Studies on pathogenic bacteria using genetic and genomic approaches and a novel small animal model

Studies on pathogenic bacteria using genetic and genomic approaches and a novel small animal model

A dissertation presented

by

Jonathan David D'Gama

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

February 2020

Dissertation Advisor: Matthew K. Waldor Jonathan David D'Gama

Studies on pathogenic bacteria using genetic and genomic approaches and a novel small animal model

Abstract

Infections with pathogenic bacteria result in disease in both humans and animals. Bacteria rely on virulence factors to cause disease in their hosts. Deepening our understanding of host-pathogen interactions requires identification and characterization of bacterial virulence factors as well as analysis of their roles in infection in vivo. Genetic and genomic approaches serve as powerful, high-throughput approaches to define these factors, while dissection of their in vivo roles requires a relevant small animal model of infection. In this thesis, I describe our work on two bacterial pathogens of global importance, *Streptococcus equi* subspecies *zooepidemicus* (SEZ), a scourge of swine, and *Shigella*, the current leading bacterial cause of diarrheal deaths worldwide.

In SEZ, a combined genetics and genomics approach led to the discovery of a novel virulence regulator, SezV, that activates expression of a neighboring gene that is a key virulence factor and immunogen, the M-like protein, SzM. Absence of SezV led to a severe attenuation of virulence in a mouse infection model. Large-scale analyses of publicly available genomes revealed that the *sezV szM* locus was present in all strains of SEZ and the closely related bacteria, *S. equi* subspecies *equi*, and some *S. pyogenes* strains, defining a distinct class of virulent streptococci. Furthermore, we found evidence that a specific monoclonal antibody to a common

microbial surface polysaccharide, poly-N-acetylglucosamine (PNAG) recognized the SzM protein, suggestive of protein glycosylation.

In our work on *Shigella*, we developed a small animal model of infection involving oral inoculation of infant rabbits. Infected animals developed disease with diarrhea and intestinal pathology reminiscent of aspects of human disease. The bacteria robustly colonized the intestine, invaded colonic epithelial cells, and foci of invasion, which represent sites where several neighboring epithelial cells contained intracellular bacteria, were observed in the colon. In situ mRNA labeling demonstrated that expression of a key chemokine, IL-8, was elevated in uninfected cells near infected cells. Finally, we found that two key *Shigella* virulence factors, IcsA, a protein required for cell-to-cell spreading, and the type III secretion system (T3SS), were required for the development of diarrhea and intestinal pathology.

Table of Contents

Title i

Copyright Page ii

Abstract iii

Table of Contents v

Acknowledgements vi

Chapter 1: Introduction 1

Chapter 2: A conserved streptococcal virulence regulator controls expression of a distinct class of M-like proteins 53

Chapter 3: An oral inoculation infant rabbit model for *Shigella* infection 119

Chapter 4: Discussion 178

Acknowledgements

I had the wonderful opportunity to be mentored and surrounded by numerous inspiring colleagues during my PhD.

My advisor, Dr. Matthew K. Waldor has been a tremendous and supportive mentor throughout my PhD. It has been an incredibly instructive and fulfilling opportunity to be in his lab. Matt fosters a lab environment that lends itself to creative and cutting-edge science. His undaunted and relentless enthusiasm and passion for science and research as a whole is inspiring and motivating, as is his range of research interests. Because of this, Matt allowed me to have the opportunity to gain exposure to and perform research in multiple different fields and projects during my PhD. Furthermore, he provided so many opportunities to learn about the many other aspects of science – writing research papers, reviews, and proposals, mentoring, and giving research presentations. As a MD-PhD, Matt has been instrumental in helping me understand how to become a successful physician scientist.

I also feel lucky to have been able to work with a PI who serves as a role model not only scientifically and professionally, but also personally. I have already been able to learn so much from Matt over the years about science, the role of a professor, mentoring, and medicine, yet always find that I still learn from him. It has been reassuring and humbling to know that Matt seeks the best for me both professionally and personally. As I move on, I know that I will greatly miss our daily chats about life and science. I am indebted to Matt for all his time, patience, scientific insights, guidance, and advice he has given and invested in me over the years to help me grow, and do hope to continue to learn from him in the future.

Another professor on the 7th floor of the Channing Laboratory, Dr. Gerald Pier, who I worked with for the project on streptococcal biology, was a constant source of support and encouragement, and I appreciated his many discussions with me over the years about science.

I thank Zhe Ma, a visiting professor from China who I worked closely with on a project for a year and a half, for introducing me to the field of streptococcal biology, his always present cordiality and humor regardless of how early or late it was, and teaching me much about science.

My daily lab experience would not have been the same without my many Waldor lab peers, both past and present. Many thanks to Carole Kuehl-Mannetho, a source of much research advice and a welcoming presence in the lab, who I rotated under and subsequently worked with for many years, for introducing me to the lab, teaching me about *Shigella*, and creating a warm, friendly lab environment through my graduate years; Brandon Sit, my lab baymate and fellow companion in the lab in the day and late nights; and Satoshi Kimura, who sat opposite me, an inimitable scientist and patient teacher, for all his guidance and source of much levity – it was incredible to learn from such a bright molecular biologist and researcher.

I also thank my fellow graduate students, who were constant sources of support and scientific advice – Brandon Sit, Alyson Warr, Troy Hubbard, Bolutife Fakoya, Karthik Hullahalli and Rachel Giorgio, and Joseph Park, a medical student who performed research for several years in the lab, for all his advice on CRISPR screens. My research would not have progressed far without the advice and support of the many brilliant and patient post-docs in the lab – Carole Kuehl-Mannetho, Satoshi Kimura, Abdelrahim Zoued, Aurore Fleurie, Gabriel Billings, Hailong Zhang, Xu Liu, Ting Zhang, Ian Campbell, Guanhua Yang, Jacob Lazarus, Veerasak Srisuknimit, and Alline Pacheco.

I have been fortunate to be part of intellectually stimulating scientific communities during my graduate school and medical training. I would like to thank the Health Sciences and Technology program (HST), Biological and Biomedical Sciences program (BBS), and the MD-PhD program at Harvard Medical School and MIT for their financial and personal support. I would like to thank my advisors, Dr. Susan Dymecki, Dr. Richard Mitchell, Dr. David Miyamoto, and Dr. David Pellman for all their constant, ongoing support, guidance, and help.

Thank you also to my PQE committee, Dr. Gerald Pier, Dr. Lee Gehrke, and Dr. Richard Malley, for their early encouragement and advice on my projects.

I thank my dissertation advisory committee, Dr. Marcia Goldberg, Dr. Tom Bernhardt, and Dr. Sarah Fortune, for all their time and effort to help guide me through my graduate years. I sincerely appreciate their support and patience as I progressed through my PhD, and their willingness to always make time to talk to me and offer advice.

I thank ahead of time my thesis examination committee, Dr. Cammie Lesser, Dr. Stephen Lory, Dr. John Leong and Dr. Marcia Goldberg for graciously giving of their time to serve on my thesis examination committee. I thank especially Dr. Marcia Goldberg for her generosity in again serving as my committee chair.

I thank my undergraduate research advisor, Dr. Victoria D'Souza, who I worked with throughout college and set me on track to become an independent scientist. In addition, I am very appreciative of Dr. Rafael Carazo-Salas, in whose lab I worked during an undergraduate summer in the other Cambridge (in England).

I am also deeply grateful to my initial research mentors, Dr. Michael Nachman, and Dr. Koen Visscher, in whose labs at the University of Arizona I began doing scientific research while I was still in high school. Their guidance, patience and enthusiasm sparked my life-long

passion for research, science and biology. I especially thank Dr. Joaquin Ruiz, then Dean of the College of Science, University of Arizona, who saw fit to give a young high school student the opportunity to do research, and chose these two nurturing labs at the University of Arizona as a starting point for my research career.

Finally, I would like to thank everyone in Leverett House, which has been my home away from home for so many years, my classmates, and my family, a constant source of support and encouragement. I have been so fortunate to have Alissa, my older sister, a fellow physician scientist and recent alumna of this program, as my role model and best friend since I was young. She is always ready to lend her ear and advise me, as is her husband, Adrian. My parents, Malti and David, have made many sacrifices to nurture us, and my successes are all products of their tireless efforts. I thank my loving God for so many blessings.

Chapter 1

Introduction

Pathogens and Virulence Factors

Bacteria are microorganisms ubiquitous in nature often found in close association with higher organisms, which they colonize in various locations. The relationships among bacteria and their host organisms are complex and range from symbiotic, such as *Bacteroides thetaiotaomicron,* the carbohydrate degrading bacteria found in the intestinal microbiome, to overtly pathogenic, such as *Bacillus anthracis*, the infectious agent of anthrax. In addition, some relationships can flip depending on the health of the host or location of bacteria in the host. For example, opportunistic pathogens are typically innocuous commensals, but can cause disease when host barriers or immune defenses are disrupted.

The factors that separate a commensal organism from a pathogen or endow a strain with pathogenic potential lie often within the organism's genome. Many bacteria become pathogenic by acquiring genetically encoded virulence factors, which can be proteins, carbohydrates, or other macromolecules that directly or indirect cause damage to the host. Classic examples include exotoxins like cholera toxin, whose actions largely account for the disease cholera, and Shiga toxin, which is responsible for hemolytic uremic syndrome. In addition, host-directed secretion systems such as the type III secretion system (T3SS) can inject intoxicating bacterial proteins into host cells. Virulence factors can also diminish the impact of the host immune response against the bacteria, such as an extracellular carbohydrate capsule or immunomodulatory surface protein, e.g. M protein in group A streptococcus.

Identification and characterization of the complete suite of virulence factors in a pathogen and their regulation allows for deepening of both our understanding of the pathogenesis of the infectious disease and host responses to infection. Modern genetic techniques, which rely heavily on next generation sequencing technologies, are often cornerstone methodologies utilized to

interrogate various aspects of the virulence of pathogenic bacteria. Furthermore, characterizing the importance of virulence factors in infection and translating findings requires the development and use of small animal models that faithfully recapitulate infection in the original host.

Applications of next generation sequencing in bacteriology

Next generation sequencing (NGS) has revolutionized the study of bacteria by enabling both rapid sequencing of thousands of strains of bacterial species and, though the use of the technology as a massively parallel molecular counting device, high-throughput measurements of other biological phenomena. Here I review some of the NGS applications utilized in the study of host-pathogen interactions.

Genetic Screens & Transposon-insertion sequencing

(Portions of this section have been adapted from D'Gama et al. Chapter 13: CRISPR Screens. CRISPR Biology and Applications. *ASM Press*. 2020)

Identification of bacterial virulence factors often begins with genetic screens. More generally, screens identify individuals (or cells) with phenotypes of interest in heterogeneous populations. Genetic screens permit the association of genetic elements or genotypes with a phenotype through the use of a population whose phenotypic heterogeneity arises from underlying genetic diversity. Forward genetic screens, through phenotypic selection, uncover causative genotype(s) of a phenotype, while reverse genetic screens uncover the phenotype(s) produced by specific genotypes. Phenotypic selection, which is the identification of a phenotype of interest, can be either positive or negative, enriching for or against the phenotype, and thus increasing or decreasing the abundance of the causative genotype(s). Genetic diversity in a

population is typically achieved by mutating a population that was originally genetically homogenous. Mutagenesis methods vary in efficacy, specificity, and the range of phenotypic and genotypic diversity produced. Ideally, mutagenesis generates a wide range of genotypic and phenotypic diversity in a population by creating only a single genetic mutation within each individual. Screens require large-sized populations to query large domains of genotypic and phenotypic space. Traditional methods used for large-scale mutagenesis, such as chemical or transposon mutagenesis, generate mutations in random or pseudo-random genomic locations.

Currently, most genetic screens in bacteria are carried out using transposon-insertion sequencing (variously abbreviated as Tn-seq, TIS, INSeq, HITS, or TraDIS). Tn-seq is a high-throughput method to perform mapping of transposon insertions, typically for the purpose of determining the frequency of transposon insertions at each site in the genome (Figure 1.1) (1). In prokaryotes, due to the feasibility of performing transposon mutagenesis in many diverse species, Tn-seq has been widely utilized as a robust, feasible technology to rapidly conduct genetic screens. Hence, Tn-seq can be used to identify genetic loci contributing to a phenotype of interest.

A few steps are required to use Tn-seq to determine changes in the frequency of transposon mutants after phenotypic selection and infer genetic loci implicated in modulating the phenotype of interest (Figure 1.1). After mutagenizing a population of bacteria with transposon mutagenesis to make a genetically diverse library, phenotypic selection is applied to the transposon library to enrich or deplete transposon mutants that influence the phenotype. Genomic DNA from both the initial and selected libraries are then collected and used for construction of a next-generation sequencing library in which genomic regions directly adjacent to sites of transposon insertions are amplified (Figure 1.2). Finally, computational analysis of the

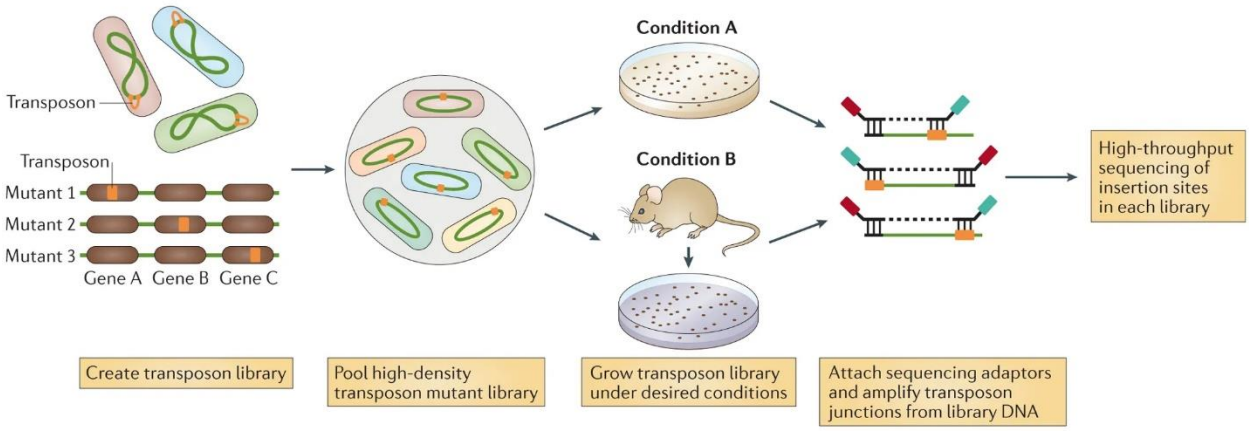


Figure 1.1. Transposon insertion sequencing (Tn-seq, TIS, HITS, TraDIS, INSeq). Transposon library generation, screening, and sequencing library construction is depicted. A high density transposon library is generated that contain mutants each harboring a single transposon insertion and the library contains mutants with transposons at multiple locations in each gene. The transposon library is grown under selective pressures and the unselected and selected libraries are both prepared for NGS sequencing by amplifying the genomic DNA adjacent to the transposon. The frequencies of the transposon insertions are compared among the unselected and selected libraries to determine genes that modulate fitness under the tested selective pressure. Adapted from (1).

sequencing reads is performed using any number of several statistical algorithms (2–5) to accurately identify changes in the frequency of transposon mutants and thus genetic loci affecting the phenotype of interest.

