaga, I. Azua, and J. Iriberri. (2001). Grazing rates of diverse morphotypes of bacterivorous ciliates feeding on four allochthonous bacteria. Lett Appl Microbiol. **33**(6): p. 455-60.
11. Backhed, F., R.E. Ley, J.L. Sonnenburg, D.A. Peterson, and J.I. Gordon. (2005). Host-bacterial mutualism in the human intestine. Science. **307**(5717): p. 1915-20.
12. Bagwell, C.B. and E.G. Adams. (1993). Fluorescence spectral overlap compensation for any number of flow cytometry parameters. Ann N Y Acad Sci. **677**: p. 167-84.
13. Baker, S. and G. Dougan. (2007). The genome of *Salmonella enterica* serovar Typhi. Clin Infect Dis. **45**(Suppl 1): p. S29-33.
14. Baker, S., J. Hardy, K.E. Sanderson, M. Quail, I. Goodhead, R.A. Kingsley, J. Parkhill, B. Stocker, and G. Dougan. (2007). A novel linear plasmid mediates flagellar variation in *Salmonella Typhi.* PLoS Pathog. **3**(5): p. e59.

15. Banning, N., S. Toze, and B.J. Mee. (2002). *Escherichia coli* survival in groundwater and effluent measured using a combination of propidium iodide and the green fluorescent protein. J Appl Microbiol. **93**(1): p. 69-76.
16. Barnes, B. and D.M. Gordon. (2004). Coliform dynamics and the implications for source tracking. Environ Microbiol. **6**(5): p. 501-9.
17. Bastin, D.A., G. Stevenson, P.K. Brown, A. Haase, and P.R. Reeves. (1993). Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol Microbiol. **7**(5): p. 725-34.
18. Baumler, A.J., R.M. Tsolis, T.A. Ficht, and L.G. Adams. (1998). Evolution of host adaptation in *Salmonella enterica.* Infect Immun. **66**(10): p. 4579-87.
19. Bayliss, C.D., D. Field, and E.R. Moxon. (2001). The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. J Clin Invest. **107**(6): p. 657-62.
20. Beavis, A.J. and R.F. Kalejta. (1999). Simultaneous analysis of the cyan, yellow and green fluorescent proteins by flow cytometry using single-laser excitation at 458 nm. Cytometry. **37**(1): p. 68-73.
21. Bedoui, S., A. Kupz, O.L. Wijburg, A.K. Walduck, M. Rescigno, and R.A. Strugnell. (2010). Different bacterial pathogens, different strategies, yet the aim is the same: evasion of intestinal dendritic cell recognition. J Immunol. **184**(5): p. 2237-42.
22. Belley, A., K. Keller, M. Gottke, and K. Chadee. (1999). Intestinal mucins in colonization and host defense against pathogens. Am J Trop Med Hyg. **60**(4 Suppl): p. 10-5.
23. Bender, R.A. and L.C. Sambucetti. (1983). Recombination-induced suppression of cell division following P1-mediated generalized transduction in *Klebsiella aerogenes*. Mol Gen Genet. **189**(2): p. 263-8.
24. Bergman, M.A., L.A. Cummings, R.C. Alaniz, L. Mayeda, I. Fellnerova, and B.T. Cookson. (2005). CD4+-T-cell responses generated during murine *Salmonella enterica* serovar Typhimurium infection are directed towards multiple epitopes within the natural antigen FliC. Infect Immun. **73**(11): p. 7226-35.
25. Bergman, M.A., L.A. Cummings, S.L. Barrett, K.D. Smith, J.C. Lara, A. Aderem, and B.T. Cookson. (2005). CD4+ T cells and toll-like receptors recognize *Salmonella* antigens expressed in bacterial surface organelles. Infect Immun. **73**(3): p. 1350-6.
26. Bergmire-Sweat, D., J. Schlegel, C. Marin, K. Winpisinger, C. Perry, M. Sotir, and J. Harris. (2008). Multistate outbreak of human *Salmonella* infections associated with exposure to turtles--United States, 2007-2008. MMWR Morb Mortal Wkly Rep. **57**(3): p. 69-72.
27. Bettarel, Y., C. Amblard, T. Sime-Ngando, J.F. Carrias, D. Sargos, F. Garabetian, and P. Lavandier. (2003). Viral lysis, flagellate grazing potential, and bacterial production in Lake Pavin. Microb Ecol. **45**(2): p. 119-27.
28. Bettarel, Y., T. Sime-Ngando, C. Amblard, and J. Dolan. (2004). Viral activity in two contrasting lake ecosystems. Appl Environ Microbiol. **70**(5): p. 2941-51.
29. *BD Fluorescence Spectrum Viewer*. [cited 2011 May 25]; Available from: http://www.bdbiosciences.com/research/multicolor/spectrum_viewer/index.jsp.
30. Bleasdale, B., P.J. Lott, A. Jagannathan, M.P. Stevens, R.J. Birtles, and P. Wigley. (2009). The *Salmonella* pathogenicity island 2-encoded type III secretion system is

essential for the survival of *Salmonella enterica* serovar Typhimurium in free-living amoebae. Appl Environ Microbiol. **75**(6): p. 1793-5.
31. Bolivar, F., R.L. Rodriguez, P.J. Greene, M.C. Betlach, H.L. Heyneker, H.W. Boyer, J.H. Crosa, and S. Falkow. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene. **2**(2): p. 95-113.
32. Bolton, A.J., M.P. Osborne, T.S. Wallis, and J. Stephen. (1999). Interaction of *Salmonella choleraesuis*, *Salmonella dublin* and *Salmonella typhimurium* with porcine and bovine terminal ileum *in vivo.* Microbiology. **145 (Pt 9)**: p. 2431-41.
33. Bonifield, H.R. and K.T. Hughes. (2003). Flagellar phase variation in *Salmonella enterica* is mediated by a posttranscriptional control mechanism. J Bacteriol. **185**(12): p. 3567-74.
34. Bowers, B. (1977). Comparison of pinocytosis and phagocytosis in *Acanthamoeba castellanii.* Exp Cell Res. **110**(2): p. 409-17.
35. Boyd, E.F., F.S. Wang, P. Beltran, S.A. Plock, K. Nelson, and R.K. Selander. (1993). *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. J Gen Microbiol. **139 Pt 6**: p. 1125-32.
36. Bradley, C.J., N.J. Griffiths, H.A. Rowe, R.S. Heyderman, and M. Virji. (2005). Critical determinants of the interactions of capsule-expressing *Neisseria meningitidis* with host cells: the role of receptor density in increased cellular targeting via the outer membrane Opa proteins. Cell Microbiol. **7**(10): p. 1490-503.
37. Brahmbhatt, H.N., P. Wyk, N.B. Quigley, and P.R. Reeves. (1988). Complete physical map of the *rfb* gene cluster encoding biosynthetic enzymes for the O antigen of *Salmonella typhimurium* LT2. J Bacteriol. **170**(1): p. 98-102.
38. Brankatschk, K., J. Blom, A. Goesmann, T.H. Smits, and B. Duffy. (2011). Genome of a European fresh-vegetable food safety outbreak strain of *Salmonella enterica* subsp. *enterica* serovar Weltevreden. J Bacteriol. **193**(8): p. 2066.
39. Briones, V., S. Tellez, J. Goyache, C. Ballesteros, M. del Pilar Lanzarot, L. Dominguez, and J.F. Fernandez-Garayzabal. (2004). *Salmonella* diversity associated with wild reptiles and amphibians in Spain. Environ Microbiol. **6**(8): p. 868-71.
40. Brown, P.K., L.K. Romana, and P.R. Reeves. (1991). Cloning of the *rfb* gene cluster of a group C2 *Salmonella* strain: comparison with the *rfb* regions of groups B and D. Mol Microbiol. **5**(8): p. 1873-81.
41. Brown, P.K., L.K. Romana, and P.R. Reeves. (1992). Molecular analysis of the *rfb* gene cluster of *Salmonella* serovar Muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B. Mol Microbiol. **6**(10): p. 1385-94.
42. Burow, L.C., K.S. Gobius, B.A. Vanselow, and A.V. Klieve. (2005). A lack of predatory interaction between rumen ciliate protozoa and Shiga-toxin producing *Escherichia coli.* Lett Appl Microbiol. **40**(2): p. 117-22.
43. Butela, K.A. and J.G. Lawrence. (2010). *Population genetics of Salmonella: Selection for antigenic diversity*, in *Bacterial Population Genetics in Infectious Disease*, D.A. Robinson, D. Falush, and E.J. Feil, Editors. John Wiley and Sons: Hoboken, NJ. p. 287-319.
44. Buzby, J.C. and T. Roberts. (2009). The economics of enteric infections: human foodborne disease costs. Gastroenterology. **136**(6): p. 1851-62.

45. Buzby, J.C., T. Roberts, C.-T.J. Lin, and J.M. MacDonald, *Bacterial foodborne disease: Medical costs and productivity losses*. 1996, U.S. Department of Agriculture: Washington, D.C. p. 93.
46. Carbonnelle, E., D.J. Hill, P. Morand, N.J. Griffiths, S. Bourdoulous, I. Murillo, X. Nassif, and M. Virji. (2009). Meningococcal interactions with the host. Vaccine. **27 Suppl 2**: p. B78-89.
47. Caugant, D.A. (2008). Genetics and evolution of *Neisseria meningitidis*: importance for the epidemiology of meningococcal disease. Infect Genet Evol. **8**(5): p. 558-65.
48. Caugant, D.A., E.A. Hoiby, P. Magnus, O. Scheel, T. Hoel, G. Bjune, E. Wedege, J. Eng, and L.O. Froholm. (1994). Asymptomatic carriage of *Neisseria meningitidis* in a randomly sampled population. J Clin Microbiol. **32**(2): p. 323-30.
49. Caugant, D.A., G. Tzanakaki, and P. Kriz. (2007). Lessons from meningococcal carriage studies. FEMS Microbiol Rev. **31**(1): p. 52-63.
50. Cerdeno-Tarraga, A.M., S. Patrick, L.C. Crossman, G. Blakely, V. Abratt, N. Lennard, I. Poxton, B. Duerden, B. Harris, M.A. Quail, A. Barron, L. Clark, C. Corton, J. Doggett, M.T. Holden, N. Larke, A. Line, A. Lord, H. Norbertczak, D. Ormond, C. Price, E. Rabbinowitsch, J. Woodward, B. Barrell, and J. Parkhill. (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. Science. **307**(5714): p. 1463-5.
51. Chambers, D.L. and A.C. Hulse. (2006). *Salmonella* serovars in the herpetofauna of Indiana County, Pennsylvania. Appl Environ Microbiol. **72**(5): p. 3771-3.
52. Chase, J.M. (1999). Food web effects of prey size refugia: variable interactions and alternative stable equilibria. Am Nat. **154**(5): p. 559-570.
53. Chen, F., C. Poppe, G.R. Liu, Y.G. Li, Y.H. Peng, K.E. Sanderson, R.N. Johnston, and S.L. Liu. (2009). A genome map of *Salmonella enterica* serovar Agona: numerous insertions and deletions reflecting the evolutionary history of a human pathogen. FEMS Microbiol Lett. **293**(2): p. 188-95.
54. Chessa, D., M.G. Winter, M. Jakomin, and A.J. Baumler. (2009). *Salmonella enterica* serotype Typhimurium Std fimbriae bind terminal alpha(1,2)fucose residues in the cecal mucosa. Mol Microbiol. **71**(4): p. 864-75.
55. Chilcott, G.S. and K.T. Hughes. (2000). Coupling of flagellar gene expression to flagellar assembly in *Salmonella enterica* serovar Typhimurium and *Escherichia coli*. Microbiol Mol Biol Rev. **64**(4): p. 694-708.
56. Chiu, C.H., P. Tang, C. Chu, S. Hu, Q. Bao, J. Yu, Y.Y. Chou, H.S. Wang, and Y.S. Lee. (2005). The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. Nucleic Acids Res. **33**(5): p. 1690-8.
57. Clement, J.M., E. Lepouce, C. Marchal, and M. Hofnung. (1983). Genetic study of a membrane protein: DNA sequence alterations due to 17 *lamB* point mutations affecting adsorption of phage lambda. EMBO J. **2**(1): p. 77-80.
58. Cohen, M.L., M. Potter, R. Pollard, and R.A. Feldman. (1980). Turtle-associated salmonellosis in the United States. Effect of Public Health Action, 1970 to 1976. JAMA. **243**(12): p. 1247-9.
59. Collins, L.V. and J. Hackett. (1991). Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. J Bacteriol. **173**(8): p. 2521-9.