Tn-seq has been widely applied to identify genetic loci important for bacterial colonization of host tissue, such as the intestinal tract (2, 6–8). In these experiments, the transposon library is used as the inoculum. The transposon library proliferates in the animal for a sufficient amount of time to allow for the selective pressure exerted by the host environment to

drive changes in genetic diversity. After proliferation, the remaining transposon mutants in the tissue of interest are isolated and the frequency of mutants in the original and animal passaged libraries are compared. Genetic loci not sustaining transposon insertions after passaging are typically required for growth in the animal tissue and hence important for host colonization, and if the microbe is pathogenic, potentially for virulence as well.

A key caveat in these experiments is the nature of the bottleneck to colonization of the tissue of interest. A bottleneck is a random loss of individuals in a population during initial colonization of the tissue of interest that can lead to large stochastic decreases in genetic diversity in a transposon library (9). These decreases in diversity can confound identification of genetic loci required for growth in animal tissue. If the bottleneck is narrow, only a few of the bacteria from the original inoculum will seed the tissue, and so after recovery of transposon mutants surviving in the tissue, genetic loci may be depleted for transposon insertions due to stochastic loss of mutants instead of due to selective pressure exerted by the host environment. This can lead to the detection of numerous false positive results in a Tn-seq experiment. Fortunately, there are some computational approaches used to analyze Tn-seq experiments that, to some extent, can mitigate the effects of the bottleneck (2).

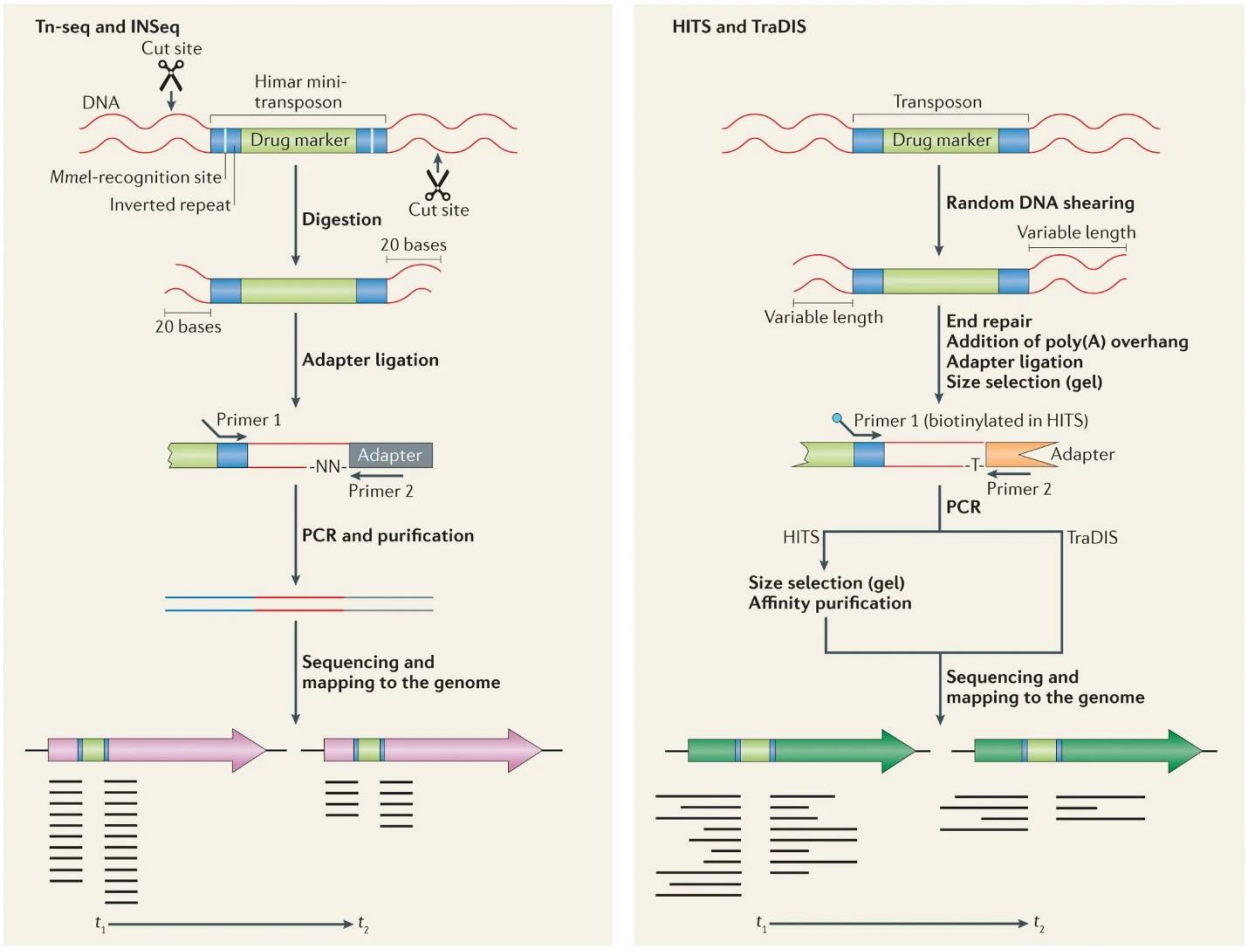


Figure 1.2. Library construction in transposon insertion sequencing.

Tn-seq/IN-seq and HITS/TraDIS utilize differing protocols for next generation sequencing library construction, but ultimately are conceptually similar and can make use of very similar post-sequencing analytic pipelines. Adapted from (10).

Bacterial genomics

A genome consists of the genetic information contained within the organism. In bacteria, the genome consists of the chromosome(s) and any associated extra-chromosomal, non-essential DNA, e.g. plasmids, chromatids. For many bacterial phenotypes, knowledge of the genotype can be very informative. The development of NGS, and the relatively new third generation (long-

read) sequencing technologies, have spurred a surge in the number of newly sequenced microbial genomes. Given the relatively small size of bacterial genomes (typically around 2-7 Mb), for many important model organism and pathogenic microbial species, sequenced genomes are available for thousands of strains. These vast amounts of data are often publicly accessible in nucleotide databases via online resources such as the NCBI in the US and the ENA in Europe. Most of these sequenced genomes have also been functionally annotated with genetic features such as the locations of protein-coding genes and key RNA loci (e.g. rRNAs and tRNAs).

The majority of available bacterial genomes are sequenced and assembled from short read (50-300 bp) technology on Illumina-based platforms. In genome assembly, short reads are aligned to each other to form longer contigs which can range in size from a few hundred nucleotides to 10s-100s of kilobases. The data gleaned from short reads results in contigs that will not approach the size of an entire chromosome (on the order of Mbs); rather, multiple contigs together span an entire chromosome, and the relative position of contigs to each other cannot be accurately determined from short reads. Hence, the genomes assembled from short reads contain both the genes in a bacterial genome and some contextual information about the relative genomic position of these genetic loci. The rise of third-generation, long read sequencing technologies, such as PacBio and Nanopore, is now enabling rapid generation of closed, completed genomes, which represent an improvement over the multi-contig, splintered genomes assembled from short reads. Long reads (1-10 kb) enable the construction of entire chromosome(s) or other replicons (e.g. plasmids) present in a bacterium. While these newer approaches are valuable, their expense and overall availability still lags behind those of short read sequencing technologies. Fortunately, for most genomics work, completed genomes are not

required, as the most valuable pieces of information in a bacterial genome are typically the sequence and function of its encoded genes, and their local genomic context.

In the current 'post-genomics' era, the amount of available genomics data provides a wealth of genetic and allelic diversity within a single species and has numerous applications. One area in which genomics is influential is in the determination of the evolutionary relationships between bacteria, often called 'phlyogenomics' (11, 12). In this area, researchers rely on genotypic differences, often among broadly conserved and shared genes, to determine the evolutionary distance between species and strains. In addition to clarifying the order of branches on the tree of life, phylogenomics has revolutionized the epidemiological study of pathogens and has been particularly useful in tracing the sources of outbreaks. Combining the genetic differences between strains of a pathogen and their associated metadata, such as geographic location and year of isolation, the spread of pathogens locally or globally can be mapped with high spatial and temporal resolution (13–15). For research on the spread of antibiotic resistance, this approach can help track the movement of antibiotic resistance genes and alleles (16, 17). More recent work has harnessed allelic diversity in candidate vaccine antigens to aid in the design of vaccines with efficacy across a broader range of strains (18). Finally, genome-wide association studies (GWAS) are now being performed in bacterial species to identify genes or variants associated with various phenotypes, such as antibiotic resistance (19–21).

The wealth of genomic data has also helped redefine the concept of a bacterial species. As genomic data for many strains of a species accumulated, it became clear that there was high intra-species diversity among bacteria (22). Such observations have given rise to the concept of the pan-genome, which represents the collection of all genetic elements found within a bacterial species (a genome refers to the genetic information contained in only a single strain). A species'

pan-genome contains three types of genes, core genes, which are near-universally present in strains of the species, shell genes, which are highly conserved among strains and occur frequently, and cloud genes, which are not well conserved across strains and occur rarely in strains (Figure 1.3) (23, 24). The pan-genome concept has helped researchers identify genes that contribute to the virulence of a pathogen or endow it with the ability to cause disease in a particular site in a host organism. This has allowed researchers to describe new groups of virulent bacterial strains based on the presence of specific genetic loci (25, 26).

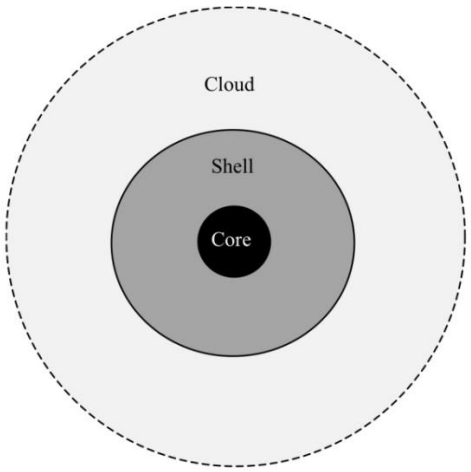


Figure 1.3. The bacterial pan-genome model.

A pan-genome consists of three types of genes – core genes, universally present in all strains; shell genes, highly conserved across strains and occur frequently; and cloud genes, not conserved across strains and occur rarely. The relative number of genes in each group is illustrated by the size of the circles, with the number of genes increasing from core to shell to cloud. Adapted from (23).

For example, the approach has been utilized to identify genes in pathogenic *E. coli* strains that colonize the intestine and are associated with necrotizing enterocolitis, a potentially fatal disease of newborns (27); and gene families and pathways associated with pathogenic O104:H4 *E. coli* strains that caused an outbreak of foodborne illnesses (e.g. bloody diarrhea) in 2011 in Germany (28). Overall, associating the presence of genes in particular strains with specific phenotypes has enabled researches to identify novel genes associated with the phenotype. These genes can subsequently be used in a molecular diagnostic setting to predict whether a host associated microbe is pathogenic or increases disease risk.

Animal models of infection

The study of host-pathogen interactions has been greatly advanced by utilizing animal models of infection to simulate disease pathogenesis and aid in the development of new therapeutics and vaccines. Such approaches bypass the requirement to study disease in the original host, often humans, which is frequently technically and/or ethically infeasible. In animal models, various species (e.g. mice, fruit flies, and zebrafish) are infected with the human infectious agent using a natural or an artificial mode of inoculation, and the resulting disease and organ pathology is monitored and characterized in the animal. A wide range of histologic and physiological measurements can be obtained during infection, and include biopsies of numerous organs and tissues, host immune responses (e.g. host transcriptional response, cytokine levels, immune cell activation), changes in host metabolites, and infectious agent burden. Infection with mutants of the pathogen allows dissection of the role of various virulence factors during disease; conversely infection of mutant hosts, e.g. knockout mice, can provide insights into host restriction factors and pathways involved in the activation of host immune responses.

The main alternative to animal models that is used to study host-pathogen interactions is tissue cultured cells. While this system is ideal for dissecting fine molecular mechanisms underlying the interactions between pathogens and host cells, key limitations include: the lack of complex tissue structure, which prevents the assessment of inflammatory responses and pathology; use of a single type of non-primary, immortalized host cell; and the inability to evaluate pathogen and host factors involved in the development of disease (e.g. diarrhea) and clinical signs (e.g. reductions in body weight). The two former limitations arise because traditional cell culture systems are composed of only one type of host cell, which is typically an immortalized cancer cell line, e.g. HeLa or HT-29, and usually do not form the defined supra-cellular structures seen in tissues in an organ. While these types of cell lines are widely utilized due to their ease of use, conclusions drawn from experiments with these cells can be erroneous due to underlying abnormalities in the host cells and lack of an organized, tissue-like structure. Furthermore, the lack of diverse cell types in most tissue cultured cell systems limits simulation of host tissue, immune, and physiological responses to infection. Recent improvements in the field have started to address some of these concerns, largely through the use of primary cells and organoids, miniaturized versions of organs that contain multiple differentiated types of primary cells and retain the three-dimensional structure of the original/parent tissue (29, 30). However, these systems still lack many features found in organs in vivo, such as numerous additional components of tissue, e.g. vasculature and lymphatics, which can provide many systemic factors during infections (acute phase proteins and cytokines), nerves, whose previously underappreciated roles in infections are now being characterized (31), and immune cells. Hence, small animal models continue to serve important roles that have yet to be superseded by traditional or modern in vitro tissue culture systems.

Chief among these has been to understand pathogen and host factors that contribute to the development of disease and pathology, largely made possible via genetic ablation and subsequent complementation of specific genes. These analyses have deepened our understanding of virulence factors of pathogenic bacteria, which are proteins or other molecules produced by bacteria that are required for the pathogen to cause disease. In addition, they have aided in the identification of novel host factors involved in various aspects of innate and adaptive immune systems or required for maintaining homeostasis. For the latter studies, animal models that utilize model organisms have been useful due to the relative ease of genetic manipulation in these organisms. Historically, animal models have also been valuable to rapidly determine the pathogenic potential of an organism – either through determining whether the organism causes disease or merely whether the organism stimulates an inflammatory response, e.g. screening for isolates of *E. coli* that can cause diarrhea (32) or the relative magnitude of virulence of closely related strains (33, 34). More modern uses of animal models leverage high-throughput and/or multi-dimensional measurements, which are often NGS based, to analyze host-pathogen interactions. For example, a common approach to identify microbial factors required for colonization of a host tissue involves conducting an in vivo Tn-seq screen. Additional approaches being pursued are the transcriptomic changes in either or both the pathogen and host tissues during the course of infection.

While animal models of infection ideally are perfect replicas of natural infections, this is only true for few diseases and in general, much care must be taken to choose a suitable model to study the topic of interest and in the generalizability of the conclusions. For example, while non-human primates (NHPs) are excellent models for many human diseases, due to their expense and many ethical issues, they are not widely used for host-pathogen infection studies. Small animal

models of infection are more heavily relied on, and the model of choice often utilizes model organisms, typically mice, zebrafish, or flies. While model organisms are convenient to use, in many instances model organisms are not the most suitable or accurate models of infection. Development and use of non-model organism models of infection can be challenging but have greatly furthered research efforts and likely will provide more translatable research findings. A few examples include: infant rabbits, which are the animal model of cholera that most completely reproduces the features of human infection (35) and have been utilized by our lab to develop a novel class of cholera vaccine (36); guinea pigs, which have been widely utilized as models of *Shigella* intestinal infection (37–39); and Golden Syrian hamsters, which are a commonly used and highly relevant model for *Clostridium difficile* infection (40). The drawback of these latter organisms is the lack of genetically modified strains, which limits analysis of host factors involved in infection. Hence, the type of experiment can constrain the animal models available for use. Notably, studies on vaccine development require a mature immune system and development of an immune response, and so are restricted to adult, immunocompetent animal models. While much research has been devoted to development and testing of animal models of infection, for certain diseases, no model yet exists that accurately captures the majority of aspects of the disease. Hence, development of improved animal models is still an active area of research.