60. Cormack, B.P., R.H. Valdivia, and S. Falkow. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene. **173**(1 Spec No): p. 33-8.
61. Cotta, M.A., T.R. Whitehead, and R.L. Zeltwanger. (2003). Isolation, characterization and comparison of bacteria from swine faeces and manure storage pits. Environ Microbiol. **5**(9): p. 737-45.
62. Coyne, M.J., M. Chatzidaki-Livanis, L.C. Paoletti, and L.E. Comstock. (2008). Role of glycan synthesis in colonization of the mammalian gut by the bacterial symbiont *Bacteroides fragilis*. Proc Natl Acad Sci USA. **105**(35): p. 13099-104.
63. Coyne, M.J., K.G. Weinacht, C.M. Krinos, and L.E. Comstock. (2003). Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. Proc Natl Acad Sci USA. **100**(18): p. 10446-51.
64. Cubitt, A.B., L.A. Woollenweber, and R. Heim. (1999). Understanding structure-function relationships in the *Aequorea victoria* green fluorescent protein. Methods Cell Biol. **58**: p. 19-30.
65. Cummings, L.A., S.L. Barrett, W.D. Wilkerson, I. Fellnerova, and B.T. Cookson. (2005). FliC-specific CD4+ T cell responses are restricted by bacterial regulation of antigen expression. J Immunol. **174**(12): p. 7929-38.
66. Cunningham, R.E. (2010). Overview of flow cytometry and fluorescent probes for flow cytometry. Methods Mol Biol. **588**: p. 319-26.
67. Curd, H., D. Liu, and P.R. Reeves. (1998). Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. J Bacteriol. **180**(4): p. 1002-7.
68. Datsenko, K.A. and B.L. Wanner. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci USA. **97**(12): p. 6640-5.
69. Davis, R.W., D. Botstein, J.R. Roth, and Cold Spring Harbor Laboratory., *Advanced bacterial genetics*. A Manual for genetic engineering, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory. 254 p.
70. Day, R.N. and M.W. Davidson. (2009). The fluorescent protein palette: tools for cellular imaging. Chem Soc Rev. **38**(10): p. 2887-921.
71. De Rosa, S.C. and M. Roederer. (2001). Eleven-color flow cytometry. A powerful tool for elucidation of the complex immune system. Clin Lab Med. **21**(4): p. 697-712, vii.
72. Dean, A.M. (1989). Selection and neutrality in lactose operons of *Escherichia coli*. Genetics. **123**(3): p. 441-54.
73. Dean, A.M., D.E. Dykhuizen, and D.L. Hartl. (1988). Fitness effects of amino acid replacements in the beta-galactosidase of *Escherichia coli*. Mol Biol Evol. **5**(5): p. 469-85.
74. Ding, H.F., I. Nakoneczna, and H.S. Hsu. (1990). Protective immunity induced in mice by detoxified *Salmonella* lipopolysaccharide. J Med Microbiol. **31**(2): p. 95-102.
75. Dodson, S.I. (1989). The ecological role of chemical stimuli for the zooplankton: predator-induced morphology in *Daphnia*. Oecologia. **78**: p. 361-367.
76. Donaldson, M., D. Heyneman, R. Dempster, and L. Garcia. (1975). Epizootic of fatal amebiasis among exhibited snakes: epidemiologic, pathologic, and chemotherapeutic considerations. Am J Vet Res. **36**(6): p. 807-17.
77. Dorsey, C.W., M.C. Laarakker, A.D. Humphries, E.H. Weening, and A.J. Baumler. (2005). Salmonella enterica serotype Typhimurium MisL is an intestinal colonization factor that binds fibronectin. Mol Microbiol. **57**(1): p. 196-211.

78. Dworkin, M.S., P.C. Shoemaker, M.J. Goldoft, and J.M. Kobayashi. (2001). Reactive arthritis and Reiter's syndrome following an outbreak of gastroenteritis caused by *Salmonella enteritidis*. Clin Infect Dis. **33**(7): p. 1010-4.
79. Dykhuizen, D.E., A.M. Dean, and D.L. Hartl. (1987). Metabolic flux and fitness. Genetics. **115**(1): p. 25-31.
80. Dykhuizen, D.E. and L. Green. (1991). Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol. **173**(22): p. 7257-68.
81. Dykova, I., J. Lom, B. Machackova, and H. Peckova. (1998). *Vexillifera expectata* sp. n. and other non-encysting amoebae isolated from organs of freshwater fish. Folia Parasitol (Praha). **45**(1): p. 17-26.
82. Ebel-Tsipis, J., D. Botstein, and M.S. Fox. (1972). Generalized transduction by phage P22 in *Salmonella typhimurium*. I. Molecular origin of transducing DNA. J Mol Biol. **71**(2): p. 433-48.
83. Edwards, R.A., R.A. Helm, and S.R. Maloy. (1999). Increasing DNA transfer efficiency by temporary inactivation of host restriction. Biotechniques. **26**(5): p. 892-4.
84. Eisenmann, H., H. Harms, R. Meckenstock, E.I. Meyer, and A.J. Zehnder. (1998). Grazing of *Tetrahymena* sp. on adhered bacteria in percolated columns monitored by in situ hybridization with fluorescent oligonucleotide probes. Appl Environ Microbiol. **64**(4): p. 1264-9.
85. Enomoto, M., K. Oosawa, and H. Momota. (1983). Mapping of the *pin* locus coding for a site-specific recombinase that causes flagellar-phase variation in *Escherichia coli* K-12. J Bacteriol. **156**(2): p. 663-8.
86. Enomoto, M. and B.A. Stocker. (1974). Transduction by phage P1kc in *Salmonella typhimurium*. Virology. **60**(2): p. 503-14.
87. Erwin, A.L., Y.A. Brewah, D.A. Couchenour, P.R. Barren, S.J. Burke, G.H. Choi, R. Lathigra, M.S. Hanson, and J.N. Weiser. (2000). Role of lipopolysaccharide phase variation in susceptibility of *Haemophilus influenzae* to bactericidal immunoglobulin M antibodies in rabbit sera. Infect Immun. **68**(5): p. 2804-7.
88. Fang, G., B. Weiser, C. Kuiken, S.M. Philpott, S. Rowland-Jones, F. Plummer, J. Kimani, B. Shi, R. Kaul, J. Bwayo, O. Anzala, and H. Burger. (2004). Recombination following superinfection by HIV-1. AIDS. **18**(2): p. 153-9.
89. Feuillet, V., S. Medjane, I. Mondor, O. Demaria, P.P. Pagni, J.E. Galan, R.A. Flavell, and L. Alexopoulou. (2006). Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. Proc Natl Acad Sci USA. **103**(33): p. 12487-92.
90. Fiala, I. and I. Dykova. (2003). Molecular characterisation of *Neoparamoeba* strains isolated from gills of *Scophthalmus maximus*. Dis Aquat Organ. **55**(1): p. 11-6.
91. Fontanals, D., D. Van Esso, I. Pons, V. Pineda, I. Sanfeliu, D. Mariscal, J.A. Vazquez, P. Coll, and G. Prats. (1996). Asymptomatic carriage of *Neisseria meningitidis* in a randomly sampled population. Serogroup, serotype and subtype distribution and associated risk factors. Clin Microbiol Infect. **2**(2): p. 145-146.
92. Franklin, N.C. (1969). Mutation in *galU* gene of *E. coli* blocks phage P1 infection. Virology. **38**(1): p. 189-91.
93. Fyda, J. and K. Wiackowski. (1992). Predator-induced morphological defences in the ciliate *Colprium*. Eur J Protistol. **28**: p. 341.
94. Galetto, R. and M. Negroni. (2005). Mechanistic features of recombination in HIV. AIDS Rev. **7**(2): p. 92-102.

95. Garmiri, P., K.E. Coles, T.J. Humphrey, and T.A. Cogan. (2008). Role of outer membrane lipopolysaccharides in the protection of *Salmonella enterica* serovar Typhimurium from desiccation damage. FEMS Microbiol Lett. **281**(2): p. 155-9.
96. German, G.J. and R. Misra. (2001). The TolC protein of *Escherichia coli* serves as a cell-surface receptor for the newly characterized TLS bacteriophage. J Mol Biol. **308**(4): p. 579-85.
97. Gill, S.R., M. Pop, R.T. Deboy, P.B. Eckburg, P.J. Turnbaugh, B.S. Samuel, J.I. Gordon, D.A. Relman, C.M. Fraser-Liggett, and K.E. Nelson. (2006). Metagenomic analysis of the human distal gut microbiome. Science. **312**(5778): p. 1355-9.
98. Golberg, D., Y. Kroupitski, E. Belausov, R. Pinto, and S. Sela. (2011). *Salmonella typhimurium* internalization is variable in leafy vegetables and fresh herbs. Int J Food Microbiol. **145**(1): p. 250-7.
99. Goldman, R.C. and F. Hunt. (1990). Mechanism of O-antigen distribution in lipopolysaccharide. J Bacteriol. **172**(9): p. 5352-9.
100. Gopaul, D.N. and G.D. Duyne. (1999). Structure and mechanism in site-specific recombination. Curr Opin Struct Biol. **9**(1): p. 14-20.
101. Gordon, D.M., S. Bauer, and J.R. Johnson. (2002). The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology. **148**(Pt 5): p. 1513-22.
102. Gordon, D.M. and A. Cowling. (2003). The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology. **149**(Pt 12): p. 3575-86.
103. Gordon, D.M. and F. FitzGibbon. (1999). The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology. **145** ((Pt 10)): p. 2663-71.
104. Greenblatt, J.J., K. Floyd, M.E. Philipps, and C.E. Frasch. (1988). Morphological differences in *Neisseria meningitidis* pili. Infect Immun. **56**(9): p. 2356-62.
105. Griffith, J.F., S.B. Weisberg, and C.D. McGee. (2003). Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. J Water Health. **1**(4): p. 141-51.
106. Grimont, P.A.D. and F.-X. Weill, *Antigenic formulae of the Salmonella serovars*. 9th ed: WHO Collaborating Centre for Reference and Research on *Salmonella*.
107. Grossman, N., M.A. Schmetz, J. Foulds, E.N. Klima, V.E. Jimenez-Lucho, L.L. Leive, and K.A. Joiner. (1987). Lipopolysaccharide size and distribution determine serum resistance in *Salmonella montevideo*. J Bacteriol. **169**(2): p. 856-63.
108. Guttman, D.S. and D.E. Dykhuizen. (1994). Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics. **138**(4): p. 993-1003.
109. Hahn, D., J. Gaertner, M.R. Forstner, and F.L. Rose. (2007). High-resolution analysis of salmonellae from turtles within a headwater spring ecosystem. FEMS Microbiol Ecol. **60**(1): p. 148-55.
110. Hahn, M.A., P.C. Keng, and T.D. Krauss. (2008). Flow cytometric analysis to detect pathogens in bacterial cell mixtures using semiconductor quantum dots. Anal Chem. **80**(3): p. 864-72.
111. Hahn, M.A., J.S. Tabb, and T.D. Krauss. (2005). Detection of single bacterial pathogens with semiconductor quantum dots. Anal Chem. **77**(15): p. 4861-9.
112. Hahn, M.W. and M.G. Hofle. (2001). Grazing of protozoa and its effect on populations of aquatic bacteria. FEMS Microbiol Ecol. **35**(2): p. 113-121.