Pathogenic bacteria

In this thesis, I will discuss research performed on two bacterial pathogens of global relevance, the *Streptococcus equi* group and *Shigella*. In the following sections, I will provide relevant background information for these two pathogens.

The *Streptococcus equi* group

Streptococcus equi group are Gram-positive, β-hemolytic, Lancefield group C streptococci (GCS) (41). These mammalian-associated bacteria can be either pathogenic or commensal organisms with pathogenic potential (i.e. opportunistic pathogens). Microbiologically, both SEE and SEZ are hemolytic on blood agar plates, and SEE is distinguished from SEZ by its inability to ferment lactose, sorbitol, and trehalose, whereas SEZ can ferment at least one (42). Taxonomically, the *S. equi* group is subdivided into two subspecies, *S. equi* subspecies *equi* (SEE) and *S. equi* subspecies *zooepidemicus* (SEZ). SEE is often associated with disease, while SEZ is a commonly found mucosal commensal of the nasopharyngeal and endometrial tissues, and an opportunistic pathogen (41). Multilocus enzyme subtyping and ribosomal (rRNA) sequences have indicated that SEE is a subtype of SEZ (43, 44), and this has been confirmed with whole genome sequences of SEZ and SEE (45, 46). The two subspecies are very closely related, with over 98% identity between their genomes (47). The genome is contained in a 2 Mb circular chromosome that contains ~2000 open reading frames (45). Evolutionarily, SEE and SEZ are closely related to *S. pyogenes* (Figure 1.4), which are Lancefield group A streptococci (GAS) (48). SEE, SEZ, and GAS bacteria are part of the pyogenic group of streptococci.

Host range and clinical presentation

SEZ and SEE have differing host range and clinical presentations. SEZ is an opportunistic pathogen that causes infections in a wide range of economically important livestock and domesticated mammals, including pigs (particularly in Asia), ruminants (e.g. sheep), horses, dogs, and cats. SEZ bacteria have been responsible for severe large-scale outbreaks in pigs, for

example an outbreak in Sichuan, China caused ~300,000 fatalities over the course of two weeks (49, 50). The genome of this outbreak strain has been sequenced (49), and its virulence mechanisms are under investigation (e.g. chapter 2). Recently, the first high mortality outbreaks of SEZ in swine in North America were reported. In Manitoba, Canada in April 2019, an outbreak resulted in many deaths and abortions (51); in Ohio and Tennessee, an outbreak occurred in September-October 2019, where mortality rates reached 50% (52). Zoonotic infections of SEZ in humans occur infrequently (53) and are typically limited to people who are exposed to unpasteurized milk or contaminated animal products (41). In contrast, SEE is a specialized pathogen that only infects horses, where it causes the disease 'strangles', one of the most highly prevalent and contagious infectious diseases in horses (53).

The clinical presentation of SEZ infections in animals is variable and can range from pneumonia, septicemia, meningitis, and, in horses, endometritis. SEZ infections are the leading cause of endometritis in mares and can cause reduced fertility or infertility (53). SEZ infections are also an emerging cause of infectious disease in dogs, where it is responsible for outbreaks (54). In pigs, SEZ can cause diarrhea, bronchopneumonia, endocarditis, and meningitis, and can be rapidly fatal (50, 55). In ruminants (e.g. sheep, goats, cattle, alpaca), SEZ predominately causes mastitis, although it can also cause deep tissue infections. Zoonotic infections of SEZ in humans results in a range of clinical presentations from purulent abscesses to endocarditis to meningitis (41); the specific disease presentation depends on the mode of transmission, the patient's immune status, and the virulence properties of the infecting SEZ strain.

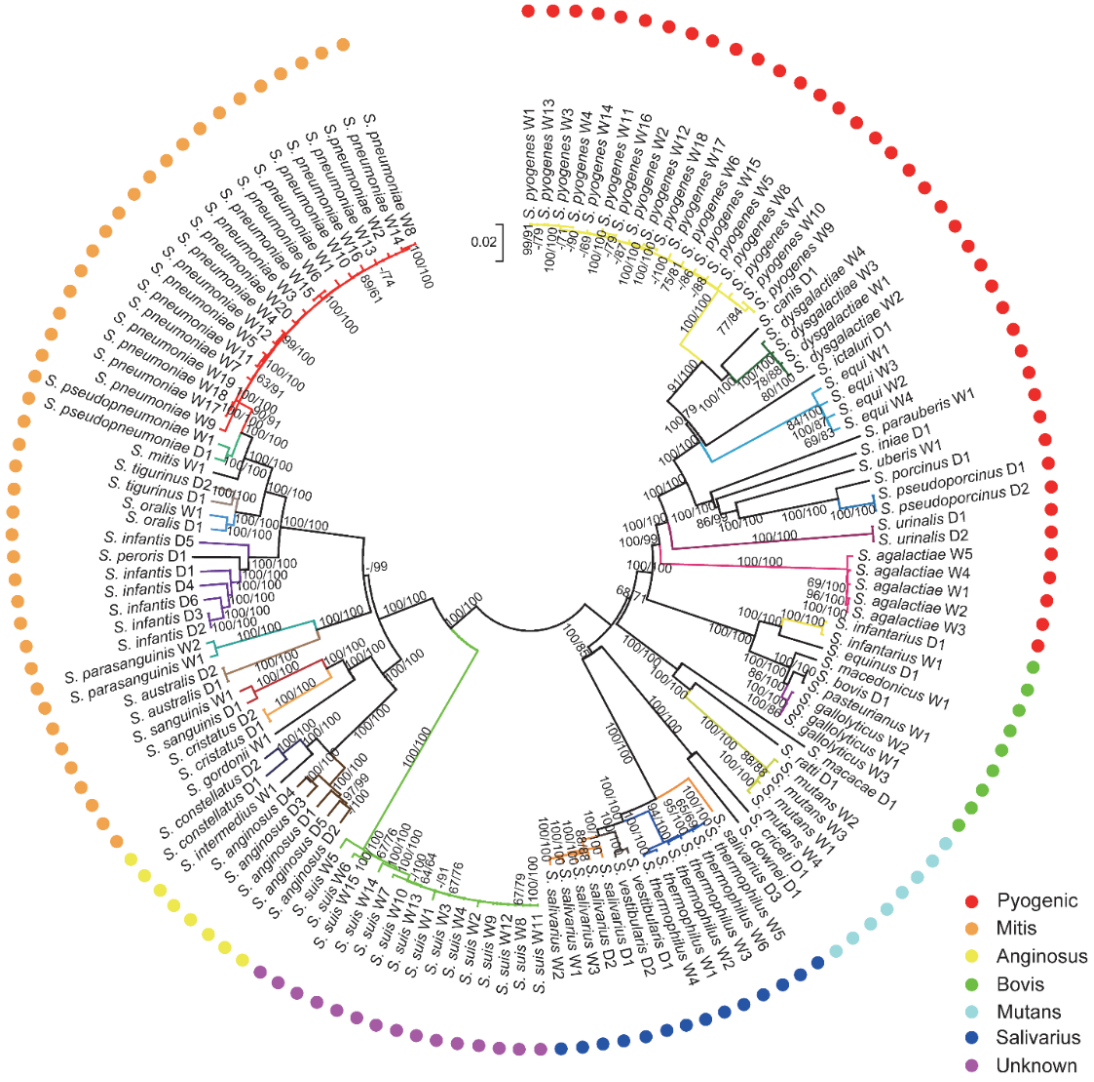


Figure 1.4. Phylogenetic tree of *Streptococcus*.

The phylogeny of 138 *Streptococcus* strains was constructed using 278 orthologous proteins, using the maximum likelihood (ML) and neighbor-joining (NJ) algorithms. The bootstrap values listed on the branches represent the ML and NJ bootstrap values (formatted as ML/NJ). The branches are colored according to species. Each strain was also assigned to a larger, phenotypic group, which is indicated with the colored circle at the periphery of the tree.

Adapted from (48).

SEE infection of horses causes the severe upper respiratory tract infection known as ‘strangles’. This is a very common infection of horses, one study found it responsible for 30% of all reported horse infections (56), and highly contagious, necessitating quarantine of the infected animals to prevent rapid spread across a herd. The clinical presentation is characterized by early nasopharyngeal signs and lymphadenopathy. This evolves into lymphadenitis, which can lead to the formation of abscesses with draining sinus tracts. Horses also present with fever, which reduces upon abscess rupture and drainage, and depressed appetite. There are two common severe complications. Rupture of abscesses, termed ‘bastard strangles’, has a 10% fatality rate (53). Development of purpura haemorrhagica, characterized by petechial hemorrhage with edema of the limbs and eye lids from trapping of immune complexes in capillary beds, can lead to circulatory failure, which can be fatal (53).

Treatment of *S. equi* group infections with antibiotics is commonly unsuccessful despite antibiotic susceptibility *in vitro* (53). Hence, clinically, treatment relies on supportive measures and abscess drainage. Prevention of disease through vaccination is an alternative strategy to curb disease; a few vaccines for SEE and SEZ infections have already been developed, but are not wholly protective and can occasionally cause disease. Development of newer vaccines is an active area of research. Both live attenuated bacterial strains and single or multivalent protein antigens have been tested as vaccines. Protein components of putative vaccines have been selected from virulence factors and immunogenic proteins discovered in *S. equi* group strains, such as SeM/SzM (57–59).

Virulence factors in SEZ and SEE

Several virulence factors have been identified in pathogenic strains of SEZ. Like GAS, all strains of SEZ contain a surface capsule made of the polysaccharide hyaluronic acid; the biosynthetic enzymes are encoded in the *has* operon (60). In streptococci, the capsule usually confers anti-phagocytic activity to the bacteria; in SEZ the capsule may also promote adherence to cells, hence promoting commensal colonization of host tissues (61). Adhesion to the extracellular matrix (ECM) is also thought to contribute to colonization of streptococci, and SEZ encodes two proteins, Fnz and Sfs, that bind fibronectin, a major ECM component. Furthermore, Fnz is a protective antigen in a mouse model of SEZ infection (59).

SEZ also contains immunomodulatory proteins that likely help it survive and evade host immune responses. SEZ encodes the surface protein Zag that binds IgG antibodies of particular host species (62) via the Fc constant region of the antibody, thus limiting opsonization. Zag also binds albumin, which has been shown to be important for inactivating antimicrobial peptides in infection of other streptococci (63). Strains of SEZ also contain two IgG endopeptidases, IdeZ and IdeZ2, that cleave IgG antibodies of certain host species (64, 65). In addition, superantigens (SAgs) similar to those found in GAS are sometimes found in SEZ strains. Most SAgs are phage-encoded virulence factors that can non-specifically bind major histocompatibility complex II (MHC II) and T cell receptors (TCRs), resulting in the activation of immune cells and production of massive amounts of proinflammatory cytokines, leading to a ‘cytokine storm’ (41). This extensive cytokine release results in circulatory shock, which can lead to multi-organ failure and ultimately death.

Key immunomodulatory virulence factors in SEZ and SEE are the two encoded M-like proteins, SzM/SeM and SzP/SzPSe (listed as the ortholog in SEZ and SEE respectively) (Figure

1.5). M-like proteins are surface proteins found primarily in group A streptococcus (GAS) that are similar in function and architecture to M protein, the canonical GAS virulence factor. SzM and SzP are not homologs of GAS M or M-like proteins, and no homologs of GAS M or M-like proteins have been found in *S. equi* group strains (41). In SEZ/SEE, the SzM/SeM proteins are immunogenic, since antibodies against the proteins have been identified in sera from convalescent and previously infected animals (66). In a mouse model of infection, SzM has been shown to be a protective antigen against challenge with a virulent strain of SEZ that encodes the same SzM protein (58). However, in horses, immunization with SzM/SeM has not been completely effective against challenge with a virulent strain (67). SzM/SeM also binds fibrinogen, which is thought to help prevent phagocytosis of *S. equi* group bacteria by macrophages (68, 69). SzP/SzPSe binds fibrinogen as well, which contributes to anti-phagocytic activity, is required for virulence of SEE in a mouse model of infection (70), and mediates adherence to tissue cultured cells (55).

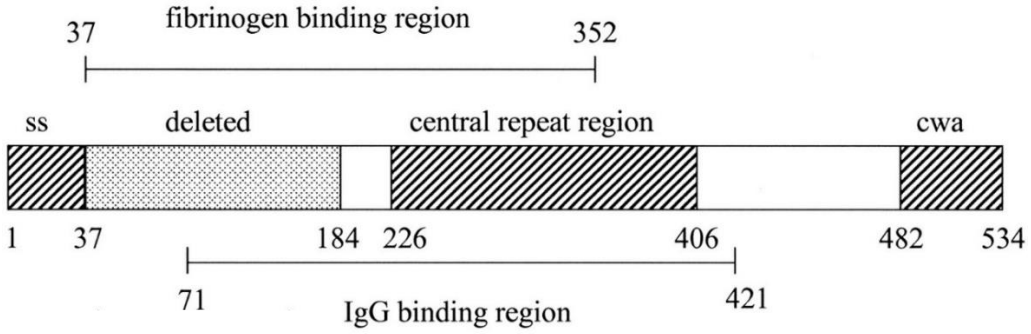


Figure 1.5. Schematic of domains of the M-like protein SeM.

SzM has a similar domain architecture. ss: signal sequence, deleted: region not found in SeM variants from SEE strains isolated from healthy animals, central repeat region: contains A and B repeat regions, cwa: cell wall spanning and anchor (i.e. LPXTG motif). Adapted from (71).

Shigella: a human-restricted pathogen

Shigella species are Gram-negative, rod-shaped, non-motile bacteria that cause bacillary dysentery in humans. Annually, the pathogen is responsible for millions of diarrheal cases and hundreds of thousands of deaths worldwide (Figure 1.6) (72). Despite its global importance, a *Shigella* vaccine does not exist and the prevalence of multidrug resistance strains is rising (73–75). Consequently, the World Health Organization has made development of a *Shigella* vaccine a priority.