113. Hammerschmidt, S., A. Muller, H. Sillmann, M. Muhlenhoff, R. Borrow, A. Fox, J. van Putten, W.D. Zollinger, R. Gerardy-Schahn, and M. Frosch. (1996). Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (*siaD*): correlation with bacterial invasion and the outbreak of meningococcal disease. Mol Microbiol. **20**(6): p. 1211-20.
114. Hammes, F. and T. Egli. (2010). Cytometric methods for measuring bacteria in water: advantages, pitfalls and applications. Anal Bioanal Chem. **397**(3): p. 1083-95.
115. Hanafusa, T., K. Saito, A. Tominaga, and M. Enomoto. (1993). Nucleotide sequence and regulated expression of the *Salmonella fljA* gene encoding a repressor of the phase 1 flagellin gene. Mol Gen Genet. **236**(2-3): p. 260-6.
116. Handeland, K., T. Refsum, B.S. Johansen, G. Holstad, G. Knutsen, I. Solberg, J. Schulze, and G. Kapperud. (2002). Prevalence of *Salmonella typhimurium* infection in Norwegian hedgehog populations associated with two human disease outbreaks. Epidemiol Infect. **128**(3): p. 523-7.
117. Hanks, M.C., B. Newman, I.R. Oliver, and M. Masters. (1988). Packaging of transducing DNA by bacteriophage P1. Mol Gen Genet. **214**(3): p. 523-32.
118. Harwood, V.J., B. Wiggins, C. Hagedorn, R.D. Ellender, J. Gooch, J. Kern, M. Samadpour, A.C. Chapman, B.J. Robinson, and B.C. Thompson. (2003). Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J Water Health. **1**(4): p. 153-66.
119. Hawley, T.S., W.G. Telford, A. Ramezani, and R.G. Hawley. (2001). Four-color flow cytometric detection of retrovirally expressed red, yellow, green, and cyan fluorescent proteins. Biotechniques. **30**(5): p. 1028-34.
120. Heim, R., A.B. Cubitt, and R.Y. Tsien. (1995). Improved green fluorescence. Nature. **373**(6516): p. 663-4.
121. Heim, R., D.C. Prasher, and R.Y. Tsien. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc Natl Acad Sci USA. **91**(26): p. 12501-4.
122. Heim, R. and R.Y. Tsien. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. Curr Biol. **6**(2): p. 178-82.
123. Heinrichs, D.E., J.A. Yethon, and C. Whitfield. (1998). Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica.* Mol Microbiol. **30**(2): p. 221-32.
124. Helander, I.M., A.P. Moran, and P.H. Makela. (1992). Separation of two lipopolysaccharide populations with different contents of O-antigen factor 122 in Salmonella enterica serovar typhimurium. Mol Microbiol. **6**(19): p. 2857-62.
125. Hendriksen, R.S., A.R. Vieira, S. Karlsmose, D.M. Lo Fo Wong, A.B. Jensen, H.C. Wegener, and F.M. Aarestrup. (2011). Global monitoring of *Salmonella* serovar distribution from the World Health Organization Global Foodborne Infections Network Country Data Bank: Results of Quality Assured Laboratories from 2001 to 2007. Foodborne Pathog Dis. **April 2011, epub ahead of print**.
126. Hiestand-Nauer, R. and S. Iida. (1983). Sequence of the site-specific recombinase gene cin and of its substrates serving in the inversion of the C segment of bacteriophage P1. EMBO J. **2**(10): p. 1733-40.

127. High, N.J., M.E. Deadman, and E.R. Moxon. (1993). The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope alpha Gal(1-4)beta Gal. Mol Microbiol. **9**(6): p. 1275-82.
128. High, N.J., M.P. Jennings, and E.R. Moxon. (1996). Tandem repeats of the tetramer 5'-CAAT-3' present in lic2A are required for phase variation but not lipopolysaccharide biosynthesis in *Haemophilus influenzae*. Mol Microbiol. **20**(1): p. 165-74.
129. Ho, T.D. and M.K. Waldor. (2007). Enterohemorrhagic *Escherichia coli* O157:H7 *gal* mutants are sensitive to bacteriophage P1 and defective in intestinal colonization. Infect Immun. **75**(4): p. 1661-6.
130. Hofnung, M., A. Jezierska, and C. Braun-Breton. (1976). *lamB* mutations in *E. coli* K12: growth of lambda host range mutants and effect of nonsense suppressors. Mol Gen Genet. **145**(2): p. 207-13.
131. Hohmann, E.L. (2001). Nontyphoidal salmonellosis. Clin Infect Dis. **32**(2): p. 263-9.
132. Holt, K.E., J. Parkhill, C.J. Mazzoni, P. Roumagnac, F.X. Weill, I. Goodhead, R. Rance, S. Baker, D.J. Maskell, J. Wain, C. Dolecek, M. Achtman, and G. Dougan. (2008). High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. Nat Genet. **40**(8): p. 987-93.
133. Hood, D.W., M.E. Deadman, T. Allen, H. Masoud, A. Martin, J.R. Brisson, R. Fleischmann, J.C. Venter, J.C. Richards, and E.R. Moxon. (1996). Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharide biosynthesis. Mol Microbiol. **22**(5): p. 951-65.
134. Hosking, S.L., J.E. Craig, and N.J. High. (1999). Phase variation of *lic1A*, *lic2A* and *lic3A* in colonization of the nasopharynx, bloodstream and cerebrospinal fluid by *Haemophilus influenzae* type b. Microbiology. **145** ((Pt 11)): p. 3005-11.
135. Huang, Y., M. Umeda, Y. Takeuchi, M. Ishizuka, K. Yano-Higuchi, and I. Ishikawa. (2003). Distribution of *Bacteroides forsythus* genotypes in a Japanese periodontitis population. Oral Microbiol Immunol. **18**(4): p. 208-14.
136. Hubalek, Z., W. Sixl, M. Mikulaskova, B. Sixl-Voigt, W. Thiel, J. Halouzka, Z. Juricova, B. Rosicky, L. Matlova, M. Honza, and et al. (1995). Salmonellae in gulls and other free-living birds in the Czech Republic. Cent Eur J Public Health. **3**(1): p. 21-4.
137. Humphries, A., S. Deridder, and A.J. Baumler. (2005). *Salmonella enterica* serotype Typhimurium fimbrial proteins serve as antigens during infection of mice. Infect Immun. **73**(9): p. 5329-38.
138. Humphries, A.D., S.M. Townsend, R.A. Kingsley, T.L. Nicholson, R.M. Tsolis, and A.J. Baumler. (2001). Role of fimbriae as antigens and intestinal colonization factors of *Salmonella* serovars. FEMS Microbiol Lett. **201**(2): p. 121-5.
139. Hyland, K.A., L. Kohrt, L. Vulchanova, and M.P. Murtaugh. (2006). Mucosal innate immune response to intragastric infection by *Salmonella enterica* serovar Choleraesuis. Mol Immunol. **43**(11): p. 1890-9.
140. Hyman, P. and S.T. Abedon. (2010). Bacteriophage host range and bacterial resistance. Adv Appl Microbiol. **70**: p. 217-48.
141. Iida, S. (1984). Bacteriophage P1 carries two related sets of genes determining its host range in the invertible C segment of its genome. Virology. **134**(2): p. 421-34.
142. Iida, S., J. Meyer, K.E. Kennedy, and W. Arber. (1982). A site-specific, conservative recombination system carried by bacteriophage P1. Mapping the recombinase gene cin

and the cross-over sites cix for the inversion of the C segment. EMBO J. **1**(11): p. 1445-53.
143. Ikeda, J.S., C.K. Schmitt, S.C. Darnell, P.R. Watson, J. Bispham, T.S. Wallis, D.L. Weinstein, E.S. Metcalf, P. Adams, C.D. O'Connor, and A.D. O'Brien. (2001). Flagellar phase variation of *Salmonella enterica* serovar Typhimurium contributes to virulence in the murine typhoid infection model but does not influence *Salmonella*-induced enteropathogenesis. Infect Immun. **69**(5): p. 3021-30.
144. Inouye, S. and F.I. Tsuji. (1994). *Aequorea* green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. FEBS Lett. **341**(2-3): p. 277-80.
145. Inzana, T.J. (1983). Electrophoretic heterogeneity and interstrain variation of the lipopolysaccharide of *Haemophilus influenzae*. J Infect Dis. **148**(3): p. 492-9.
146. Inzana, T.J. (1987). Lipopolysaccharide gel profiles of *Haemophilus influenzae* type b for epidemiologic analysis. J Clin Microbiol. **25**(11): p. 2252.
147. Israel, J.V., T.F. Anderson, and M. Levine. (1967). *in vitro* morphogenesis of phage P22 from heads and base-plate parts. Proc Natl Acad Sci USA. **57**(2): p. 284-91.
148. Iwabu, Y., H. Mizuta, M. Kawase, M. Kameoka, T. Goto, and K. Ikuta. (2008). Superinfection of defective human immunodeficiency virus type 1 with different subtypes of wild-type virus efficiently produces infectious variants with the initial viral phenotypes by complementation followed by recombination. Microbes Infect. **10**(5): p. 504-13.
149. Izumiya, H., T. Sekizuka, H. Nakaya, M. Taguchi, A. Oguchi, N. Ichikawa, R. Nishiko, S. Yamazaki, N. Fujita, H. Watanabe, M. Ohnishi, and M. Kuroda. (2011). Whole-genome analysis of *Salmonella enterica* serovar Typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. Antimicrob Agents Chemother. **55**(2): p. 623-30.
150. Jackson, E.N., D.A. Jackson, and R.J. Deans. (1978). EcoRI analysis of bacteriophage P22 DNA packaging. J Mol Biol. **118**(3): p. 365-88.
151. Jennings, M.P., D.W. Hood, I.R. Peak, M. Virji, and E.R. Moxon. (1995). Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. Mol Microbiol. **18**(4): p. 729-40.
152. Jennings, M.P., Y.N. Srikhanta, E.R. Moxon, M. Kramer, J.T. Poolman, B. Kuipers, and P. van der Ley. (1999). The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in *Neisseria meningitidis*. Microbiology. **145** ((Pt 11)): p. 3013-21.
153. Jiang, X.M., B. Neal, F. Santiago, S.J. Lee, L.K. Romana, and P.R. Reeves. (1991). Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol Microbiol. **5**(3): p. 695-713.
154. Johansson, M.E., J.M. Larsson, and G.C. Hansson. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci USA. **108**(Suppl 1): p. 4659-65.
155. Johnson, I.D. and M.T.Z. Spence, eds. *The Molecular Probes® Handbook—A Guide to Fluorescent Probes and Labeling Technologies*. 11th ed. 2010, Life Technologies: Carlsbad, CA.

156. Johnson, L.K., M.B. Brown, E.A. Carruthers, J.A. Ferguson, P.E. Dombek, and M.J. Sadowsky. (2004). Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Appl Environ Microbiol. **70**(8): p. 4478-85.
157. Kajimura, J., A. Rahman, and P.D. Rick. (2005). Assembly of cyclic enterobacterial common antigen in *Escherichia coli* K-12. J Bacteriol. **187**(20): p. 6917-27.
158. Kalir, S., J. McClure, K. Pabbaraju, C. Southward, M. Ronen, S. Leibler, M.G. Surette, and U. Alon. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. Science. **292**(5524): p. 2080-3.
159. Kalish, M.L., K.E. Robbins, D. Pieniazek, A. Schaefer, N. Nzilambi, T.C. Quinn, M.E. St Louis, A.S. Youngpairoj, J. Phillips, H.W. Jaffe, and T.M. Folks. (2004). Recombinant viruses and early global HIV-1 epidemic. Emerg Infect Dis. **10**(7): p. 1227-34.
160. Kamp, D., R. Kahmann, D. Zipser, T.R. Broker, and L.T. Chow. (1978). Inversion of the G DNA segment of phage Mu controls phage infectivity. Nature. **271**(5645): p. 577-80.
161. Kaniuk, N.A., M.A. Monteiro, C.T. Parker, and C. Whitfield. (2002). Molecular diversity of the genetic loci responsible for lipopolysaccharide core oligosaccharide assembly within the genus *Salmonella*. Mol Microbiol. **46**(5): p. 1305-18.
162. Keestra, A.M., M.R. de Zoete, R.A. van Aubel, and J.P. van Putten. (2008). Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. Mol Immunol. **45**(5): p. 1298-307.
163. Kerker, M., H. Chew, P.J. McNulty, J.P. Kratohvil, D.D. Cooke, M. Sculley, and M.P. Lee. (1979). Light scattering and fluorescence by small particles having internal structure. J Histochem Cytochem. **27**(1): p. 250-63.
164. Khlebnikov, A., O. Risa, T. Skaug, T.A. Carrier, and J.D. Keasling. (2000). Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. J Bacteriol. **182**(24): p. 7029-34.
165. Kimura, M., *The neutral theory of molecular evolution*, Cambridge: Cambridge University Press.
166. Kingsley, R.A. and A.J. Baumler. (2000). Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. Mol Microbiol. **36**(5): p. 1006-14.
167. Klein, J.R., T.F. Fahlen, and B.D. Jones. (2000). Transcriptional organization and function of invasion genes within *Salmonella enterica* serovar Typhimurium pathogenicity island 1, including the *prgH*, *prgI*, *prgJ*, *prgK*, *orgA*, *orgB*, and *orgC* genes. Infect Immun. **68**(6): p. 3368-76.
168. Klein, J.R. and B.D. Jones. (2001). *Salmonella* pathogenicity island 2-encoded proteins SseC and SseD are essential for virulence and are substrates of the type III secretion system. Infect Immun. **69**(2): p. 737-43.
169. Klobutcher, L.A., K. Ragkousi, and P. Setlow. (2006). The *Bacillus subtilis* spore coat provides "eat resistance" during phagocytic predation by the protozoan Tetrahymena thermophila. Proc Natl Acad Sci USA. **103**(1): p. 165-70.
170. Kojimoto, A., K. Uchida, Y. Horii, S. Okumura, R. Yamaguch, and S. Tateyama. (2001). Amebiasis in four ball pythons, *Python reginus*. J Vet Med Sci. **63**(12): p. 1365-8.
171. Kolko, M.M., L.A. Kapetanovich, and J.G. Lawrence. (2001). Alternative pathways for siroheme synthesis in *Klebsiella aerogenes*. J Bacteriol. **183**(1): p. 328-35.

172. Koomey, M., E.C. Gotschlich, K. Robbins, S. Bergstrom, and J. Swanson. (1987). Effects of recA mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae.* Genetics. **117**(3): p. 391-8.
173. Kourany, M., L. Bowdre, and A. Herrer. (1976). Panamanian forest mammals as carriers of *Salmonella.* Am J Trop Med Hyg. **25**(3): p. 449-55.
174. Krinos, C.M., M.J. Coyne, K.G. Weinacht, A.O. Tzianabos, D.L. Kasper, and L.E. Comstock. (2001). Extensive surface diversity of a commensal microorganism by multiple DNA inversions. Nature. **414**(6863): p. 555-8.
175. Kuhlmann, H.-W. and K. Heckmann. (1985). Interspecies morphogens regulating prey-predator relationships in protozoa. Science. **227**: p. 1347-1349.
176. Kurzai, O., C. Schmitt, H. Claus, U. Vogel, M. Frosch, and A. Kolb-Maurer. (2005). Carbohydrate composition of meningococcal lipopolysaccharide modulates the interaction of *Neisseria meningitidis* with human dendritic cells. Cell Microbiol. **7**(9): p. 1319-34.
177. Kusch, J. (1993). Behavioural and morphological changes in ciliates induced by the predator *Amoeba proteus*. Oecologia. **96**: p. 354-359.
178. Kutsukake, K. (1997). Autogenous and global control of the flagellar master operon, *flhD*, in *Salmonella typhimurium.* Mol Gen Genet. **254**(4): p. 440-8.
179. Kutsukake, K., H. Nakashima, A. Tominaga, and T. Abo. (2006). Two DNA invertases contribute to flagellar phase variation in *Salmonella enterica* serovar Typhimurium strain LT2. J Bacteriol. **188**(3): p. 950-7.
180. Kuwahara, T., A. Yamashita, H. Hirakawa, H. Nakayama, H. Toh, N. Okada, S. Kuhara, M. Hattori, T. Hayashi, and Y. Ohnishi. (2004). Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. Proc Natl Acad Sci USA. **101**(41): p. 14919-24.
181. Lavitola, A., C. Bucci, P. Salvatore, G. Maresca, C.B. Bruni, and P. Alifano. (1999). Intracistronic transcription termination in polysialyltransferase gene (*siaD*) affects phase variation in *Neisseria meningitidis.* Mol Microbiol. **33**(1): p. 119-27.
182. Lawrence, J.G. and A.C. Retchless. (2009). The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. Methods Mol Biol. **532**: p. 29-53.
183. Le Minor, L., M. Veron, and M. Popoff. (1982). [A proposal for *Salmonella* nomenclature]. Ann Microbiol (Paris). **133**(2): p. 245-54.
184. Le Minor, L., M. Veron, and M. Popoff. (1982). [The taxonomy of *Salmonella*]. Ann Microbiol (Paris). **133**(2): p. 223-43.
185. Leber, A.L. (1999). Intestinal amebae. Clin. Lab. Med. **19**(3): p. 601-619.
186. Lederberg, J. and P.R. Edwards. (1953). Sero-typic recombination in *Salmonella.* J Immunol. **71**(4): p. 232-40.
187. Lederberg, J. and T. Iino. (1956). Phase variation in *Salmonella.* Genetics. **41**(5): p. 743-57.
188. Lee, S.J., L.K. Romana, and P.R. Reeves. (1992). Cloning and structure of group C1 O antigen (*rfb* gene cluster) from *Salmonella enterica* serovar montevideo. J Gen Microbiol. **138**(2): p. 305-12.
189. Lee, S.J., L.K. Romana, and P.R. Reeves. (1992). Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. J Gen Microbiol. **138**(9): p. 1843-55.

190. Leff, L.G. and A.A. Leff. (1996). Use of green fluorescent protein to monitor survival of genetically engineered bacteria in aquatic environments. Appl Environ Microbiol. **62**(9): p. 3486-8.
191. Lennox, E.S. (1955). Transduction of linked genetic characters of the host by bacteriophage P1. Virology. **1**(2): p. 190-206.
192. Levin, B.R. (1988). Frequency-dependent selection in bacterial populations. Philos Trans R Soc Lond B Biol Sci. **319**(1196): p. 459-72.
193. Levinson, G. and G.A. Gutman. (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol. **4**(3): p. 203-21.
194. Ley, R.E., D.A. Peterson, and J.I. Gordon. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell. **124**(4): p. 837-48.
195. Li, Q. and P.R. Reeves. (2000). Genetic variation of dTDP-L-rhamnose pathway genes in *Salmonella enterica*. Microbiology. **146 (Pt 9)**: p. 2291-307.
196. Li, W.H., C.I. Wu, and C.C. Luo. (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol. **2**(2): p. 150-74.
197. Liu, C.H., S.M. Lee, J.M. Vanlare, D.L. Kasper, and S.K. Mazmanian. (2008). Regulation of surface architecture by symbiotic bacteria mediates host colonization. Proc Natl Acad Sci USA. **105**(10): p. 3951-6.
198. Liu, D., A.M. Haase, L. Lindqvist, A.A. Lindberg, and P.R. Reeves. (1993). Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. J Bacteriol. **175**(11): p. 3408-13.
199. Liu, D., L. Lindqvist, and P.R. Reeves. (1995). Transferases of O-antigen biosynthesis in *Salmonella enterica*: dideoxyhexosyltransferases of groups B and C2 and acetyltransferase of group C2. J Bacteriol. **177**(14): p. 4084-8.
200. Liu, D., N.K. Verma, L.K. Romana, and P.R. Reeves. (1991). Relationships among the *rfb* regions of *Salmonella* serovars A, B, and D. J Bacteriol. **173**(15): p. 4814-9.
201. Macpherson, A.J. (2006). IgA adaptation to the presence of commensal bacteria in the intestine. Curr Top Microbiol Immunol. **308**: p. 117-36.
202. Macpherson, A.J. and E. Slack. (2007). The functional interactions of commensal bacteria with intestinal secretory IgA. Curr Opin Gastroenterol. **23**(6): p. 673-8.
203. Macpherson, A.J. and T. Uhr. (2004). Induction of protective IgA by intestinal dendritic cells carrying commensal bacteria. Science. **303**(5664): p. 1662-1665.
204. Manning, P.A., A. Kaufmann, U. Roll, J. Pohlner, T.F. Meyer, and R. Haas. (1991). L-pilin variants of *Neisseria gonorrhoeae* MS11. Mol Microbiol. **5**(4): p. 917-26.
205. Marolda, C.L., L.D. Tatar, C. Alaimo, M. Aebi, and M.A. Valvano. (2006). Interplay of the Wzx translocase and the corresponding polymerase and chain length regulator proteins in the translocation and periplasmic assembly of lipopolysaccharide O-antigen. J Bacteriol. **188**(14): p. 5124-35.
206. Martinez-Diaz, R.A., S. Herrera, A. Castro, and F. Ponce. (2000). *Entamoeba* sp. (Sarcomastigophora: Endamoebidae) from ostriches (*Struthio camelus*) (Aves: Struthionidae). Vet. Parasitol. **92**(3): p. 173-179.
207. Masters, M. (1977). The frequency of P1 transduction of the genes of *Escherichia coli* as a function of chromosomal position: preferential transduction of the origin of replication. Mol Gen Genet. **155**(2): p. 197-202.