Shigella infections, which can be caused by ingestion of even ≈100 bacteria (76), presents clinically as a debilitating, bloody diarrhea. The disease symptoms typically last around a week and are self-resolving, but in certain patient populations, e.g. malnourished children in developing nations, can lead to death. Pathology is marked by pathogen and immune-mediated

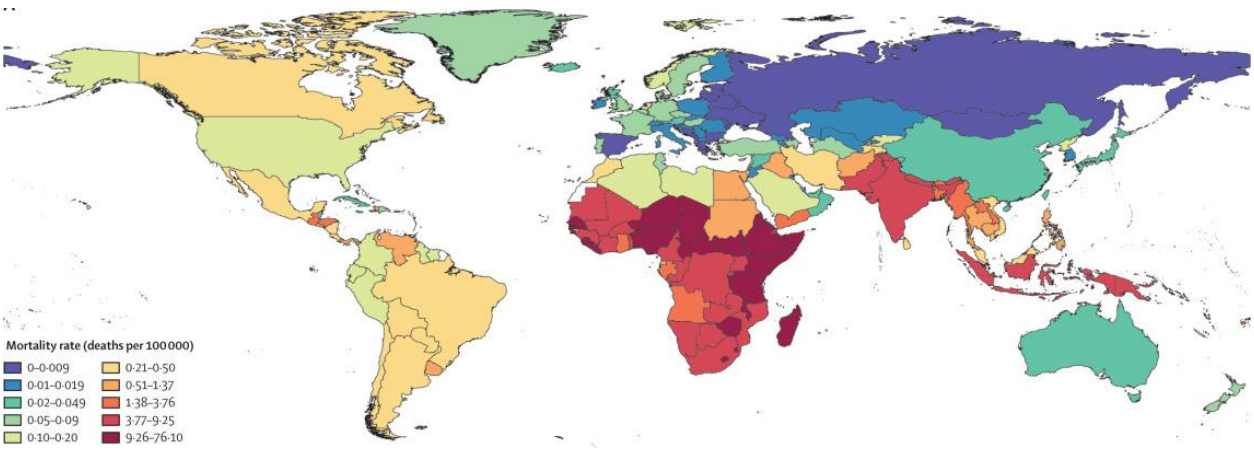


Figure 1.6. Distribution of mortality rate per 100,000 people of *Shigella* infections in 2016. Adapted from (72).

tissue destruction that is primarily restricted to the colon and largely due to *Shigella* infection of and spread through colonic epithelial cells. Colonic infection results in epithelial cell destruction and a severe acute inflammatory response.

There are four species of *Shigella.* From an epidemiology standpoint, *S. flexneri* is responsible for most cases worldwide (77), while *S. sonnei* is prevalent in developed nations and its prevalence is rising in transition nations (78). There are no known vectors for *Shigella*. The pathogen only causes disease in humans, and occasionally higher NHPs, and is spread through the fecal-oral route.

From an evolutionary and genomics perspective, *Shigella* are closely related to *E. coli*, but the four S*higella* species do not comprise a monophyletic group (Figure 1.7). A large (>200 kb), single copy virulence plasmid, and various chromosomal alterations distinguish *Shigella* from *E. coli* and render the bacteria pathogenic (79). The virulence plasmid carries genetic loci that are necessary and sufficient for invasion of host cells (80, 81). The eponymous Shiga toxin, which intoxicates eukaryotic cells by inhibiting the ribosome, is only found in strains of one of the fifteen serotypes of *S. dysenteriae*, serotype 1 (82).

Shigella **Epidemiology**

Amidst devastating outbreaks in Japan and other nations at the turn of the century, *Shigella* bacteria were identified as the infectious agent of bacillary dysentery in 1898 by Kiyoshi Shiga (83). *S. dysenteriae* serotype 1, which encodes Shiga toxin, was the original Shiga bacillus (83, 84). Currently, *Shigella* are separated into four species – *S. flexneri*, *S. sonnei*, *S. dysenteriae,* and *S. boydii*, but only a small fraction of all *Shigella* cases are attributable to *S. dysenteriae* and *S. boydii* (85). *S. flexneri* is the most predominate species worldwide and is

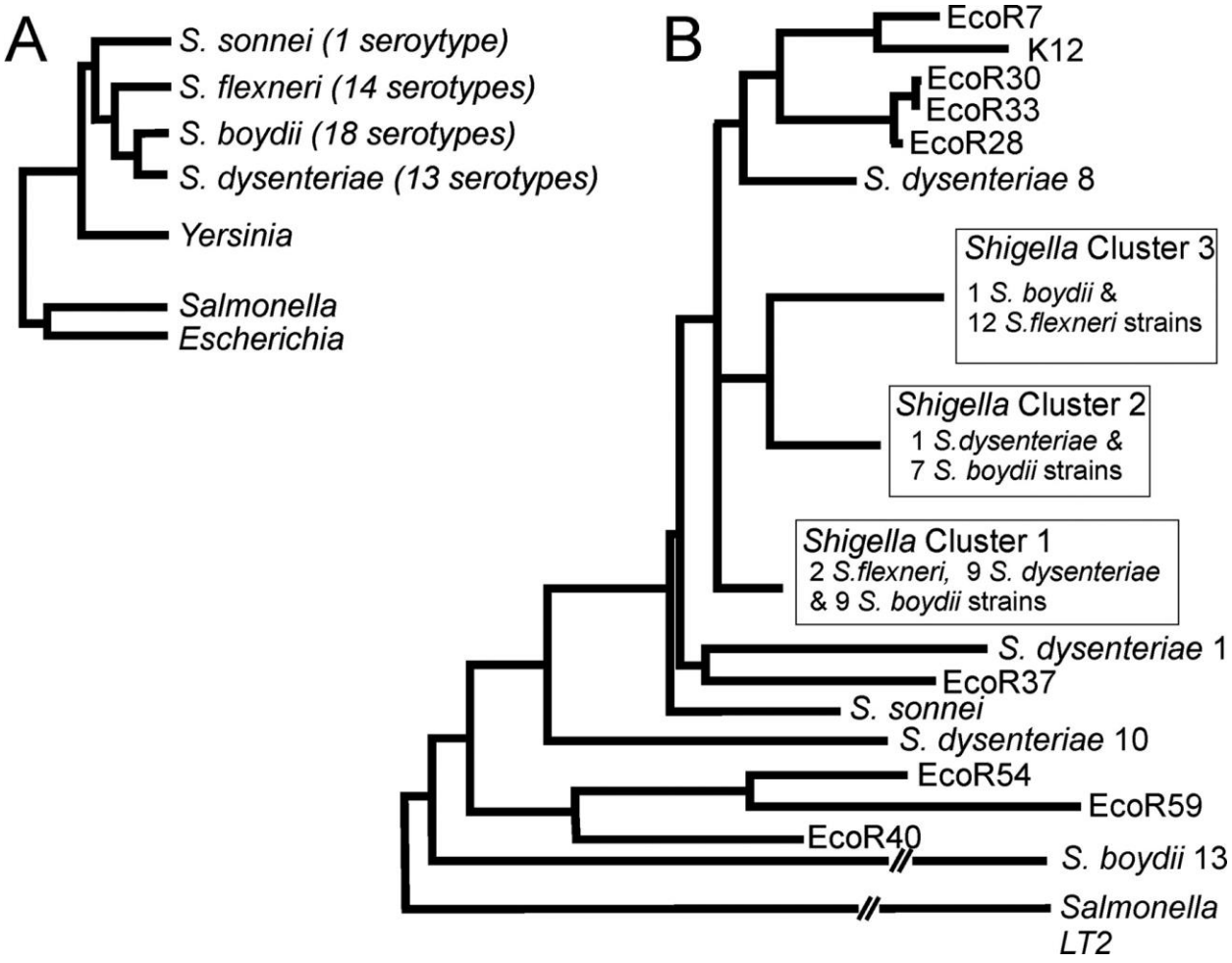
Figure 1.7. Phylogeny of *Shigella* and related Enterobacteriaceae species.

A. Classical phylogeny based on phenotypic characteristics. Note - there are now considered to be 15 serotypes of *S. dysenteriae*, 15 of *S. flexneri*, and 19 of *S. boydii* (82).

B. Modern phylogeny of *Shigella* species and several *E. coli* (EcoR) strains based on comparative genomics of several housekeeping genes. *Salmonella enterica* strain LT2 is used as an outgroup.

Adapted from (79, 86–88)

Figure 1.7. (Continued)



responsible for the largest number of cases, which primarily occur in Africa, and East and Southeast Asia (Figure 1.5). Historically, *S. dysenteriae* was widely prevalent and responsible for devastating outbreaks with high fatality rates (82). *S. dysenteriae* is no longer the global scourge it once was and is now responsible for only a small percentage of *Shigella* infections worldwide (85). Nevertheless, it is still a public health threat; over the past few decades, the species has periodically caused explosive outbreaks in unstable regions (82). The predilection for *S. sonnei* for developed nations (78) is thought to be due to an increase in a lack of immunity to the O-antigen of *S. sonnei*. Immunity against the O-antigen is thought to be generated through contact with contaminated water containing the distantly related Gram-negative bacteria *Plesiomonas shigelloides*, which has the same O-antigen as *S. sonnei* (89, 90).

Outbreaks of shigellosis occur worldwide and in developed countries certain groups are at risk. These include children and staff at day care centers, due to inadequate hand hygiene; closed communities, e.g. certain Ashkenazi Jewish populations (91); the immunocompromised; and men who have sex with men (MSM) (92, 93). Outbreaks involving multi-drug resistant strains of *S. sonnei* and *S. flexneri* have been reported (73–75).

Shigella **Disease and Pathology**

Shigella infection typically causes a self-resolving gastroenteritis with a colonic predominance. The course of infection typically begins with constitutional symptoms such as fever, and develops into overt diarrhea, which often contains blood and mucus, tenesmus (rectal pain and straining), and occasionally emesis (vomiting) (94). Typically, patients experience up to ten episodes of bowel movements per day. Patients have a watery diarrhea that precedes bloody diarrhea, but occasionally, the diarrhea can remain watery and non-bloody throughout the course

of the disease. The infection is self-resolving in most cases, and symptoms generally last for a week. Rarely, there can be more severe complications of disease such as rectal prolapse, colonic perforation, bacteremia, and reactive arthritis, specifically with *S. flexneri* infections (94). These complications typically occur in younger patients such as children or infants, or patients with risk factors, e.g. malnutrition. A notable complication of infection is seizures, which occur almost exclusively in children, are typically seen in infections with *S. flexneri* and *S. sonnei*, and can have a high prevalence; one study reported that 45% of patients experienced this complication (95). Patients infected with *S. dysenteriae* serotype 1 strains that encode Shiga toxin are at risk for developing hemolytic uremic syndrome, which is characterized by microangiopathic hemolytic anemia, thrombocytopenia, and acute renal failure (94). In pre-WWII Japan, *S. sonnei* infections caused a highly lethal disease called Ekiri syndrome, which was characterized by high fever, seizures, coma, and few diarrheal symptoms, and was responsible for 15,000 deaths annually (96, 97).

Histologically, *Shigella* infections are characterized by pathology generally limited to the colon. Much of our understanding of the pathology during infection is drawn from biopsy specimens of the colon and rectum of patients, typically from developed countries, and from experimental infections of non-human primates. Characteristic findings include mucosal erosions and ulcers, massive recruitment of neutrophils and occasionally mononuclear cells, congestion of blood vessels, hemorrhage, and distortion of colonic crypt architecture (98, 99). Pathology is largely limited to the superficial layers of the colon; pathogen invasion and damage of deeper layers is uncommon. Furthermore, bacterial infection and pathology is typically limited to the colon, with a lack of bloodstream invasion, dissemination, and seeding of other organs as seen in infections with *Salmonella*. Detailed studies on the progression and resolution of pathology

caused by *Shigella* infections are lacking, although it is likely that these proceed similarly to other types of acute inflammatory reactions of the intestine.

Lifecycle of the *Shigella* pathogen

At the cellular level, the interaction between *Shigella* and host cells has been elucidated in great detail over the past 30 years (Figure 1.8) (79). Upon attachment to epithelial cell surfaces and recognition of the plasma membrane, *S. flexneri* invades host cells by secreting effector proteins through the Type III Secretion System (T3SS). The bacteria induce local membrane ruffling by remodeling the actin cytoskeleton to produce a phagocytosis-like uptake into the host cell (79). Once inside the cell, the bacteria are initially located in a membrane bound, vacuole-like structure that is rapidly (<15 min) ruptured, a process that requires the transolocon proteins (79, 100–103). The bacteria then replicate in the host cytoplasm and display intracellular actin-based motility. The force generated by actin polymerization at one pole of the bacteria is sufficient to propel the bacteria into neighboring cells by forming membrane bound protrusions. Upon uptake by neighboring cells, the bacteria are contained in a double membrane vacuole, which is rapidly ruptured, permitting the bacterial replication cycle to begin anew inside another cell cytoplasm. Many host factors have been implicated in *Shigella* cellular invasion, inter-cellular spreading, and killing of epithelial cells. For example, host factors required for infection include key components of the actin cytoskeletal network such as Cdc42 (104) and ELMO (105), which are necessary for bacterial invasion; the inter-cellular adherens junction protein E-cadherin, which facilitates cell-to-cell spreading (106); and the intermediate filament vimentin, which is required for docking of the T3SS to the cell, an early step in the cellular invasion process (107).

In contrast to the stereotyped interactions with epithelial cells, interactions with cells of the immune system result in varied outcomes. *Shigella* are thought to encounter and invade macrophages in the lamina propria of the intestine, which they rapidly kill (108).

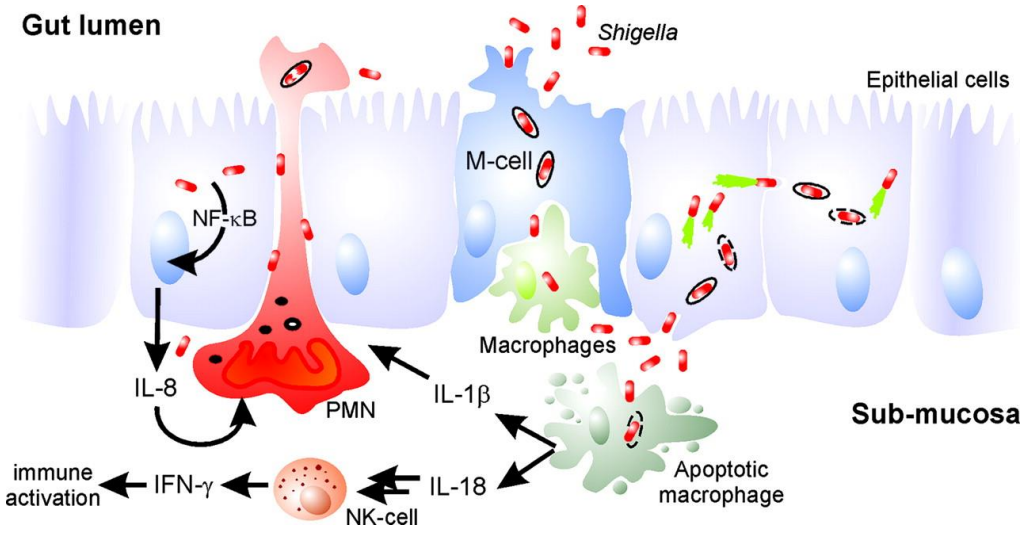


Figure 1.8. Lifecycle of *Shigella* in the colonic epithelium.

Shigella (red) invade the colonic epithelial layer, often through Microfold (M) cells, and are taken up by macrophages, which the bacteria rapidly kill. The bacteria then invade the colonic epithelial cells using the T3SS from the basolateral surfaces. The bacteria replicate and exhibit intracellular movement within the cells. The force of the actin-based movement propels the bacteria to form membrane protrusions, which are taken up by neighboring cells, allowing for cell-to-cell spread and secondary infection. In epithelial cells, bacterial products stimulate the activation of the NF-kB signaling pathway, resulting in the expression of innate immune proteins and cytokines such as IL-8. This chemokine recruits neutrophils (polymorphonuclear leukocytes [PMNs]) to the site of invasion, and these immune cells can kill *Shigella*. Other cytokines released at the site of infection activate NK cells, which produce IFN-gamma, which further stimulates the immune system. Adapted from (79).