208. Mazmanian, S.K., J.L. Round, and D.L. Kasper. (2008). A microbial symbiosis factor prevents intestinal inflammatory disease. Nature. **453**(7195): p. 620-625.
209. McClelland, M., K.E. Sanderson, J. Spieth, S.W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R.K. Wilson. (2001). Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature. **413**(6858): p. 852-6.
210. McCutchan, F.E., M. Hoelscher, S. Tovanabutra, S. Piyasirisilp, E. Sanders-Buell, G. Ramos, L. Jagodzinski, V. Polonis, L. Maboko, D. Mmbando, O. Hoffmann, G. Riedner, F. von Sonnenburg, M. Robb, and D.L. Birx. (2005). In-depth analysis of a heterosexually acquired human immunodeficiency virus type 1 superinfection: evolution, temporal fluctuation, and intercompartment dynamics from the seronegative window period through 30 months postinfection. J Virol. **79**(18): p. 11693-704.
211. McGee, Z.A., D.S. Stephens, L.H. Hoffman, W.F. Schlech, 3rd, and R.G. Horn. (1983). Mechanisms of mucosal invasion by pathogenic *Neisseria.* Rev Infect Dis. **5 Suppl 4**: p. S708-14.
212. McGuckin, M.A., S.K. Linden, P. Sutton, and T.H. Florin. (2011). Mucin dynamics and enteric pathogens. Nat Rev Microbiol. **9**(4): p. 265-78.
213. McQuiston, J.R., R. Parrenas, M. Ortiz-Rivera, L. Gheesling, F. Brenner, and P.I. Fields. (2004). Sequencing and comparative analysis of flagellin genes *fliC*, *fljB*, and *flpA* from *Salmonella.* J Clin Microbiol. **42**(5): p. 1923-1932.
214. Mead, P.S., L. Slutsker, V. Dietz, L.F. McCaig, J.S. Bresee, C. Shapiro, P.M. Griffin, and R.V. Tauxe. (1999). Food-related illness and death in the United States. Emerg Infect Dis. **5**(5): p. 607-25.
215. Mergeay, M. and J. Gerits. (1983). Transduction of *Escherichia coli trp* genes in *Salmonella typhimurium* and effect of N-methyl-N'-nitro-N-nitrosoguanidine on transduction with heterogenotic DNA. J Gen Microbiol. **129**(2): p. 321-35.
216. Meyerholz, D.K. and T.J. Stabel. (2003). Comparison of early ileal invasion by *Salmonella enterica* serovars Choleraesuis and Typhimurium. Vet Pathol. **40**(4): p. 371-375.
217. Meyers, L.A., B.R. Levin, A.R. Richardson, and I. Stojiljkovic. (2003). Epidemiology, hypermutation, within-host evolution and the virulence of *Neisseria meningitidis.* Proc Biol Sci. **270**(1525): p. 1667-77.
218. Michaels, R.H. and C.W. Norden. (1977). Pharyngeal colonization with *Haemophilus influenzae* type b: a longitudinal study of families with a child with meningitis or epiglottitis due to *H. influenzae* type b. J Infect Dis. **136**(2): p. 222-8.
219. Milkman, R., E. Jaeger, and R.D. McBride. (2003). Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. Genetics. **163**(2): p. 475-483.
220. Miller, R.D., D.E. Dykhuizen, and D.L. Hartl. (1988). Fitness effects of a deletion mutation increasing transcription of the 6-phosphogluconate dehydrogenase gene in *Escherichia coli.* Mol Biol Evol. **5**(6): p. 691-703.
221. Mitra, R.D., C.M. Silva, and D.C. Youvan. (1996). Fluorescence resonance energy transfer between blue-emitting and red-shifted excitation derivatives of the green fluorescent protein. Gene. **173**(1 Spec No): p. 13-7.

222. Morita, M., C.R. Fischer, K. Mizoguchi, M. Yoichi, M. Oda, Y. Tanji, and H. Unno. (2002). Amino acid alterations in Gp38 of host range mutants of PP01 and evidence for their infection of an *ompC* null mutant of *Escherichia coli* O157:H7. FEMS Microbiol Lett. **216**(2): p. 243-8.
223. Moxon, E.R. (1985). The molecular basis of *Haemophilus influenzae* virulence. J R Coll Physicians Lond. **19**(3): p. 174-8.
224. Moxon, E.R., P.B. Rainey, M.A. Nowak, and R.E. Lenski. (1994). Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr Biol. **4**(1): p. 24-33.
225. Moxon, R., C. Bayliss, and D. Hood. (2006). Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. **40**: p. 307-33.
226. Mullaney, P.F. and P.N. Dean. (1969). Cell sizing: a small-angle light-scattering method for sizing particles of low relative refractive index. Appl Opt. **8**(11): p. 2361-2.
227. Mullaney, P.F., M.A. Van Dilla, J.R. Coulter, and P.N. Dean. (1969). Cell sizing: a light scattering photometer for rapid volume determination. Rev Sci Instrum. **40**(8): p. 1029-32.
228. Mumby, P.J., C.P. Dahlgren, A.R. Harborne, C.V. Kappel, F. Micheli, D.R. Brumbaugh, K.E. Holmes, J.M. Mendes, K. Broad, J.N. Sanchirico, K. Buch, S. Box, R.W. Stoffle, and A.B. Gill. (2006). Fishing, trophic cascades, and the process of grazing on coral reefs. Science. **311**(5757): p. 98-101.
229. Muroi, M. and K. Tanamoto. (2002). The polysaccharide portion plays an indispensable role in *Salmonella* lipopolysaccharide-induced activation of NF-kappaB through human toll-like receptor 4. Infect Immun. **70**(11): p. 6043-7.
230. Murphy, K.C., K.G. Campellone, and A.R. Poteete. (2000). PCR-mediated gene replacement in *Escherichia coli*. Gene. **246**(1-2): p. 321-30.
231. Murray, G.L., S.R. Attridge, and R. Morona. (2003). Regulation of *Salmonella typhimurium* lipopolysaccharide O-antigen chain length is required for virulence; identification of FepE as a second Wzz. Mol Microbiol. **47**(5): p. 1395-406.
232. Murray, G.L., S.R. Attridge, and R. Morona. (2005). Inducible serum resistance in *Salmonella typhimurium* is dependent on *wzz*(*fepE*)-regulated very long O-antigen chains. Microbes Infect. **7**(13): p. 1296-304.
233. Murray, G.L., S.R. Attridge, and R. Morona. (2006). Altering the length of the lipopolysaccharide O-antigen has an impact on the interaction of *Salmonella enterica* serovar Typhimurium with macrophages and complement. J Bacteriol. **188**(7): p. 2735-2739.
234. Musovic, S., G. Oregaard, N. Kroer, and S.J. Sorensen. (2006). Cultivation-independent examination of horizontal transfer and host range of an IncP-1 plasmid among gram-positive and gram-negative bacteria indigenous to the barley rhizosphere. Appl Environ Microbiol. **72**(10): p. 6687-92.
235. Muthukkumar, S. and V.R. Muthukkaruppan. (1993). Mechanism of protective immunity induced by porin-lipopolysaccharide against murine salmonellosis. Infect Immun. **61**(7): p. 3017-25.
236. Myoda, S.P., C.A. Carson, J.J. Fuhrmann, B.K. Hahm, P.G. Hartel, H. Yampara-Lquise, L. Johnson, R.L. Kuntz, C.H. Nakatsu, M.J. Sadowsky, and M. Samadpour. (2003). Comparison of genotypic-based microbial source tracking methods requiring a host origin database. J Water Health. **1**(4): p. 167-180.

237. National Library of Medicine, N.C.B.I., *The NCBI handbook [Internet]*, in *Chapter 18, The Reference Sequence (RefSeq) Project*. 2002, National Library of Medicine (US), National Center for Biotechnology Information: Bethesda (MD).
238. Neal, B.L., P.K. Brown, and P.R. Reeves. (1993). Use of *Salmonella* phage P22 for transduction in *Escherichia coli.* J Bacteriol. **175**(21): p. 7115-8.
239. Nelson, K. and R.K. Selander. (1994). Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. Proc Natl Acad Sci USA. **91**(21): p. 10227-10231.
240. Nempont, C., D. Cayet, M. Rumbo, C. Bompard, V. Villeret, and J.C. Sirard. (2008). Deletion of flagellin's hypervariable region abrogates antibody-mediated neutralization and systemic activation of TLR5-dependent immunity. J Immunol. **181**(3): p. 2036-43.
241. Nevola, J.J., B.A. Stocker, D.C. Laux, and P.S. Cohen. (1985). Colonization of the mouse intestine by an avirulent *Salmonella typhimurium* strain and its lipopolysaccharide-defective mutants. Infect Immun. **50**(1): p. 152-9.
242. Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiol Mol Biol Rev. **67**(4): p. 593-656.
243. Nikaido, H. and M. Vaara. (1985). Molecular basis of bacterial outer membrane permeability. Microbiol Rev. **49**(1): p. 1-32.
244. Norris, T.L. and A.J. Baumler. (1999). Phase variation of the *lpf* operon is a mechanism to evade cross-immunity between *Salmonella* serotypes. Proc Natl Acad Sci USA. **96**(23): p. 13393-8.
245. Ochman, H. and E.A. Groisman. (1996). Distribution of pathogenicity islands in *Salmonella* spp. Infect Immun. **64**(12): p. 5410-2.
246. Olenych, S.G., N.S. Claxton, G.K. Ottenberg, and M.W. Davidson. (2007). The fluorescent protein color palette. Curr Protoc Cell Biol. **Chapter 21**: p. Unit 21 5.
247. Ormo, M., A.B. Cubitt, K. Kallio, L.A. Gross, R.Y. Tsien, and S.J. Remington. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. Science. **273**(5280): p. 1392-5.
248. Ornellas, E.P. and B.A. Stocker. (1974). Relation of lipopolysaccharide character to P1 sensitivity in *Salmonella typhimurium*. Virology. **60**(2): p. 491-502.
249. Parker, C.T. and J. Guard-Petter. (2001). Contribution of flagella and invasion proteins to pathogenesis of *Salmonella enterica* serovar enteritidis in chicks. FEMS Microbiol Lett. **204**(2): p. 287-91.
250. Parkhill, J., G. Dougan, K.D. James, N.R. Thomson, D. Pickard, J. Wain, C. Churcher, K.L. Mungall, S.D. Bentley, M.T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R.M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T.T. Hien, S. Holroyd, K. Jagels, A. Krogh, T.S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B.G. Barrell. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature. **413**(6858): p. 848-52.
251. Patrick, S., J. Parkhill, L.J. McCoy, N. Lennard, M.J. Larkin, M. Collins, M. Sczaniecka, and G. Blakely. (2003). Multiple inverted DNA repeats of *Bacteroides fragilis* that control polysaccharide antigenic variation are similar to the hin region inverted repeats of *Salmonella typhimurium*. Microbiology. **149**(Pt 4): p. 915-24.

252. Patsos, G. and A. Corfield. (2009). Management of the human mucosal defensive barrier: evidence for glycan legislation. Biol Chem. **390**(7): p. 581-90.
253. Peduzzi, P. and F. Schiemer. (2004). Bacteria and viruses in the water column of tropical freshwater reservoirs. Environ Microbiol. **6**(7): p. 707-15.
254. Perfetto, S.P., P.K. Chattopadhyay, and M. Roederer. (2004). Seventeen-colour flow cytometry: unravelling the immune system. Nat Rev Immunol. **4**(8): p. 648-55.
255. Perry, A.C., C.A. Hart, I.J. Nicolson, J.E. Heckels, and J.R. Saunders. (1987). Inter-strain homology of pilin gene sequences in *Neisseria meningitidis* isolates that express markedly different antigenic pilus types. J Gen Microbiol. **133**(6): p. 1409-18.
256. Perry, A.C., I.J. Nicolson, and J.R. Saunders. (1988). *Neisseria meningitidis* C114 contains silent, truncated pilin genes that are homologous to *Neisseria gonorrhoeae* pil sequences. J Bacteriol. **170**(4): p. 1691-7.
257. Petrovska, L., R.J. Aspinall, L. Barber, S. Clare, C.P. Simmons, R. Stratford, S.A. Khan, N.R. Lemoine, G. Frankel, D.W. Holden, and G. Dougan. (2004). *Salmonella enterica* serovar Typhimurium interaction with dendritic cells: impact of the *sifA* gene. Cell Microbiol. **6**(11): p. 1071-84.
258. Piuri, M., W.R. Jacobs, Jr., and G.F. Hatfull. (2009). Fluoromycobacteriophages for rapid, specific, and sensitive antibiotic susceptibility testing of *Mycobacterium tuberculosis*. PLoS One. **4**(3): p. e4870.
259. Plasterk, R.H., A. Brinkman, and P. van de Putte. (1983). DNA inversions in the chromosome of *Escherichia coli* and in bacteriophage Mu: relationship to other site-specific recombination systems. Proc Natl Acad Sci USA. **80**(17): p. 5355-5358.
260. Poteete, A.R. and A.C. Fenton. (2000). Genetic requirements of phage lambda red-mediated gene replacement in *Escherichia coli* K-12. J Bacteriol. **182**(8): p. 2336-40.
261. Potts, W.J. and J.R. Saunders. (1988). Nucleotide sequence of the structural gene for class I pilin from *Neisseria meningitidis*: homologies with the *pilE* locus of *Neisseria gonorrhoeae*. Mol Microbiol. **2**(5): p. 647-53.
262. Pozio, E. (2003). Foodborne and waterborne parasites. Acta Microbiol. Pol. **52**(Suppl): p. 83-96.
263. Price, B.J., V.H. Kollman, and G.C. Salzman. (1978). Light-scatter analysis of microalgae. Correlation of scatter patterns from pure and mixed asynchronous cultures. Biophys J. **22**(1): p. 29-36.
264. Rabsch, W., H.L. Andrews, R.A. Kingsley, R. Prager, H. Tschape, L.G. Adams, and A.J. Baumler. (2002). Salmonella enterica serotype Typhimurium and its host-adapted variants. Infect Immun. **70**(5): p. 2249-55.
265. Rabsch, W., L. Ma, G. Wiley, F.Z. Najar, W. Kaserer, D.W. Schuerch, J.E. Klebba, B.A. Roe, J.A. Laverde Gomez, M. Schallmey, S.M. Newton, and P.E. Klebba. (2007). FepA- and TonB-dependent bacteriophage H8: receptor binding and genomic sequence. J Bacteriol. **189**(15): p. 5658-74.
266. Ram, J.L., R.P. Ritchie, J. Fang, F.S. Gonzales, and J.P. Selegean. (2004). Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J Environ Qual. **33**(3): p. 1024-1032.
267. Rambaut, A., D. Posada, K.A. Crandall, and E.C. Holmes. (2004). The causes and consequences of HIV evolution. Nat Rev Genet. **5**(1): p. 52-61.
268. Ramirez, B.C., E. Simon-Loriere, R. Galetto, and M. Negroni. (2008). Implications of recombination for HIV diversity. Virus Research. **134**(1-2): p. 64-73.

269. Raser, J.M. and E.K. O'Shea. (2004). Control of stochasticity in eukaryotic gene expression. Science. **304**(5678): p. 1811-4.
270. Reeves, P. (1993). Evolution of *Salmonella* O-antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9**(1): p. 17-22.
271. Reeves, P. (1995). Role of O-antigen variation in the immune response. Trends Microbiol. **3**(10): p. 381-6.
272. Reeves, P.P. and L. Wang. (2002). Genomic organization of LPS-specific loci. Curr Top Microbiol Immunol. **264**(1): p. 109-35.
273. Reeves, P.R., M. Hobbs, M.A. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C.R. Raetz, and P.D. Rick. (1996). Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. **4**(12): p. 495-503.
274. Refsum, T., E. Heir, G. Kapperud, T. Vardund, and G. Holstad. (2002). Molecular epidemiology of *Salmonella enterica* serovar typhimurium isolates determined by pulsed-field gel electrophoresis: comparison of isolates from avian wildlife, domestic animals, and the environment in Norway. Appl Environ Microbiol. **68**(11): p. 5600-6.
275. Retchless, A.C. and J.G. Lawrence. (2007). Temporal fragmentation of speciation in bacteria. Science. **317**(5841): p. 1093-6.
276. Retchless, A.C. and J.G. Lawrence. (2010). Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. Proc Natl Acad Sci USA. **107**(25): p. 11453-8.
277. Richardson, E.J., B. Limaye, H. Inamdar, A. Datta, K.S. Manjari, G.D. Pullinger, N.R. Thomson, R.R. Joshi, M. Watson, and M.P. Stevens. (2011). Genome sequences of *Salmonella enterica* serovar Typhimurium, Choleraesuis, Dublin and Gallinarum strains of highly defined virulence in food-producing animals. J Bacteriol.
278. Rigottier-Gois, L., V. Rochet, N. Garrec, A. Suau, and J. Dore. (2003). Enumeration of *Bacteroides* species in human faeces by fluorescent *in situ* hybridisation combined with flow cytometry using 16S rRNA probes. Syst Appl Microbiol. **26**(1): p. 110-118.
279. Robbe, C., C. Capon, B. Coddeville, and J.C. Michalski. (2004). Structural diversity and specific distribution of O-glycans in normal human mucins along the intestinal tract. Biochem J. **384**(Pt 2): p. 307-16.
280. Robbins, J.B., C. Chu, and R. Schneerson. (1992). Hypothesis for vaccine development: protective immunity to enteric diseases caused by nontyphoidal salmonellae and shigellae may be conferred by serum IgG antibodies to the O-specific polysaccharide of their lipopolysaccharides. Clin Infect Dis. **15**(2): p. 346-61.
281. Roche, R.J., N.J. High, and E.R. Moxon. (1994). Phase variation of *Haemophilus influenzae* lipopolysaccharide: characterization of lipopolysaccharide from individual colonies. FEMS Microbiol Lett. **120**(3): p. 279-83.
282. Roche, R.J. and E.R. Moxon. (1995). Phenotypic variation of carbohydrate surface antigens and the pathogenesis of *Haemophilus influenzae* infections. Trends Microbiol. **3**(8): p. 304-9.
283. Rodriguez-Zaragoza, S. (1994). Ecology of free-living amoebae. Crit Rev Microbiol. **20**(3): p. 225-41.
284. Roederer, M. (2001). Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. Cytometry. **45**(3): p. 194-205.
285. Roederer, M. (2002). Compensation in flow cytometry. Curr Protoc Cytom. **Chapter 1**: p. Unit 1 14.

286. Rondon, L., M. Piuri, W.R. Jacobs, Jr., J. de Waard, G.F. Hatfull, and H.E. Takiff. (2011). Evaluation of fluoromycobacteriophages for detecting drug resistance in *Mycobacterium tuberculosis*. J Clin Microbiol. **49**(5): p. 1838-42.
287. Ronn, R., A.E. McCaig, B.S. Griffiths, and J.I. Prosser. (2002). Impact of protozoan grazing on bacterial community structure in soil microcosms. Appl Environ Microbiol. **68**(12): p. 6094-105.
288. Royle, M.C., S. Totemeyer, L.C. Alldridge, D.J. Maskell, and C.E. Bryant. (2003). Stimulation of Toll-like receptor 4 by lipopolysaccharide during cellular invasion by live *Salmonella typhimurium* is a critical but not exclusive event leading to macrophage responses. J Immunol. **170**(11): p. 5445-54.
289. Rytkonen, A., B. Albiger, P. Hansson-Palo, H. Kallstrom, P. Olcen, H. Fredlund, and A.B. Jonsson. (2004). *Neisseria meningitidis* undergoes PilC phase variation and PilE sequence variation during invasive disease. J Infect Dis. **189**(3): p. 402-9.
290. Saitou, N. and M. Nei. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. **4**(4): p. 406-25.
291. Salazar-Gonzalez, R.M. and S.J. McSorley. (2005). *Salmonella* flagellin, a microbial target of the innate and adaptive immune system. Immunol Lett. **101**(2): p. 117-22.
292. Salzman, G.C., J.M. Crowell, J.C. Martin, T.T. Trujillo, A. Romero, P.F. Mullaney, and P.M. LaBauve. (1975). Cell classification by laser light scattering: identification and separation of unstained leukocytes. Acta Cytol. **19**(4): p. 374-7.
293. Samuel, G. and P. Reeves. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydr Res. **338**(23): p. 2503-19.
294. Sanders, C.J., L. Franchi, F. Yarovinsky, S. Uematsu, S. Akira, G. Nunez, and A.T. Gewirtz. (2009). Induction of adaptive immunity by flagellin does not require robust activation of innate immunity. Eur J Immunol. **39**(2): p. 359-71.
295. Sanders, C.J., D.A. Moore, 3rd, I.R. Williams, and A.T. Gewirtz. (2008). Both radioresistant and hemopoietic cells promote innate and adaptive immune responses to flagellin. J Immunol. **180**(11): p. 7184-92.
296. Sanders, C.J., Y. Yu, D.A. Moore, 3rd, I.R. Williams, and A.T. Gewirtz. (2006). Humoral immune response to flagellin requires T cells and activation of innate immunity. J Immunol. **177**(5): p. 2810-8.
297. Sandulache, R., P. Prehm, D. Expert, A. Toussaint, and D. Kamp. (1985). The cell wall receptor for bacteriophage Mu G(−) in *Erwinia* and *Escherichia coli* C. FEMS Microbiol Lett. **28**(3): p. 307-310.
298. Sandulache, R., P. Prehm, and D. Kamp. (1984). Cell wall receptor for bacteriophage Mu G(+). J Bacteriol. **160**(1): p. 299-303.
299. Santiviago, C.A., C.S. Toro, A.A. Hidalgo, P. Youderian, and G.C. Mora. (2003). Global regulation of the *Salmonella enterica* serovar typhimurium major porin, OmpD. J Bacteriol. **185**(19): p. 5901-5.
300. Savage, D.C. (1977). Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol. **31**: p. 107-33.
301. Schafer, I.A., A.M. Jamieson, M. Petrelli, B.J. Price, and G.C. Salzman. (1979). Multiangle light scattering flow photometry of cultured human fibroblasts: comparison of normal cells with a mutant line containing cytoplasmic inclusions. J Histochem Cytochem. **27**(1): p. 359-65.