Shigella kills lymphocytes by injecting effector proteins into the cells via the T3SS, which does not require cellular invasion (109). In contrast, *Shigella* is killed by neutrophils (110), likely through the action of the enzyme neutrophil elastase, which degrades components of the T3SS (111). The host factors mediating interaction between cells of the immune system and *Shigella* have not been completely characterized, but likely partially overlap with those required for epithelial cell invasion and death.

Virulence factors in *Shigella*

Pathogenesis relies on several key virulence factors; the vast majority are encoded in the virulence plasmid. A primary factor is the T3SS, a large, multi-component protein complex that acts as a molecular syringe to inject bacteria effector proteins from the bacterial cytoplasm into the cytoplasm of host cells (Figure 1.9). T3SS are commonly found in Gram-negative pathogens; some contain multiple T3SSs, which each have a separate suite of effector proteins (112). *Shigella* species contain only one T3SS, which is necessary and sufficient for invasion into host cells (103). Structurally the T3SS is composed of three parts – the stator anchor that spans both the inner and outer lipid membranes, the needle, and the translocon (113, 114). In *Shigella* infections, secreted effector proteins or structural components are required or promote most aspects of infection including: invasion into host cells (79), rewiring host pathways such as dampening the innate immune response (115), and cell-to-cell spreading (116). Expression of the T3SS is tightly regulated by environmental temperature, with the apparatus only expressed at the higher temperatures (37°C) typical of the human intestine and not at lower, environmental temperatures (25°C) (117). Many of the 30+ effector proteins secreted by the *Shigella* T3SS have been intensely studied and have varied functions such as serving as ubiquitin ligases to target

host proteins involved in innate immune responses for degradation (118) or dampening the activity of intermediary mediators in signaling pathways in innate immune responses, potentially by catalyzing unique, previously undescribed enzymatic reactions (119) (Figure 1.9).

Another crucial *Shigella* virulence factor is the surface protein IcsA. This outer membrane protein, which is localized to a single pole of the bacteria, is necessary and sufficient for intracellular actin-based motility of the bacteria (120). IcsA is a member of the type V secretion systems, autotransporter proteins that promote secretion of their own extracellular domains through a beta-barrel channel formed by their own membrane spanning domains (121, 122). The extracellular domain of IcsA stimulates polymerization of host actin, which form actin tails at one pole of the bacteria that permits their intracellular movement and provides the force for cell-to-cell spreading (120, 123, 124). More recently, Zumsteg et al. demonstrated that IcsA has adhesin-like activity, promoting attachment to epithelial cells, a process that requires T3SS activity and is promoted by the binding of a common human colonic bile salt, deoxycholate (125).

Shigella contains other virulence factors in addition to the T3SS and IcsA. The eponymous exotoxin, Shiga toxin, is uncommon in infecting strains as it is only found in *S. dysenteriae* serotype 1 strains, currently a rare cause of infections. Infections involving Shiga toxin are more likely to be caused by Enterohemorrhagic *E. coli* (EHEC), which encoded two toxins similar to Shiga toxin. The O-antigen of lipopolysaccharide (LPS) in *Shigella* is a key virulence factor and is immunogenic in humans (126–128). The O-antigen is the main protective immunogen in humans and antibodies against the O-antigen are thought to be effective, since infection with a specific serotype leads to resistance against repeated infection with the same serotype, but not other serotypes (126). Many *Shigella* strains contain up to two enterotoxins,

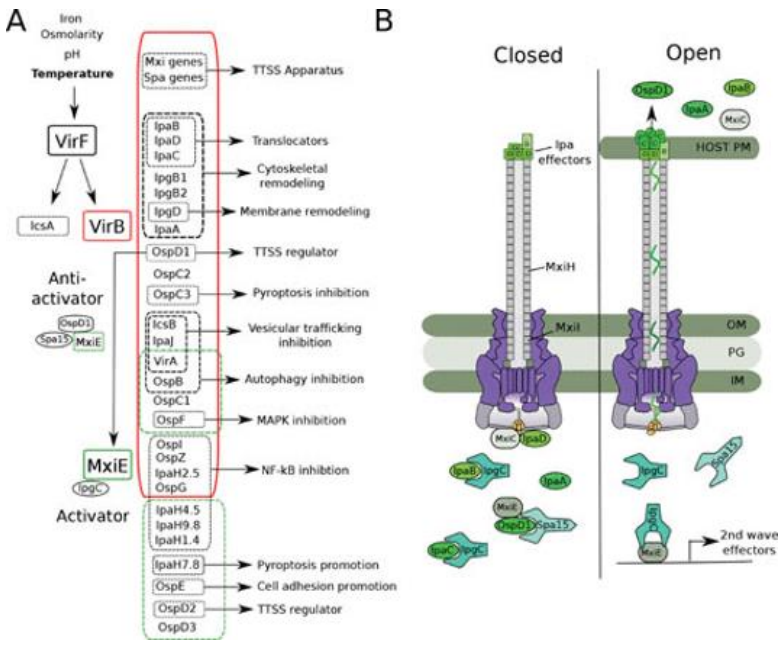


Figure 1.9. The regulation, function, and structure of *Shigella* type III secretion system (T3SS) and effector proteins.

A. General virulence regulation and functions of *Shigella* T3SS effector proteins. (Left) Simplified virulence regulation network in *Shigella*. (Middle) T3SS effectors, grouped based on their activation by a shared transcriptional regulator, VirB or MxiE. (Right) Function of T3SS effector protein. TTSS = T3SS.

B. Overall architecture and states of *Shigella* T3SS. Upon encountering and recognizing a eukaryotic plasma membrane, the translocon forms a pore on the host cell membrane permitting efflux of effector proteins from the bacterial cytoplasm into the host cell cytoplasm.

Adapted from (129).

called ShET-1 and ShET-2, which are encoded on the chromosome and virulence plasmid, respectively, and thought to cause fluid leakage in the intestine and potentially the watery diarrhea that precedes dysentery in most *Shigella* infections (130, 131). Additional virulence factors include aerobactin, an siderophore promoting iron acquisition that is encoded by the *iuc* operon (132), and the secreted serine proteases SigA and Pic, which are thought to cleave host factors, such as αII-spectrin (133) and human leukocyte adhesin proteins (e.g. CD43, CD44, CD45) (134), respectively.

While the molecular mechanisms of some of these virulence factors have been well studied, the connection between putative virulence factors and the development of disease and pathology has been less well characterized due to a lack of an appropriate small animal model of infection that recapitulates disease seen in human infections. Hence, there are many open questions regarding the roles of virulence factors in pathogenesis. Recently, new models of infection in infant rabbits by our group, which is the basis of chapter 3, and Yum et al. (135) have started providing deeper insights into the molecular pathogenesis of disease caused by *Shigella* infection.

Small animal models of *Shigella* infection

Investigation of the virulence factors that contribute to disease pathogenesis, the testing of novel therapeutics, and the development of vaccines to prevent *Shigella* infection have been hindered by the lack of a small animal model of infection that accurately captures disease seen in human infections. Numerous types of small animal models of *Shigella* infection have been described over the past several decades, but all are lacking in one or more aspects when compared with human infections. Finding a suitable model remains a challenge in the field,

which is partially attributable to the highly selective nature of *Shigella* infections which have both limited host range and tissue tropism. *Shigella* infections only occur in humans and cause intestinal pathology, primarily in the colon, since dissemination to other organs, e.g. via bacteremia, is rare. Key hurdles in the development of a small animal model have been the lack of development of disease (e.g. diarrhea) or intestinal pathology when animals are inoculated orally, and the resistance of mice, one of the most genetically tractable model organisms, to *Shigella* infection.

Shigella animal models can be roughly grouped based on the route of inoculation used; unfortunately, most models do not use the natural oral route due to the frequent lack of disease and pathology when animals are inoculated orally. The original small animal model of *Shigella* infection was the Sereny test (136), developed in 1955, in which *Shigella* are inoculated into the conjunctiva of an eye of an adult guinea pig, which results in keratoconjunctivitis (inflammation of the conjunctiva and cornea of eye) when virulent bacteria are used. While unnatural in many aspects, the Sereny test provided a reliable method to determine key virulence factors in *Shigella* stimulating an inflammatory response in the host (137). Another early model of *Shigella* infection was the orally infected adult guinea pig, however, to ensure development of pathology and disease, animals had to be starved for 3-4 days (138) or treated with carbon tetrachloride prior to inoculation (139). This model was not pursued further.

The next common model of *Shigella* infection, after the Sereny test, was the adult rabbit ligated ileal loop model. This technically complex model requires specialized abdominal surgery of large, adult rabbits; bacteria are directly injected into an isolated loop of the small intestine (81). The model was an improvement compared with the Sereny test since it queried a more anatomically relevant tissue, so infection-based pathology was more informative, and provided a

readout of host inflammatory and immune responses. The model was used for an in vivo genetic screen that utilized signature tag mutagenesis (STM), a precursor to Tn-seq. The research identified genes required for survival in the presence of a host inflammatory response and thus discovered an unusual role for prophage genes in pathogenesis (140). In addition to adult rabbits, young (four-week-old) rabbits have been used as models of *Shigella* infection, in which animals were inoculated via an orogastric tube and subsequently develop small intestinal enteritis without diarrhea. The model was used to demonstrate the protective effects of proteins in milk (e.g. lactoferrin) (141).

Several model organisms have been used as models of *Shigella* infection, with varying degrees of success. Mice, the most widely used and genetically tractable mammalian small animal model, are resistant to oral *Shigella* inoculation. All adult mice, including conventional (i.e. specific pathogen free), streptomycin treated mice, and germ-free mice, are resistant, although GF mice can be colonized without developing disease (142). The lack of susceptibility is thought to be partially attributable to the absence of the CXC chemokine interleukin 8 (IL-8), a primary neutrophil recruitment factor, and its receptor in mice (143). A human intestine xenograft model did provide a technically complex model of *Shigella* infection (144), as xenograft tissue developed appropriate pathology after infection. Unlike adult mice, infant mice are susceptible to *Shigella* infection, and oral infection of infant mice results in intestinal pathology. However, pathology occurs primarily in the small intestine, and furthermore, all infected animals die within several hours of inoculation. Death due to infection requires the presence of the virulence plasmid in the infecting strain and thus may be dependent on canonical virulence factors. While the mechanism of death has not been clarified, the rapidity at which death occurs suggests that mortality arises through bacteremia and not intestinal disease.

More recently, a non-mammalian, zebrafish larvae model has gained traction. In this model, zebrafish larvae are infected via intra venous (IV) injection, which results in a dose-dependent dissemination of the bacteria and T3SS-dependent lethality (145). Despite the many differences compared with natural infections, the value of this model lies in the ability to perform live imaging of host-pathogen interactions in vivo due to the larvae's translucent body. Hence, the model permits detailed live microscopic analysis of host innate immune responses to *Shigella* infection in vivo (146–148), a much more relevant and complex environment than in vitro tissue cultured cells.

In addition to zebrafish, another non-mammalian model organism, *C. elegans*, has been explored as a model of *Shigella* infection. Oral feeding of virulent *Shigella* to *C. elegans* is lethal to the nematode and the bacteria accumulate in the lumen, but likely do not invade into host cells (149, 150).

The small animal models of *Shigella* infection that have most accurately replicated the bloody diarrhea seen in human infections have been those utilizing intra-rectal modes of inoculation. Two intra-rectal models have been developed, in adult guinea pigs and infant rabbits. In the first model, developed in adult guinea pigs, animals infected with virulent strains have bloody liquid feces and rectal pathology characterized by an acute inflammatory response (37). Hence, the model captures several salient aspects of human infection, notably bacillary dysentery. The model has been widely used to deepen our understanding of many aspects of *Shigella* infection in vivo, e.g. high-resolution imaging analysis of *Shigella* localization in the colon (38), dynamics of T3SS activity in vivo (151), function of various host factors, neutrophils, and oxygen during infection (39, 152). The second intra-rectal model, which utilized infant rabbits, was more recently developed. In the paper, Yum et al. were able to extend our

understanding of the molecular mechanisms of *Shigella* pathogenesis by utilizing isogenic single gene deletion mutant strains of bacteria to define the contributions of two individual virulence factors in disease pathogenesis (135). The individual roles of these virulence factors had not been studied in the guinea pig model. In the infant rabbit model, infected animals develop grossly bloody diarrhea and rectal pathology reminiscent of those seen in human infections. The paper provided evidence that both IcsA and T3SS were required for diarrhea, and that IcsA was required for widespread epithelial cell destruction and sloughing, but not for stimulating an acute inflammatory response in the rectum. Hence, the authors concluded that tissue destruction caused by the pathogen plays a more important role in the development of diarrhea and pathology than the host immune response. This was unexpected since prior research, using other animal models, had suggested that the host acute inflammatory response to infection drove pathology and disease.

The majority of the small animal models of *Shigella* infection have focused on studying disease pathogenesis, and only a few are available to evaluate candidate vaccines. Models used for vaccine development often require immunization of adult animals, as infant animals lack mature immune systems and cannot mount strong adaptive immune responses. Furthermore, since adaptive immune responses take time to develop (typically days to weeks), animals will have progressed from the infant development phase when it is appropriate to challenge the vaccinated animal. Among the previous described animal models, only the adult guinea pig model has been shown to provide a platform for testing vaccines. Notably, immunization also occurs at the intestinal mucosal surface, since vaccination is performed via intra-rectal delivery of the vaccine (37). The commonly used vaccine model for *Shigella* infection is the pulmonary infection model in adult mice where intra-nasal inoculation with virulent strains results in lung

disease and mortality (153, 154). In this model, mice develop pneumonitis and an acute inflammatory response in the lungs that resembles responses seen in intestinal infections in humans; furthermore, *Shigella* bacteria can invade the alveolar epithelial cells. By leveraging the power of mouse genetics, the model has been transformative in furthering our understanding of specific factors and cells of the immune system that are involved in overcoming *Shigella* infections (155). Another potential adult mice model for vaccine development is the intra-peritoneal model, in which animals develop intestinal pathology and almost all mice succumb to infection; despite the ability to test vaccines using this model, which can protect mice from lethal infection, this model has not been widely adopted for vaccine testing (156). Finally, human challenge studies are still common and viable options for testing leading vaccine candidates, and these clinical trials have been very informative for *Shigella* vaccine development (126, 157–159).

Apart from humans, higher NHPs, e.g. rhesus macaques (160), chimpanzees (161), gibbons, and gorillas (162), are the only other animals that can naturally develop shigellosis and spontaneous cases of shigellosis have been reported in zoo animals. Experimental infection of NHPs with *Shigella* has contributed substantially to our understanding of the intestinal pathology caused by infection and provided a platform for the development of early vaccines. However, due to their high cost, NHPs are not feasible models of *Shigella* infection that can be used to study nuanced and detailed aspects of pathogenesis or test additional vaccine candidates. Furthermore, while NHPs can spontaneously contract shigellosis, in experimental settings, NHPs require very large inoculation doses to reliably develop disease. This is in stark contrast to the situation in humans, where even healthy adults can develop disease from an extremely low dose, e.g. one report demonstrated that ≈100 organisms can cause disease in 40% of volunteers, while

≈1000 organisms resulted in a disease rate >50% (76). The human-specific nature of *Shigella* infections likely contributes to the challenge of developing small animal models of infection. The surprisingly small dose required for human disease strongly suggests that there is a specific molecular basis for host restriction, yet despite decades of research it remains largely unknown. Recent evidence has shed some light on this issue by demonstrating that a human-specific intestinal factor, an antimicrobial peptide that is a member of the defensin family of proteins (defensin 5), may dramatically stimulate adhesion of *Shigella* to intestinal epithelial cells. In doing so, this process greatly promotes cellular invasion and disease, and permits lower doses of *Shigella* to cause disease in an animal model (intra-rectal guinea pig model) (39).

Conclusions

Bacterial pathogens cause various infectious diseases in a range of hosts through multiple different mechanisms utilizing diverse virulence factors. Understanding these host-pathogen interactions holds potential to deepen our understanding of basic biological processes and to develop strategies to tackle these infections. Necessarily, this research requires thorough investigation of the nature and regulation of virulence factors as well as the use of relevant small animal models that accurately mimic disease in the original host. More than ever, such research requires interdisciplinary approaches.

In this thesis, I discuss our discoveries on host-pathogen interactions of two globally important bacteria, *Streptococcus equi* subspecies *zooepidemicus* (SEZ) and *Shigella*. In chapter 2, I discuss our findings regarding virulence regulation in SEZ and discovery of a conserved locus that defines a group of pathogenic streptococci whose composition crosses traditional taxonomic classifications. In chapter 3, I discuss the development of a new small animal model

of *Shigella* infection involving oral inoculation of infant rabbits and insights gained regarding pathogenesis using defined mutant strains. In chapter 4, I summarize our results and discuss attractive potential future directions based on our work.

References

1. **Chao MC, Abel S, Davis BM, Waldor MK**. 2016. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol* **14**:119–128.

2. **Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJ, Rubin EJ, Waldor MK**. 2014. ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet* **10**:e1004782.

3. **Hubbard TP, D'Gama JD, Billings G, Davis BM, Waldor MK**. 2019. Unsupervised Learning Approach for Comparing Multiple Transposon Insertion Sequencing Studies. *mSphere* **4**:e00031-19.

4. **Zomer A, Burghout P, Bootsma HJ, Hermans PWM, van Hijum SAFT**. 2012. ESSENTIALS: Software for Rapid Analysis of High Throughput Transposon Insertion Sequencing Data. *PLoS One* **7**:e43012.

5. **Barquist L, Mayho M, Cummins C, Cain AK, Boinett CJ, Page AJ, Langridge GC, Quail MA, Keane JA, Parkhill J**. 2016. The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics* **32**:1109–11.

6. **Hubbard TP, Chao MC, Abel S, Blondel CJ, Abel zur Wiesch P, Zhou X, Davis BM, Waldor MK**. 2016. Genetic analysis of *Vibrio parahaemolyticus* intestinal colonization. *Proc Natl Acad Sci* **113**:6283–6288.

7. **Warr AR, Hubbard TP, Munera D, Blondel CJ, Abel Zur Wiesch P, Abel S, Wang X, Davis BM, Waldor MK**. 2019. Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization. *PLoS Pathog* **15**:e1007652.

8. **McCarthy AJ, Stabler RA, Taylor PW**. 2018. Genome-wide identification by transposon insertion sequencing of Escherichia coli K1 genes essential for in vitro growth, gastrointestinal colonizing capacity, and survival in serum. *J Bacteriol* **200**:1–19.

9. **Abel S, Abel zur Wiesch P, Davis BM, Waldor MK**. 2015. Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog* **11**:e1004823.

10. **van Opijnen T, Camilli A**. 2013. Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. *Nat Rev Microbiol* **11**:435–42.

11. **Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF**. 2016. A new view of the tree of life. *Nat Microbiol* **1**:16048.

12. **Zhu Q, Mai U, Pfeiffer W, Janssen S, Asnicar F, Sanders JG, Belda-Ferre P, Al-Ghalith GA, Kopylova E, McDonald D, Kosciolek T, Yin JB, Huang S, Salam N, Jiao J-Y, Wu Z, Xu ZZ, Cantrell K, Yang Y, Sayyari E, Rabiee M, Morton JT, Podell S, Knights D, Li W-J, Huttenhower C, Segata N, Smarr L, Mirarab S, Knight R**. 2019. Phylogenomics of 10,575 genomes reveals evolutionary proximity between domains Bacteria and Archaea. *Nat Commun* **10**:5477.

13. **Weill F-X, Domman D, Njamkepo E, Tarr C, Rauzier J, Fawal N, Keddy KH, Salje**

H, Moore S, Mukhopadhyay AK, Bercion R, Luquero FJ, Ngandjio A, Dosso M, Monakhova E, Garin B, Bouchier C, Pazzani C, Mutreja A, Grunow R, Sidikou F, Bonte L, Breurec S, Damian M, Njanpop-Lafourcade B-M, Sapriel G, Page A-L, Hamze M, Henkens M, Chowdhury G, Mengel M, Koeck J-L, Fournier J-M, Dougan G, Grimont PAD, Parkhill J, Holt KE, Piarroux R, Ramamurthy T, Quilici M-L, Thomson NR. 2017. Genomic history of the seventh pandemic of cholera in Africa. *Science* **358**:785–789.

14. **Domman D, Quilici M-L, Dorman MJ, Njamkepo E, Mutreja A, Mather AE, Delgado G, Morales-Espinosa R, Grimont PAD, Lizárraga-Partida ML, Bouchier C, Aanensen DM, Kuri-Morales P, Tarr CL, Dougan G, Parkhill J, Campos J, Cravioto A, Weill F-X, Thomson NR**. 2017. Integrated view of Vibrio cholerae in the Americas. *Science* **358**:789–793.

15. **Domman D, Chowdhury F, Khan AI, Dorman MJ, Mutreja A, Uddin MI, Paul A, Begum YA, Charles RC, Calderwood SB, Bhuiyan TR, Harris JB, LaRocque RC, Ryan ET, Qadri F, Thomson NR**. 2018. Defining endemic cholera at three levels of spatiotemporal resolution within Bangladesh. *Nat Genet* **50**:951–955.

16. **Baker S, Thomson N, Weill F-X, Holt KE**. 2018. Genomic insights into the emergence and spread of antimicrobial-resistant bacterial pathogens. *Science* **360**:733–738.

17. **Holden MTG, Hsu L-Y, Kurt K, Weinert LA, Mather AE, Harris SR, Strommenger B, Layer F, Witte W, de Lencastre H, Skov R, Westh H, Zemlicková H, Coombs G, Kearns AM, Hill RLR, Edgeworth J, Gould I, Gant V, Cooke J, Edwards GF, McAdam PR, Templeton KE, McCann A, Zhou Z, Castillo-Ramírez S, Feil EJ, Hudson LO, Enright MC, Balloux F, Aanensen DM, Spratt BG, Fitzgerald JR, Parkhill J, Achtman M, Bentley SD, Nübel U**. 2013. A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic. *Genome Res* **23**:653–64.

18. **Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, Towers RJ, Duchêne S, Smeesters PR, Frost HR, Price DJ, Holden MTGG, David S, Giffard PM, Worthing KA, Seale AC, Berkley JA, Harris SR, Rivera-Hernandez T, Berking O, Cork AJ, Torres RSLALA, Lithgow T, Strugnell RA, Bergmann R, Nitsche-Schmitz P, Chhatwal GS, Bentley SD, Fraser JD, Moreland NJ, Carapetis JR, Steer AC, Parkhill J, Saul A, Williamson DA, Currie BJ, Tong SYCC, Dougan G, Walker MJ**. 2019. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat Genet* **51**:1035–1043.

19. **Wadsworth CB, Arnold BJ, Sater MRA, Grad YH**. 2018. Azithromycin Resistance through Interspecific Acquisition of an Epistasis-Dependent Efflux Pump Component and Transcriptional Regulator in Neisseria gonorrhoeae. *MBio* **9**:e01419-18.

20. **Hicks ND, Yang J, Zhang X, Zhao B, Grad YH, Liu L, Ou X, Chang Z, Xia H, Zhou Y, Wang S, Dong J, Sun L, Zhu Y, Zhao Y, Jin Q, Fortune SM**. 2018. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. *Nat Microbiol* **3**:1032–1042.

21. **Li Y, Metcalf BJ, Chochua S, Li Z, Walker H, Tran T, Hawkins PA, Gierke R,**

Pilishvili T, McGee L, Beall BW. 2019. Genome-wide association analyses of invasive pneumococcal isolates identify a missense bacterial mutation associated with meningitis. *Nat Commun* **10**:178.

22. **McInerney JO, McNally A, O'Connell MJ**. 2017. Why prokaryotes have pangenomes. *Nat Microbiol* **2**:17040.

23. **Snipen L, Ussery DW**. 2010. Standard operating procedure for computing pangenome trees. *Stand Genomic Sci* **2**:135–41.

24. **Koonin E V., Wolf YI**. 2012. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? *Front Cell Infect Microbiol* **2**:119.

25. **Rouli L, Merhej V, Fournier P-E, Raoult D**. 2015. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New microbes new Infect* **7**:72–85.

26. **Brüggemann H, Jensen A, Nazipi S, Aslan H, Meyer RL, Poehlein A, Brzuszkiewicz E, Al-Zeer MA, Brinkmann V, Söderquist B**. 2018. Pan-genome analysis of the genus Finegoldia identifies two distinct clades, strain-specific heterogeneity, and putative virulence factors. *Sci Rep* **8**:266.

27. **Ward D V, Scholz M, Zolfo M, Taft DH, Schibler KR, Tett A, Segata N, Morrow AL**. 2016. Metagenomic Sequencing with Strain-Level Resolution Implicates Uropathogenic E. coli in Necrotizing Enterocolitis and Mortality in Preterm Infants. *Cell Rep* **14**:2912–24.

28. **Scholz M, Ward D V, Pasolli E, Tolio T, Zolfo M, Asnicar F, Truong DT, Tett A, Morrow AL, Segata N**. 2016. Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat Methods* **13**:435–8.

29. **Bartfeld S**. 2016. Modeling infectious diseases and host-microbe interactions in gastrointestinal organoids. *Dev Biol* **420**:262–270.

30. **Co JY, Margalef-Català M, Li X, Mah AT, Kuo CJ, Monack DM, Amieva MR**. 2019. Controlling Epithelial Polarity: A Human Enteroid Model for Host-Pathogen Interactions. *Cell Rep* **26**:2509-2520.e4.

31. **Lai NY, Musser MA, Pinho-Ribeiro FA, Baral P, Jacobson A, Ma P, Potts DE, Chen Z, Paik D, Soualhi S, Yan Y, Misra A, Goldstein K, Lagomarsino VN, Nordstrom A, Sivanathan KN, Wallrapp A, Kuchroo VK, Nowarski R, Starnbach MN, Shi H, Surana NK, An D, Wu C, Huh JR, Rao M, Chiu IM**. 2020. Gut-Innervating Nociceptor Neurons Regulate Peyer's Patch Microfold Cells and SFB Levels to Mediate Salmonella Host Defense. *Cell* **180**:33-49.e22.

32. **Duchet-Suchaux M, Le Maitre C, Bertin A**. 1990. Differences in susceptibility of inbred and outbred infant mice to enterotoxigenic Escherichia coli of bovine, porcine and human origin. *J Med Microbiol* **31**:185–90.

33. **Kamareddine L, Wong ACN, Vanhove AS, Hang S, Purdy AE, Kierek-Pearson K, Asara JM, Ali A, Morris JG, Watnick PI**. 2018. Activation of Vibrio cholerae quorum sensing promotes survival of an arthropod host. *Nat Microbiol* **3**:243–252.

34. **Sit B, Zhang T, Fakoya B, Akter A, Biswas R, Ryan ET, Waldor MK**. 2019. Oral immunization with a probiotic cholera vaccine induces broad protective immunity against Vibrio cholerae colonization and disease in mice. *PLoS Negl Trop Dis* **13**:e0007417.

35. **Ritchie JM, Rui H, Bronson RT, Waldor MK**. 2010. Back to the future: studying cholera pathogenesis using infant rabbits. *MBio* **1**:e00047-10.

36. **Hubbard TP, Billings G, Dörr T, Sit B, Warr AR, Kuehl CJ, Kim M, Delgado F, Mekalanos JJ, Lewnard JA, Waldor MK**. 2018. A live vaccine rapidly protects against cholera in an infant rabbit model. *Sci Transl Med* **10**:1–11.

37. **Shim D-H, Suzuki T, Chang S-Y, Park S-M, Sansonetti PJ, Sasakawa C, Kweon M-N**. 2007. New Animal Model of Shigellosis in the Guinea Pig: Its Usefulness for Protective Efficacy Studies. *J Immunol* **178**:2476–2482.

38. **Arena ET, Campbell-Valois FX, Tinevez JY, Nigro G, Sachse M, Moya-Nilges M, Nothelfer K, Marteyn B, Shorte SL, Sansonetti PJ**. 2015. Bioimage analysis of Shigella infection reveals targeting of colonic crypts. *Proc Natl Acad Sci U S A* **112**:E3282–E3290.

39. **Xu D, Liao C, Zhang B, Tolbert WD, He W, Dai Z, Zhang W, Yuan W, Pazgier M, Liu J, Yu J, Sansonetti PJ, Bevins CL, Shao Y, Lu W**. 2018. Human Enteric α-Defensin 5 Promotes Shigella Infection by Enhancing Bacterial Adhesion and Invasion. *Immunity* **48**:1233-1244.e6.

40. **Douce G, Goulding D**. 2010. Refinement of the hamster model of Clostridium difficile disease. *Methods Mol Biol* **646**:215–27.

41. **Fulde M, Valentin-Weigand P**. 2013. Epidemiology and pathogenicity of zoonotic streptococci. *Curr Top Microbiol Immunol* **368**:49–81.

42. **Grant ST, Efstratiou A, Chanter N**. 1993. Laboratory diagnosis of strangles and the isolation of atypical Streptococcus equi. *Vet Rec* **133**:215–6.

43. **Plowright W, Nakajima H, Keibakai. NC**. 1994. Equine infectious diseases VII: proceedings of the Seventh International Conference, Tokyo, 8th-11th June 1994. R & W Publications, Newmarket.

44. **Chanter N, Collin N, Holmes N, Binns M, Mumford J**. 1997. Characterization of the Lancefield group C streptococcus 16S-23S RNA gene intergenic spacer and its potential for identification and sub-specific typing. *Epidemiol Infect* **118**:125–35.

45. **Holden MTG, Heather Z, Paillot R, Steward KF, Webb K, Ainslie F, Jourdan T, Bason NC, Holroyd NE, Mungall K, Quail MA, Sanders M, Simmonds M, Willey D, Brooks K, Aanensen DM, Spratt BG, Jolley KA, Maiden MCJ, Kehoe M, Chanter N, Bentley SD, Robinson C, Maskell DJ, Parkhill J, Waller AS**. 2009. Genomic Evidence for the Evolution of Streptococcus equi: Host Restriction, Increased Virulence, and Genetic Exchange with Human Pathogens. *PLoS Pathog* **5**:e1000346.

46. **Webb K, Jolley KA, Mitchell Z, Robinson C, Newton JR, Maiden MCJ, Waller A**. 2008. Development of an unambiguous and discriminatory multilocus sequence typing scheme for the Streptococcus zooepidemicus group. *Microbiology* **154**:3016–3024.

47. **Timoney JF**. 2004. The pathogenic equine streptococci. *Vet Res* **35**:397–409.