302. Schmieger, H. (1972). Phage P22-mutants with increased or decreased transduction abilities. Mol. Gen. Genet. **119**: p. 75-88.
303. Schmieger, H. and P. Schicklmaier. (1999). Transduction of multiple drug resistance of *Salmonella enterica* serovar typhimurium DT104. FEMS Microbiol Lett. **170**(1): p. 251-6.
304. Schmitt, C.K., J.S. Ikeda, S.C. Darnell, P.R. Watson, J. Bispham, T.S. Wallis, D.L. Weinstein, E.S. Metcalf, and A.D. O'Brien. (2001). Absence of all components of the flagellar export and synthesis machinery differentially alters virulence of *Salmonella enterica* serovar Typhimurium in models of typhoid fever, survival in macrophages, tissue culture invasiveness, and calf enterocolitis. Infect Immun. **69**(9): p. 5619-5625.
305. Schnaitman, C.A. and J.D. Klena. (1993). Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol. Rev. **57**(3): p. 655-82.
306. Schoen, C., J. Blom, H. Claus, A. Schramm-Gluck, P. Brandt, T. Muller, A. Goesmann, B. Joseph, S. Konietzny, O. Kurzai, C. Schmitt, T. Friedrich, B. Linke, U. Vogel, and M. Frosch. (2008). Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*. Proc Natl Acad Sci USA. **105**(9): p. 3473-8.
307. Schumacher, W.C., C.A. Storozuk, P.K. Dutta, and A.J. Phipps. (2008). Identification and characterization of *Bacillus anthracis* spores by multiparameter flow cytometry. Appl Environ Microbiol. **74**(16): p. 5220-3.
308. Schuster, F.L., J.F. De Jonckheere, H. Moura, R. Sriram, M.M. Garner, and G.S. Visvesvara. (2003). Isolation of a thermotolerant *Paravahlkampfia* sp. from lizard intestine: biology and molecular identification. J Euk Microbiol. **50**(5): p. 373-378.
309. Scott, T.M., J.B. Rose, T.M. Jenkins, S.R. Farrah, and J. Lukasik. (2002). Microbial source tracking: current methodology and future directions. Appl Environ Microbiol. **68**(12): p. 5796-5803.
310. Sechman, E.V., K.A. Kline, and H.S. Seifert. (2006). Loss of both Holliday junction processing pathways is synthetically lethal in the presence of gonococcal pilin antigenic variation. Mol Microbiol. **61**(1): p. 185-93.
311. Secundino, I., C. Lopez-Macias, L. Cervantes-Barragan, C. Gil-Cruz, N. Rios-Sarabia, R. Pastelin-Palacios, M.A. Villasis-Keever, I. Becker, J.L. Puente, E. Calva, and A. Isibasi. (2006). Salmonella porins induce a sustained, lifelong specific bactericidal antibody memory response. Immunology. **117**(1): p. 59-70.
312. Seiflein, T.A. and J.G. Lawrence. (2001). Methionine-to-cysteine recycling in *Klebsiella aerogenes*. J Bacteriol. **183**(1): p. 336-46.
313. Seiflein, T.A. and J.G. Lawrence. (2006). Two transsulfurylation pathways in *Klebsiella pneumoniae*. J Bacteriol. **188**(16): p. 5762-74.
314. Shapiro, H.M. (1977). Fluorescent dyes for differential counts by flow cytometry: does histochemistry tell us much more than cell geometry? J Histochem Cytochem. **25**(8): p. 976-89.
315. Shapiro, H.M. (2001). Multiparameter flow cytometry of bacteria: implications for diagnostics and therapeutics. Cytometry. **43**(3): p. 223-6.
316. Shapiro, H.M., *Practical flow cytometry*. 4th ed, New York: Wiley-Liss. l, 681 p.
317. Shapiro, H.M. and G. Nebe-von-Caron. (2004). Multiparameter flow cytometry of bacteria. Methods Mol Biol. **263**: p. 33-44.

318. Shapiro, O.H., A. Kushmaro, and A. Brenner. (2010). Bacteriophage predation regulates microbial abundance and diversity in a full-scale bioreactor treating industrial wastewater. ISME J. **4**(3): p. 327-36.
319. Sharp, R., G.P. Hazlewood, H.J. Gilbert, and A.G. O'Donnell. (1994). Unmodified and recombinant strains of *Lactobacillus plantarum* are rapidly lost from the rumen by protozoal predation. J Appl Bacteriol. **76**(2): p. 110-7.
320. Sharpless, T.K., M. Bartholdi, and M.R. Melamed. (1977). Size and refractive index dependence of simple forward angle scattering measurements in a flow system using sharply-focused illumination. J Histochem Cytochem. **25**(7): p. 845-56.
321. Sharpless, T.K. and M.R. Melamed. (1976). Estimation of cell size from pulse shape in flow cytofluorometry. J Histochem Cytochem. **24**(1): p. 257-64.
322. Shea, J.E., M. Hensel, C. Gleeson, and D.W. Holden. (1996). Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium.* Proc Natl Acad Sci USA. **93**(6): p. 2593-7.
323. Shimomura, O., F.H. Johnson, and Y. Saiga. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea.* J Cell Comp Physiol. **59**: p. 223-39.
324. Simpson, J.M., J.W. Santo Domingo, and D.J. Reasoner. (2002). Microbial source tracking: state of the science. Environ Sci Technol. **36**(24): p. 5279-5288.
325. Singh, S.P., S. Miller, Y.U. Williams, K.E. Rudd, and H. Nikaido. (1996). Immunochemical structure of the OmpD porin from *Salmonella typhimurium.* Microbiology. **142**(Pt 11): p. 3201-10.
326. Singh, S.P., S.R. Singh, Y.U. Williams, L. Jones, and T. Abdullah. (1995). Antigenic determinants of the OmpC porin from *Salmonella typhimurium.* Infect Immun. **63**(12): p. 4600-5.
327. Singh, S.P., Y. Upshaw, T. Abdullah, S.R. Singh, and P.E. Klebba. (1992). Structural relatedness of enteric bacterial porins assessed with monoclonal antibodies to *Salmonella typhimurium* OmpD and OmpC. J Bacteriol. **174**(6): p. 1965-73.
328. Slauch, J.M., A.A. Lee, M.J. Mahan, and J.J. Mekalanos. (1996). Molecular characterization of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: *oafA* is a member of a family of integral membrane trans-acylases. J Bacteriol. **178**(20): p. 5904-9.
329. Smith, M.C. and H.M. Thorpe. (2002). Diversity in the serine recombinases. Mol Microbiol. **44**(2): p. 299-307.
330. Smith, N.H., P. Beltran, and R.K. Selander. (1990). Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. J Bacteriol. **172**(5): p. 2209-16.
331. Smith, N.H. and R.K. Selander. (1990). Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella typhimurium.* J Bacteriol. **172**(2): p. 603-609.
332. Smith, N.H. and R.K. Selander. (1991). Molecular genetic basis for complex flagellar antigen expression in a triphasic serovar of *Salmonella.* Proc Natl Acad Sci USA. **88**(3): p. 956-60.
333. Snyder, L.A., S.A. Butcher, and N.J. Saunders. (2001). Comparative whole-genome analyses reveal over 100 putative phase-variable genes in the pathogenic *Neisseria* spp. Microbiology. **147**(Pt 8): p. 2321-32.

334. Sonne-Hansen, J. and S.M. Jenabian. (2005). Molecular serotyping of *Salmonella*: identification of the phase 1 H antigen based on partial sequencing of the *fliC* gene. APMIS. **113**(5): p. 340-8.
335. Sorensen, S.J., A.H. Sorensen, L.H. Hansen, G. Oregaard, and D. Veal. (2003). Direct detection and quantification of horizontal gene transfer by using flow cytometry and *gfp* as a reporter gene. Curr Microbiol. **47**(2): p. 129-33.
336. Spierings, G., R. Elders, B. van Lith, H. Hofstra, and J. Tommassen. (1992). Characterization of the *Salmonella typhimurium phoE* gene and development of *Salmonella*-specific DNA probes. Gene. **122**(1): p. 45-52.
337. Stephens, D.S., Z.A. McGee, M.A. Melly, L.H. Hoffman, and C.R. Gregg. (1982). Attachment of pathogenic *Neisseria* to human mucosal surfaces: role in pathogenesis. Infection. **10**(3): p. 192-5.
338. Stern, A. and T.F. Meyer. (1987). Common mechanism controlling phase and antigenic variation in pathogenic neisseriae. Mol Microbiol. **1**(1): p. 5-12.
339. Sternberg, N. (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA. **87**(1): p. 103-7.
340. Stoddard, L.I., J.B. Martiny, and M.F. Marston. (2007). Selection and characterization of cyanophage resistance in marine *Synechococcus* strains. Appl Environ Microbiol. **73**(17): p. 5516-22.
341. Stoeckel, D.M., M.V. Mathes, K.E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T.L. O'Brien, T.W. Fenger, M. Samadpour, K.M. Strickler, and B.A. Wiggins. (2004). Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. Environ Sci Technol. **38**(22): p. 6109-6117.
342. Strack, R.L., B. Hein, D. Bhattacharyya, S.W. Hell, R.J. Keenan, and B.S. Glick. (2009). A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. Biochemistry. **48**(35): p. 8279-81.
343. Streicher, S., E. Gurney, and R.C. Valentine. (1971). Transduction of the nitrogen-fixation genes in *Klebsiella pneumoniae*. Proc Natl Acad Sci USA. **68**(6): p. 1174-7.
344. Suiter, A.M. and A.M. Dean. (2005). Selection in a cyclical environment: possible impact of phenotypic lag on Darwinian fitness. J Mol Evol. **61**(2): p. 153-70.
345. Tettelin, H., N.J. Saunders, J. Heidelberg, A.C. Jeffries, K.E. Nelson, J.A. Eisen, K.A. Ketchum, D.W. Hood, J.F. Peden, R.J. Dodson, W.C. Nelson, M.L. Gwinn, R. DeBoy, J.D. Peterson, E.K. Hickey, D.H. Haft, S.L. Salzberg, O. White, R.D. Fleischmann, B.A. Dougherty, T. Mason, A. Ciecko, D.S. Parksey, E. Blair, H. Cittone, E.B. Clark, M.D. Cotton, T.R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H.O. Smith, C.M. Fraser, E.R. Moxon, R. Rappuoli, and J.C. Venter. (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science. **287**(5459): p. 1809-15.
346. Thierauf, A., G. Perez, and A.S. Maloy. (2009). Generalized transduction. Methods Mol Biol. **501**: p. 267-86.
347. Thigpen, J.E., J.A. Moore, B.N. Gupta, and D.B. Feldman. (1975). Opossums as a reservoir for Salmonellae. J Am Vet Med Assoc. **167**(7): p. 590-2.
348. Thirion, J.P. and M. Hofnung. (1972). On some genetic aspects of phage lambda resistance in *E. coli* K12. Genetics. **71**(2): p. 207-16.

349. Thomason, L.C., N. Costantino, and D.L. Court. (2007). *E. coli* genome manipulation by P1 transduction. Curr Protoc Mol Biol. **Chapter 1**: p. Unit 1 17.
350. Thomsen, L.E., M.S. Chadfield, J. Bispham, T.S. Wallis, J.E. Olsen, and H. Ingmer. (2003). Reduced amounts of LPS affect both stress tolerance and virulence of *Salmonella enterica* serovar Dublin. FEMS Microbiol Lett. **228**(2): p. 225-31.
351. Thomson, N.R., D.J. Clayton, D. Windhorst, G. Vernikos, S. Davidson, C. Churcher, M.A. Quail, M. Stevens, M.A. Jones, M. Watson, A. Barron, A. Layton, D. Pickard, R.A. Kingsley, A. Bignell, L. Clark, B. Harris, D. Ormond, Z. Abdellah, K. Brooks, I. Cherevach, T. Chillingworth, J. Woodward, H. Norberczak, A. Lord, C. Arrowsmith, K. Jagels, S. Moule, K. Mungall, M. Sanders, S. Whitehead, J.A. Chabalgoity, D. Maskell, T. Humphrey, M. Roberts, P.A. Barrow, G. Dougan, and J. Parkhill. (2008). Comparative genome analysis of *Salmonella Enteritidis* PT4 and *Salmonella Gallinarum* 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. **18**(10): p. 1624-37.
352. Thurman, J., J.D. Parry, P.J. Hill, and J. Laybourn-Parry. (2010). The filter-feeding ciliates *Colpidium striatum* and *Tetrahymena pyriformis* display selective feeding behaviours in the presence of mixed, equally-sized, bacterial prey. Protist. **161**(4): p. 577-88.
353. Tindall, B.J., P.A. Grimont, G.M. Garrity, and J.P. Euzeby. (2005). Nomenclature and taxonomy of the genus *Salmonella.* Int J Syst Evol Microbiol. **55**(Pt 1): p. 521-4.
354. Tinsley, C.R. and J.E. Heckels. (1986). Variation in the expression of pili and outer membrane protein by *Neisseria meningitidis* during the course of meningococcal infection. J Gen Microbiol. **132**(9): p. 2483-90.
355. Tolan, R.W., Jr., R.S. Munson, Jr., and D.M. Granoff. (1986). Lipopolysaccharide gel profiles of *Haemophilus influenzae* type b are not stable epidemiologic markers. J Clin Microbiol. **24**(2): p. 223-7.
356. Tsien, R.Y. (1998). The green fluorescent protein. Annu Rev Biochem. **67**: p. 509-44.
357. Tyler, B.M. and R.B. Goldberg. (1976). Transduction of chromosomal genes between enteric bacteria by bacteriophage P1. J Bacteriol. **125**(3): p. 1105-11.
358. Uzzau, S., G.S. Leori, V. Petruzzi, P.R. Watson, G. Schianchi, D. Bacciu, V. Mazzarello, T.S. Wallis, and S. Rubino. (2001). *Salmonella enterica* serovar-host specificity does not correlate with the magnitude of intestinal invasion in sheep. Infect Immun. **69**(5): p. 3092-9.
359. van de Putte, P., S. Cramer, and M. Giphart-Gassler. (1980). Invertible DNA determines host specificity of bacteriophage Mu. Nature. **286**(5770): p. 218-222.
360. van de Putte, P. and N. Goosen. (1992). DNA inversions in phages and bacteria. Trends Genet. **8**(12): p. 457-62.
361. van der Ende, A., C.T. Hopman, S. Zaat, B.B. Essink, B. Berkhout, and J. Dankert. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and - 35 regions of the promoter. J Bacteriol. **177**(9): p. 2475-2480.
362. van der Ende, A., C.T.P. Hopman, and J. Dankert. (2000). Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. Infect Immun. **68**(12): p. 6685-6690.
363. van der Velden, A.W., M.K. Copass, and M.N. Starnbach. (2005). *Salmonella* inhibit T cell proliferation by a direct, contact-dependent immunosuppressive effect. Proc Natl Acad Sci USA. **102**(49): p. 17769-74.

364. Van Dilla, M.A., T.T. Trujillo, P.F. Mullaney, and J.R. Coulter. (1969). Cell microfluorometry: a method for rapid fluorescence measurement. Science. **163**(872): p. 1213-4.
365. Vazquez-Torres, A., B.A. Vallance, M.A. Bergman, B.B. Finlay, B.T. Cookson, J. Jones-Carson, and F.C. Fang. (2004). Toll-like receptor 4 dependence of innate and adaptive immunity to *Salmonella*: importance of the Kupffer cell network. J Immunol. **172**(10): p. 6202-8.
366. Verma, N.K., N.B. Quigley, and P.R. Reeves. (1988). O-antigen variation in *Salmonella* spp.: *rfb* gene clusters of three strains. J Bacteriol. **170**(1): p. 103-7.
367. Verma, N.K. and P. Reeves. (1989). Identification and sequence of *rfbS* and *rfbE*, which determine antigenic specificity of group A and group D salmonellae. J Bacteriol. **171**(10): p. 5694-701.
368. Virji, M., H. Kayhty, D.J. Ferguson, C. Alexandrescu, and E.R. Moxon. (1991). Interactions of *Haemophilus influenzae* with cultured human endothelial cells. Microb Pathog. **10**(3): p. 231-45.
369. Virji, M., H. Kayhty, D.J. Ferguson, C. Alexandrescu, and E.R. Moxon. (1992). Interactions of *Haemophilus influenzae* with human endothelial cells *in vitro*. J Infect Dis. **165 Suppl 1**: p. S115-6.
370. Voetsch, A.C., T.J. Van Gilder, F.J. Angulo, M.M. Farley, S. Shallow, R. Marcus, P.R. Cieslak, V.C. Deneen, and R.V. Tauxe. (2004). FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin Infect Dis. **38**(Suppl 3): p. S127-134.
371. Vogel, H.J. and D.M. Bonner. (1956). Acetylornithinase of Escherichia coli: partial purification and some properties. J Biol Chem. **218**(1): p. 97-106.
372. Wachter, R.M., B.A. King, R. Heim, K. Kallio, R.Y. Tsien, S.G. Boxer, and S.J. Remington. (1997). Crystal structure and photodynamic behavior of the blue emission variant Y66H/Y145F of green fluorescent protein. Biochemistry. **36**(32): p. 9759-65.
373. Wang, L., K. Andrianopoulos, D. Liu, M.Y. Popoff, and P.R. Reeves. (2002). Extensive variation in the O-antigen gene cluster within one *Salmonella enterica* serogroup reveals an unexpected complex history. J Bacteriol. **184**(6): p. 1669-77.
374. Wang, L., L.K. Romana, and P.R. Reeves. (1992). Molecular analysis of a *Salmonella enterica* group E1 *rfb* gene cluster: O-antigen and the genetic basis of the major polymorphism. Genetics. **130**(3): p. 429-43.
375. Wang, L., D. Rothemund, H. Curd, and P.R. Reeves. (2003). Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. J Bacteriol. **185**(9): p. 2936-43.
376. Wang, W., A.V. Perepelov, L. Feng, S.D. Shevelev, Q. Wang, S.N. Senchenkova, W. Han, Y. Li, A.S. Shashkov, Y.A. Knirel, P.R. Reeves, and L. Wang. (2007). A group of *Escherichia coli* and *Salmonella enterica* O-antigens sharing a common backbone structure. Microbiology. **153**(Pt 7): p. 2159-67.
377. Wasson, K. and R.L. Peper. (2000). Mammalian microsporidiosis. Vet. Pathol. **37**(2): p. 113-128.
378. Weening, E.H., J.D. Barker, M.C. Laarakker, A.D. Humphries, R.M. Tsolis, and A.J. Baumler. (2005). The *Salmonella enterica* serotype Typhimurium *lpf*, *bcf*, *stb*, *stc*, *std*, and *sth* fimbrial operons are required for intestinal persistence in mice. Infect Immun. **73**(6): p. 3358-66.

379. Weiser, J.N., J.M. Love, and E.R. Moxon. (1989). The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. Cell. **59**(4): p. 657-65.
380. Weiser, J.N., D.J. Maskell, P.D. Butler, A.A. Lindberg, and E.R. Moxon. (1990). Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. J Bacteriol. **172**(6): p. 3304-9.
381. Weiser, J.N. and N. Pan. (1998). Adaptation of *Haemophilus influenzae* to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. Mol Microbiol. **30**(4): p. 767-75.
382. Weitere, M., T. Bergfeld, S.A. Rice, C. Matz, and S. Kjelleberg. (2005). Grazing resistance of *Pseudomonas aeruginosa* biofilms depends on type of protective mechanism, developmental stage and protozoan feeding mode. Environ Microbiol. **7**(10): p. 1593-601.
383. Wheeler, A.L., P.G. Hartel, D.G. Godfrey, J.L. Hill, and W.I. Segars. (2002). Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J Environ Qual. **31**(4): p. 1286-1293.
384. Wicklow, B.J. (1988). Developmental polymorphism induced by introspecific predation in the ciliated protozooan *Onychodromus quadricornutus*. J Protozool. **35**: p. 137-141.
385. Wildschutte, H. and K.A. Butela, *Protocol for use of fluorescent labeled cells in protozoan predation experiments*. 2011, University of Pittsburgh.
386. Wildschutte, H. and J.G. Lawrence. (2007). Differential Salmonella survival against communities of intestinal amoebae. Microbiology. **153**(Pt 6): p. 1781-9.
387. Wildschutte, H., D.M. Wolfe, A. Tamewitz, and J.G. Lawrence. (2004). Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella.* Proc Natl Acad Sci USA. **101**(29): p. 10644-9.
388. Williams, K.P., J.J. Gillespie, B.W. Sobral, E.K. Nordberg, E.E. Snyder, J.M. Shallom, and A.W. Dickerman. (2010). Phylogeny of gammaproteobacteria. J Bacteriol. **192**(9): p. 2305-14.
389. Wisniewski-Dye, F. and L. Vial. (2008). Phase and antigenic variation mediated by genome modifications. Antonie Van Leeuwenhoek. **94**(4): p. 493-515.
390. Wu, K.Y., G.R. Liu, W.Q. Liu, A.Q. Wang, S. Zhan, K.E. Sanderson, R.N. Johnston, and S.L. Liu. (2005). The genome of *Salmonella enterica* serovar gallinarum: distinct insertions/deletions and rare rearrangements. J Bacteriol. **187**(14): p. 4720-7.
391. Xiang, S.H., A.M. Haase, and P.R. Reeves. (1993). Variation of the *rfb* gene clusters in *Salmonella enterica.* J Bacteriol. **175**(15): p. 4877-84.
392. Xiang, S.H., M. Hobbs, and P.R. Reeves. (1994). Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. J Bacteriol. **176**(14): p. 4357-65.
393. Xu, J., M.A. Mahowald, R.E. Ley, C.A. Lozupone, M. Hamady, E.C. Martens, B. Henrissat, P.M. Coutinho, P. Minx, P. Latreille, H. Cordum, A. Van Brunt, K. Kim, R.S. Fulton, L.A. Fulton, S.W. Clifton, R.K. Wilson, R.D. Knight, and J.I. Gordon. (2007). Evolution of symbiotic bacteria in the distal human intestine. PLoS Biol. **5**(7): p. e156.
394. Yamamoto, S. and K. Kutsukake. (2006). FljA-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica* serovar Typhimurium. J Bacteriol. **188**(3): p. 958-67.

395. Yanagihara, S., S. Iyoda, K. Ohnishi, T. Iino, and K. Kutsukake. (1999). Structure and transcriptional control of the flagellar master operon of *Salmonella typhimurium*. Genes Genet Syst. **74**(3): p. 105-11.
396. Yang, G., J.R. Rich, M. Gilbert, W.W. Wakarchuk, Y. Feng, and S.G. Withers. (2010). Fluorescence activated cell sorting as a general ultra-high-throughput screening method for directed evolution of glycosyltransferases. J Am Chem Soc. **132**(30): p. 10570-7.
397. Yang, T.T., L. Cheng, and S.R. Kain. (1996). Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. Nucleic Acids Res. **24**(22): p. 4592-3.
398. Yarmolinsky, M.B. and N. Sternberg. (1988). *Bacteriophage P1*, in *The Bacteriophages*, R. Calendar, Editor. Plenum Publishing Corporation: New York, NY. p. 291-438.
399. Zeph, L.R., M.A. Onaga, and G. Stotzky. (1988). Transduction of *Escherichia coli* by bacteriophage P1 in soil. Appl Environ Microbiol. **54**(7): p. 1731-7.
400. Zhang, R., M.G. Weinbauer, and P.Y. Qian. (2007). Viruses and flagellates sustain apparent richness and reduce biomass accumulation of bacterioplankton in coastal marine waters. Environ Microbiol. **9**(12): p. 3008-18.
401. Zhuang, J., A.E. Jetzt, G. Sun, H. Yu, G. Klarmann, Y. Ron, B.D. Preston, and J.P. Dougherty. (2002). Human Immunodeficiency Virus Type 1 recombination: rate, fidelity, and putative hot spots. J. Virol. **76**(22): p. 11273-11282.
402. Zieg, J. and M. Simon. (1980). Analysis of the nucleotide sequence of an invertible controlling element. Proc Natl Acad Sci USA. **77**(7): p. 4196-200.
403. Zwahlen, A., L.G. Rubin, and E.R. Moxon. (1986). Contribution of lipopolysaccharide to pathogenicity of *Haemophilus influenzae*: comparative virulence of genetically-related strains in rats. Microb Pathog. **1**(5): p. 465-73.