48. **Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, Li W-J**. 2014. Comparative Genomics of the Bacterial Genus Streptococcus Illuminates Evolutionary Implications of Species Groups. *PLoS One* **9**:e101229.

49. **Ma Z, Geng J, Zhang H, Yu H, Yi L, Lei M, Lu C, Fan H, Hu S**. 2011. Complete genome sequence of Streptococcus equi subsp. zooepidemicus strain ATCC 35246. *J Bacteriol* **193**:5583–4.

50. **Feng Z, Hu J, Feng, Z.G. HJS**. 1977. Outbreak of swine streptococcosis in Sichan province and identification of pathogen. *Anim Husb Vet Med Lett* **2**:7–12.

51. **de Costa MO, Lage B**. 2019. Streptococcus equi subsp. zooepidemicus associated with sudden death of swine in North America. *bioRxiv* 2019.10.25.812636.

52. **Chen X, Resende-De-Macedo N, Sitthicharoenchai P, Sahin O, Burrough E, Clavijo M, Derscheid R, Schwartz K, Lantz K, Robbe-Austerman S, Main R, Li G**. 2019. Genetic characterization of Streptococcus equi subspecies zooepidemicus associated with high swine mortality in United States. *bioRxiv* 2019.12.12.874644.

53. **Harrington DJ, Sutcliffe IC, Chanter N**. 2002. The molecular basis of Streptococcus equi infection and disease. *Microbes Infect* **4**:501–510.

54. **Priestnall S, Erles K**. 2011. Streptococcus zooepidemicus: an emerging canine pathogen. *Vet J* **188**:142–8.

55. **Fan H, Wang Y, Tang F, Lu C**. 2008. Determination of the mimic epitope of the M-like protein adhesin in swine Streptococcus equi subsp. zooepidemicus. *BMC Microbiol* **8**:170.

56. **Chanter N**. 1997. Streptococci and enterococci as animal pathogens. *Soc Appl Bacteriol Symp Ser* **26**:100S-109S.

57. **Velineni S, Timoney JF**. 2013. Identification of novel immunoreactive proteins of streptococcus zooepidemicus with potential as vaccine components. *Vaccine* **31**:4129–4135.

58. **Velineni S, Timoney JF**. 2013. Characterization and protective immunogenicity of the szm protein of Streptococcus zooepidemicus NC78 from a clonal outbreak of equine respiratory disease. *Clin Vaccine Immunol* **20**:1181–1188.

59. **Flock M, Karlström A, Lannergård J, Guss B, Flock J-I**. 2006. Protective effect of vaccination with recombinant proteins from Streptococcus equi subspecies equi in a strangles model in the mouse. *Vaccine* **24**:4144–51.

60. **Blank LM, Hugenholtz P, Nielsen LK**. 2008. Evolution of the hyaluronic acid synthesis (has) operon in Streptococcus zooepidemicus and other pathogenic streptococci. *J Mol Evol* **67**:13–22.

61. **Wibawan IW, Pasaribu FH, Utama IH, Abdulmawjood A, Lämmler C**. 1999. The role of hyaluronic acid capsular material of Streptococcus equi subsp. zooepidemicus in mediating adherence to HeLa cells and in resisting phagocytosis. *Res Vet Sci* **67**:131–5.

62. **Jonsson H, Lindmark H, Guss B**. 1995. A protein G-related cell surface protein in Streptococcus zooepidemicus. *Infect Immun* **63**:2968–75.

63. **Egesten A, Frick I-M, Mörgelin M, Olin AI, Björck L**. 2011. Binding of albumin promotes bacterial survival at the epithelial surface. *J Biol Chem* **286**:2469–76.

64. **Hulting G, Flock M, Frykberg L, Lannergård J, Flock J-I, Guss B**. 2009. Two novel IgG endopeptidases of Streptococcus equi. *FEMS Microbiol Lett* **298**:44–50.

65. **Lannergård J, Guss B**. 2006. IdeE, an IgG-endopeptidase of Streptococcus equi ssp. equi. *FEMS Microbiol Lett* **262**:230–5.

66. **Timoney JF, DeNegri R, Sheoran A, Forster N**. 2010. Affects of N-terminal variation in the SeM protein of Streptococcus equi on antibody and fibrinogen binding. *Vaccine* **28**:1522–1527.

67. **Sheoran AS, Artiushin S, Timoney JF**. 2002. Nasal mucosal immunogenicity for the horse of a SeM peptide of Streptococcus equi genetically coupled to cholera toxin. *Vaccine* **20**:1653–9.

68. **Meehan M, Nowlan P, Owen P**. 1998. Affinity purification and characterization of a fibrinogen-binding protein complex which protects mice against lethal challenge with Streptococcus equi subsp. equi. *Microbiology* **144**:993–1003.

69. **Bergmann R, Jentsch M-C, Uhlig A, Müller U, van der Linden M, Rasmussen M, Waller A, von Köckritz-Blickwede M, Baums CG**. 2019. Prominent binding of human and equine fibrinogen to Streptococcus equi subsp. zooepidemicus is mediated by specific SzM-types and is a distinct phenotype of zoonotic isolates. *Infect Immun* **125**:82–88.

70. **Timoney JF, Artiushin SC, Boschwitz JS**. 1997. Comparison of the sequences and functions of Streptococcus equi M-like proteins SeM and SzPSe. *Infect Immun* **65**:3600–3605.

71. **Kelly C, Bugg M, Robinson C, Mitchell Z, Davis-Poynter N, Newton JR, Jolley KA, Maiden MCJ, Waller AS**. 2006. Sequence variation of the SeM gene of Streptococcus equi allows discrimination of the source of strangles outbreaks. *J Clin Microbiol* **44**:480–486.

72. **Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK, Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD, Walson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC**. 2018. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect Dis* **18**:1229–1240.

73. **Abbasi E, Abtahi H, van Belkum A, Ghaznavi-Rad E**. 2019. Multidrug-resistant Shigella infection in pediatric patients with diarrhea from central Iran. *Infect Drug Resist* **12**:1535–1544.

74. **Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, Sangal V, Brown DJ, Coia JE, Kim DW, Choi SY, Kim SH, Da Silveira WD, Pickard DJ, Farrar JJ,**

Parkhill J, Dougan G, Thomson NR. 2012. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet* **44**:1056–1059.

75. **Puzari M, Sharma M, Chetia P**. 2018. Emergence of antibiotic resistant Shigella species: A matter of concern. *J Infect Public Health* **11**:451–454.

76. **DuPont, Herbert L; Levine Myron M; Hornick Richard B; Formal SB**. 1989. Inoculum Size in Shigellosis and Implications for Expected Mode of Transmission Author (s): Herbert L . DuPont , Myron M . Levine , Richard B . Hornick and Samuel B . Formal Reviewed work (s): Published by : Oxford University Press Stable URL : http:/. *J Infect Dis* **159**:1126–1128.

77. **Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, Antonio M, Hossain A, Mandomando I, Ochieng JB, Oundo JO, Qureshi S, Ramamurthy T, Tamboura B, Adegbola RA, Hossain MJ, Saha D, Sen S, Faruque ASG, Alonso PL, Breiman RF, Zaidi AKM, Sur D, Sow SO, Berkeley LY, O'Reilly CE, Mintz ED, Biswas K, Cohen D, Farag TH, Nasrin D, Wu Y, Blackwelder WC, Kotloff KL, Nataro JP, Levine MM**. 2014. Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin Infect Dis* **59**:933–941.

78. **Thompson CN, Duy PT, Baker S**. 2015. The rising dominance of Shigella sonnei: An intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis* **9**:1–13.

79. **Schroeder GN, Hilbi H**. 2008. Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* **21**:134–56.

80. **Sansonetti PJ, Kopecko DJ, Formal SB**. 1982. Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun* **35**:852–60.

81. **Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Collins HH, Formal SB**. 1983. Alterations in the pathogenicity of Escherichia coli K-12 after transfer of plasmid and chromosomal genes from Shigella flexneri. *Infect Immun* **39**:1392–402.

82. **Kotloff KL, Riddle MS, Platts-Mills JA, Pavlinac P, Zaidi AKM**. 2018. Shigellosis. *Lancet* **391**:801–812.

83. **Shiga K**. 1989. Ueber den Erreger der Dysenterie in Japan. *Zentralbl Bakteriol Microbiol Hyg* **23**:599–600.

84. **Trofa AF, Ueno-Olsen H, Oiwa R, Yoshikawa M**. 1999. Dr. Kiyoshi Shiga: Discoverer of the Dysentery Bacillus. *Clin Infect Dis* **29**:1303–1306.

85. **Gu B, Cao Y, Pan S, Zhuang L, Yu R, Peng Z, Qian H, Wei Y, Zhao L, Liu G, Tong M**. 2012. Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents* **40**:9–17.

86. **Dodd CE, Jones D**. 1982. A numerical taxonomic study of the genus Shigella. *J Gen Microbiol* **128**:1933–57.

87. **Pupo GM, Lan R, Reeves PR**. 2000. Multiple independent origins of Shigella clones of

Escherichia coli and convergent evolution of many of their characteristics. *Proc Natl Acad Sci U S A* **97**:10567–72.

88. **Lan R, Reeves PR**. 2002. Escherichia coli in disguise: molecular origins of Shigella. *Microbes Infect* **4**:1125–32.

89. **Sack DA, Hoque AT, Huq A, Etheridge M**. 1994. Is protection against shigellosis induced by natural infection with Plesiomonas shigelloides? *Lancet (London, England)* **343**:1413–5.

90. **Shepherd JG, Wang L, Reeves PR**. 2000. Comparison of O-antigen gene clusters of Escherichia coli (Shigella) sonnei and Plesiomonas shigelloides O17: sonnei gained its current plasmid-borne O-antigen genes from P. shigelloides in a recent event. *Infect Immun* **68**:6056–61.

91. **Sobel J, Cameron DN, Ismail J, Strockbine N, Williams M, Diaz PS, Westley B, Rittmann M, DiCristina J, Ragazzoni H, Tauxe R V, Mintz ED**. 1998. A prolonged outbreak of Shigella sonnei infections in traditionally observant Jewish communities in North America caused by a molecularly distinct bacterial subtype. *J Infect Dis* **177**:1405–9.

92. **Wu HH, Shen YT, Chiou CS, Fang CT, Lo YC**. 2019. Shigellosis outbreak among MSM living with HIV: A case-control study in Taiwan, 2015-2016. *Sex Transm Infect* **95**:67–70.

93. **Serafino Wani RL, Filson SA, Chattaway MA, Godbole G**. 2016. Invasive shigellosis in MSM. *Int J STD AIDS* **27**:917–919.

94. **Agha R, Goldberg MB**. 2018. Shigella infection: Clinical manifestations and diagnosis, p. . *In* Post, TW (ed.), UpToDate. UpToDate, Waltham, MA.

95. **Avital A, Maayan C, Goitein KJ**. 1982. Incidence of convulsions and encephalopathy in childhood Shigella infections. Survey of 117 hospitalized patients. *Clin Pediatr (Phila)* **21**:645–8.

96. **Dodd K, Buddingh J, Rapoport S**. 1949. The etiology of ekiri, a highly fatal disease of Japanese children. *Pediatrics* **3**:9–19.

97. **Bennish ML**. 1991. Potentially Lethal Complications of Shigellosis. *Clin Infect Dis* **13**:S319–S324.

98. **Anand BS, Malhotra V, Bhattacharya SK, Datta P, Datta D, Sen D, Bhattacharya MK, Mukherjee PP, Pal SC**. 1986. Rectal histology in acute bacillary dysentery. *Gastroenterology* **90**:654–60.

99. **Mathan MM, Mathan VI**. 1991. Morphology of rectal mucosa of patients with shigellosis. *Rev Infect Dis* **13 Suppl 4**:S314-8.

100. **Sansonetti PJ, Ryter A, Clerc P, Maurelli AT, Mounier J**. 1986. Multiplication of Shigella flexneri within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. *Infect Immun* **51**:461–9.

101. **High N, Mounier J, Prévost MC, Sansonetti PJ**. 1992. IpaB of Shigella flexneri causes entry into epithelial cells and escape from the phagocytic vacuole. *EMBO J* **11**:1991–9.

102. **Osiecki JC, Barker J, Picking WL, Serfis AB, Berring E, Shah S, Harrington A, Picking WD**. 2001. IpaC from Shigella and SipC from Salmonella possess similar biochemical properties but are functionally distinct. *Mol Microbiol* **42**:469–81.

103. **Du J, Reeves AZ, Klein JA, Twedt DJ, Knodler LA, Lesser CF**. 2016. The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *Proc Natl Acad Sci U S A* **113**:4794–9.

104. **Suzuki T, Mimuro H, Miki H, Takenawa T, Sasaki T, Nakanishi H, Takai Y, Sasakawa C**. 2000. Rho family GTPase Cdc42 is essential for the actin-based motility of Shigella in mammalian cells. *J Exp Med* **191**:1905–20.

105. **Handa Y, Suzuki M, Ohya K, Iwai H, Ishijima N, Koleske AJ, Fukui Y, Sasakawa C**. 2007. Shigella IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery. *Nat Cell Biol* **9**:121–8.

106. **Sansonetti PJ, Mounier J, Prévost MC, Mège RM**. 1994. Cadherin expression is required for the spread of Shigella flexneri between epithelial cells. *Cell* **76**:829–39.

107. **Russo BC, Stamm LM, Raaben M, Kim CM, Kahoud E, Robinson LR, Bose S, Queiroz AL, Herrera BB, Baxt LA, Mor-Vaknin N, Fu Y, Molina G, Markovitz DM, Whelan SP, Goldberg MB**. 2016. Intermediate filaments enable pathogen docking to trigger type 3 effector translocation. *Nat Microbiol* **1**:16025.

108. **Zychlinsky A, Prevost MC, Sansonetti PJ**. 1992. Shigella flexneri induces apoptosis in infected macrophages. *Nature* **358**:167–9.

109. **Pinaud L, Samassa F, Porat Z, Ferrari ML, Belotserkovsky I, Parsot C, Sansonetti PJ, Campbell-Valois F-X, Phalipon A**. 2017. Injection of T3SS effectors not resulting in invasion is the main targeting mechanism of Shigella toward human lymphocytes. *Proc Natl Acad Sci U S A* **114**:9954–9959.

110. **Mandic-Mulec I, Weiss J, Zychlinsky A**. 1997. Shigella flexneri is trapped in polymorphonuclear leukocyte vacuoles and efficiently killed. *Infect Immun* **65**:110–5.

111. **Weinrauch Y, Drujan D, Shapiro SD, Weiss J, Zychlinsky A**. 2002. Neutrophil elastase targets virulence factors of enterobacteria. *Nature* **417**:91–4.

112. **Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H, Hattori M, Iida T**. 2003. Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism distinct from that of V cholerae. *Lancet (London, England)* **361**:743–9.

113. **Wagner S, Grin I, Malmsheimer S, Singh N, Torres-Vargas CE, Westerhausen S**. 2018. Bacterial type III secretion systems: a complex device for the delivery of bacterial effector proteins into eukaryotic host cells. *FEMS Microbiol Lett* **365**:1–13.

114. **Deng W, Marshall NC, Rowland JL, McCoy JM, Worrall LJ, Santos AS, Strynadka**

NCJ, Finlay BB. 2017. Assembly, structure, function and regulation of type III secretion systems. *Nat Rev Microbiol* **15**:323–337.

115. **Ashida H, Mimuro H, Sasakawa C**. 2015. Shigella manipulates host immune responses by delivering effector proteins with specific roles. *Front Immunol* **6**:219.

116. **Kuehl CJ, Dragoi A-M, Agaisse H**. 2014. The Shigella flexneri type 3 secretion system is required for tyrosine kinase-dependent protrusion resolution, and vacuole escape during bacterial dissemination. *PLoS One* **9**:e112738.

117. **Maurelli AT, Sansonetti PJ**. 1988. Identification of a chromosomal gene controlling temperature-regulated expression of Shigella virulence. *Proc Natl Acad Sci U S A* **85**:2820–4.

118. **Li P, Jiang W, Yu Q, Liu W, Zhou P, Li J, Xu J, Xu B, Wang F, Shao F**. 2017. Ubiquitination and degradation of GBPs by a Shigella effector to suppress host defence. *Nature* **551**:378–383.

119. **Li H, Xu H, Zhou Y, Zhang J, Long C, Li S, Chen S, Zhou J-M, Shao F**. 2007. The phosphothreonine lyase activity of a bacterial type III effector family. *Science* **315**:1000–3.

120. **Goldberg MB, Theriot JA**. 1995. Shigella flexneri surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci U S A* **92**:6572–6.

121. **Suzuki T, Lett MC, Sasakawa C**. 1995. Extracellular transport of VirG protein in Shigella. *J Biol Chem* **270**:30874–80.

122. **Henderson IR, Nataro JP, Kaper JB, Meyer TF, Farrand SK, Burns DL, Finlay BB, St Geme JW**. 2000. Renaming protein secretion in the gram-negative bacteria. *Trends Microbiol* **8**:352.

123. **Makino S, Sasakawa C, Kamata K, Kurata T, Yoshikawa M**. 1986. A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in S. flexneri 2a. *Cell* **46**:551–5.

124. **Bernardini ML, Mounier J, D'Hauteville H, Coquis-Rondon M, Sansonetti PJ**. 1989. Identification of icsA, a plasmid locus of Shigella flexneri that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proc Natl Acad Sci U S A* **86**:3867–3871.

125. **Brotcke Zumsteg A, Goosmann C, Brinkmann V, Morona R, Zychlinsky A**. 2014. IcsA is a Shigella flexneri adhesion regulated by the type III secretion system and required for pathogenesis. *Cell Host Microbe* **15**:435–445.

126. **Barry EM, Pasetti MF, Sztein MB, Fasano A, Kotloff KL, Levine MM**. 2013. Progress and pitfalls in Shigella vaccine research. *Nat Rev Gastroenterol Hepatol* **10**:245–255.

127. **Okada N, Sasakawa C, Tobe T, Yamada M, Nagai S, Talukder KA, Komatsu K, Kanegasaki S, Yoshikawa M**. 1991. Virulence-associated chromosomal loci of Shigella flexneri identified by random Tn5 insertion mutagenesis. *Mol Microbiol* **5**:187–95.

128. **Okamura N, Nagai T, Nakaya R, Kondo S, Murakami M, Hisatsune K**. 1983. HeLa cell invasiveness and O antigen of Shigella flexneri as separate and prerequisite attributes of virulence to evoke keratoconjunctivitis in guinea pigs. *Infect Immun* **39**:505–13.

129. **Schnupf P, Sansonetti PJ**. 2019. Shigella Pathogenesis: New Insights through Advanced Methodologies. *Microbiol Spectr* **7**:BAI-0023-2019.

130. **Faherty C, Harper JM, Shea-Donohue T, Barry EM, Kaper JB, Fasano A, Nataro JP**. 2012. Chromosomal and Plasmid-Encoded Factors of Shigella flexneri Induce Secretogenic Activity Ex Vivo. *PLoS One* **7**:e49980.

131. **Vargas M, Gascon J, Jimenez De Anta MT, Vila J**. 1999. Prevalence of Shigella enterotoxins 1 and 2 among Shigella strains isolated from patients with traveler's diarrhea. *J Clin Microbiol* **37**:3608–3611.

132. **Lawlor KM, Payne SM**. 1984. Aerobactin genes in Shigella spp. *J Bacteriol* **160**:266–72.

133. **Al-Hasani K, Navarro-Garcia F, Huerta J, Sakellaris H, Adler B**. 2009. The immunogenic SigA enterotoxin of Shigella flexneri 2a binds to HEp-2 cells and induces fodrin redistribution in intoxicated epithelial cells. *PLoS One* **4**:e8223.

134. **Ruiz-Perez F, Wahid R, Faherty CS, Kolappaswamy K, Rodriguez L, Santiago A, Murphy E, Cross A, Sztein MB, Nataro JP**. 2011. Serine protease autotransporters from Shigella flexneri and pathogenic Escherichia coli target a broad range of leukocyte glycoproteins. *Proc Natl Acad Sci U S A* **108**:12881–6.

135. **Yum LK, Byndloss MX, Feldman SH, Agaisse H**. 2019. Critical role of bacterial dissemination in an infant rabbit model of bacillary dysentery. *Nat Commun* **10**:1–10.

136. **Sereny B**. 1955. Experimental shigella keratoconjunctivitis; a preliminary report. *Acta Microbiol Acad Sci Hung* **2**:293–6.

137. **D'Hauteville H, Khan S, Maskell DJ, Kussak A, Weintraub A, Mathison J, Ulevitch RJ, Wuscher N, Parsot C, Sansonetti PJ**. 2002. Two msbB genes encoding maximal acylation of lipid A are required for invasive Shigella flexneri to mediate inflammatory rupture and destruction of the intestinal epithelium. *J Immunol* **168**:5240–51.

138. **Formal SB, Dammin GJ, Labrec EH, Schneider H**. 1958. Experimental Shigella infections: characteristics of a fatal infection produced in guinea pigs. *J Bacteriol* **75**:604–10.

139. **Formal SB, Dammin GJ, Schneider H, Labrec EH**. 1959. Experimental Shigella infections. II. Characteristics of a fatal enteric infection in guinea pigs following the subcutaneous inoculation of carbon tetrachloride. *J Bacteriol* **78**:800–4.

140. **West NP**. 2005. Optimization of Virulence Functions Through Glucosylation of Shigella LPS. *Science (80-)* **307**:1313–1317.

141. **Gomez HF, Ochoa TJ, Herrera-Insua I, Carlin LG, Cleary TG**. 2002. Lactoferrin protects rabbits from Shigella flexneri-induced inflammatory enteritis. *Infect Immun* **70**:7050–3.

142. **Maier BR, Hentges DJ**. 1972. Experimental Shigella Infections in Laboratory Animals. *Infect Immun* **6**:168–173.

143. **Singer M, Sansonetti PJ**. 2004. IL-8 Is a Key Chemokine Regulating Neutrophil Recruitment in a New Mouse Model of Shigella- Induced Colitis . *J Immunol* **173**:4197–4206.

144. **Zhang Z, Jin L, Champion G, Seydel KB, Stanley J**. 2001. Shigella infection in a SCID mouse-human intestinal xenograft model: Role for neutrophils in containing bacterial dissemination in human intestine. *Infect Immun* **69**:3240–3247.

145. **Mostowy S, Boucontet L, Mazon Moya MJ, Sirianni A, Boudinot P, Hollinshead M, Cossart P, Herbomel P, Levraud J-P, Colucci-Guyon E**. 2013. The zebrafish as a new model for the in vivo study of Shigella flexneri interaction with phagocytes and bacterial autophagy. *PLoS Pathog* **9**:e1003588.

146. **Duggan GM, Mostowy S**. 2018. Use of zebrafish to study Shigella infection. *Dis Model Mech* **11**:1–11.

147. **Willis AR, Torraca V, Gomes MC, Shelley J, Mazon-Moya M, Filloux A, Lo Celso C, Mostowy S**. 2018. Shigella -Induced Emergency Granulopoiesis Protects Zebrafish Larvae from Secondary Infection. *MBio* **9**:e00933-18.

148. **Mazon-Moya MJ, Willis AR, Torraca V, Boucontet L, Shenoy AR, Colucci-Guyon E, Mostowy S**. 2017. Septins restrict inflammation and protect zebrafish larvae from Shigella infection. *PLoS Pathog* **13**:e1006467.

149. **Burton EA, Pendergast AM, Aballay A**. 2006. The Caenorhabditis elegans ABL-1 tyrosine kinase is required for Shigella flexneri pathogenesis. *Appl Environ Microbiol* **72**:5043–5051.

150. **George DT, Behm CA, Hall DH, Mathesius U, Rug M, Nguyen KCQ, Verma NK**. 2014. Shigella flexneri Infection in Caenorhabditis elegans: Cytopathological Examination and Identification of Host Responses. *PLoS One* **9**:e106085.

151. **Campbell-Valois FO-X, Schnupf P, Nigro G, Sachse M, Sansonetti PJ, Parsot C**. 2014. A Fluorescent Reporter Reveals On/Off Regulation of the Shigella Type III Secretion Apparatus during Entry and Cell-to-Cell Spread. *Cell Host Microbe* **15**:177–189.

152. **Tinevez J-Y, Arena ET, Anderson M, Nigro G, Injarabian L, André A, Ferrari M, Campbell-Valois F-X, Devin A, Shorte SL, Sansonetti PJ, Marteyn BS**. 2019. Shigella-mediated oxygen depletion is essential for intestinal mucosa colonization. *Nat Microbiol* **4**:2001–2009.

153. **Mallett CP, VanDeVerg L, Collins HH, Hale TL**. 1993. Evaluation of Shigella vaccine safety and efficacy in an intranasally challenged mouse model. *Vaccine* **11**:190–196.

154. **Van de Verg LL, Mallett CP, Collins HH, Larsen T, Hammack C, Hale TL**. 1995. Antibody and cytokine responses in a mouse pulmonary model of Shigella flexneri serotype 2a infection. *Infect Immun* **63**:1947–1954.

155. **Way SS, Borczuk AC, Dominitz R, Goldberg MB**. 1998. An Essential Role for Gamma Interferon in Innate Resistance to Shigella flexneri Infection. *Infect Immun* **66**:1342–1348.

156. **Yang JY, Lee SN, Chang SY, Ko HJ, Ryu S, Kweon MN**. 2014. A mouse model of shigellosis by intraperitoneal infection. *J Infect Dis* **209**:203–215.

157. **Coster TS, Hoge CW, VanDeVerg LL, Hartman AB, Oaks E V., Venkatesan MM, Cohen D, Robin G, Fontaine-Thompson A, Sansonetti PJ, Hale TL**. 1999. Vaccination against shigellosis with attenuated Shigella flexneri 2a strain SC602. *Infect Immun* **67**:3437–43.

158. **Porter CK, Lynen A, Riddle MS, Talaat K, Sack D, Gutiérrez RL, McKenzie R, DeNearing B, Feijoo B, Kaminski RW, Taylor DN, Kirkpatrick BD, Bourgeois AL**. 2018. Clinical endpoints in the controlled human challenge model for Shigella: A call for standardization and the development of a disease severity score. *PLoS One* **13**:e0194325.

159. **Kotloff KL, Taylor DN, Sztein MB, Wasserman SS, Losonsky GA, Nataro JP, Venkatesan M, Hartman A, Picking WD, Katz DE, Campbell JD, Levine MM, Hale TL**. 2002. Phase I evaluation of ΔvirG Shigella sonnei live, attenuated, oral vaccine strain WRSS1 in healthy adults. *Infect Immun* **70**:2016–2021.

160. **Lee J-I, Kim S-J, Park C-G**. 2011. Shigella flexneri infection in a newly acquired rhesus macaque (Macaca mulatta). *Lab Anim Res* **bosch**:343.

161. **Enurah LU, Uche EM, Nawathe DR**. 1988. Fatal shigellosis in a chimpanzee (Pan troglodytes) in the Jos Zoo, Nigeria. *J Wildl Dis* **24**:178–9.

162. **Higgins R, Sauvageau R, Bonin P**. 1985. Shigella flexneri Type 2 Infection in Captive Nonhuman Primates. *Can Vet J = La Rev Vet Can* **26**:402–3.

Chapter 2

A conserved streptococcal virulence regulator controls expression of a distinct class of M-like proteins

Adapted from a manuscript in mBio (October 22, 2019)

A conserved streptococcal virulence regulator controls expression of a distinct class of M-like proteins

Jonathan D. D'Gama[*1,2], Zhe Ma[*#1,2,3,4,5], Hailong Zhang[1,2], Xu Liu[1,2], Hongjie Fan[3,4,5], Ellen Ruth A. Morris[6], Noah D. Cohen[6], Colette Cywes-Bentley[1,2], Gerald B. Pier[1,2], Matthew K. Waldor[1,2,7]

Affiliations

1. Department of Microbiology, Harvard Medical School, Boston, Massachusetts, USA

2. Division of Infectious Diseases, Brigham & Women's Hospital, Boston, Massachusetts, USA

3. MOE Joint International Research Laboratory of Animal Health and Food Safety, College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, Jiangsu, China

4. Ministry of Agriculture Key Laboratory of Animal Bacteriology, Nanjing, Jiangsu, China

5. Jiangsu Co-Innovation Center for Prevention and Control of Important Animal Infectious Diseases and Zoonoses, Yangzhou, Jiangsu, China

6. Department of Large Animal Clinical Sciences, College of Veterinary Medicine & Biomedical Sciences, College Station, Texas, USA

7. Howard Hughes Medical Institute, Boston, Massachusetts, USA

[*]Contributed equally

[#]For correspondence: mazhe@njau.edu.cn

Author Contributions

This chapter was adapted from a publication in *mBio* (October 2019). Jonathan D'Gama contributed to designing the study, performing experiments, analyzing the data, and wrote the manuscript. Zhe Ma contributed to designing the study, performing experiments, analyzing the data, and edited the manuscript. Hailong Zhang and Xu Liu helped with experiments. Ellen Ruth Morris and Noah Cohen contributed assembled whole genome sequences of novel streptococcal strains that we analyzed in the study. Hongjie Fan edited the manuscript. Colette Cywes-Bentley contributed to experiments. Gerald Pier contributed to designing the study and edited the manuscript. Matthew Waldor contributed to designing the study and edited the manuscript. Matthew Waldor and Zhe Ma supervised the study.

Abstract

Streptococcus equi subspecies *zooepidemicus* (SEZ) are group C streptococci that are important pathogens of economically valuable animals such as horses and pigs. Here, we found that many SEZ isolates bind to a monoclonal antibody that recognizes poly-*N*-acetyl-glucosamine (PNAG), a polymer that is found as a surface capsule-like structure on diverse microbes. A FACS-based transposon insertion sequencing (Tn-seq) screen coupled with whole genome sequencing was used to search for genes for PNAG biosynthesis. Surprisingly, mutations in a gene encoding an M-like protein, *szM*, and the adjacent transcription factor, designated *sezV*, rendered strains PNAG-negative. SezV was required for *szM* expression and transcriptome analysis showed that SezV has a small regulon. SEZ strains with inactivating mutations in either *sezV* or *szM* were highly attenuated in a mouse model of infection. Comparative genomic analyses revealed that linked *sezV* and *szM* homologues are present in all SEZ, *S. equi subspecies equi* (SEE) and M18 group A streptococcal genomes in the database, but not in other streptococci. The antibody to PNAG bound to a wide-range of SEZ, SEE and M18 GAS strains. Immunochemical studies suggest that the SzM protein may be decorated with a PNAG-like oligosaccharide although an intact oligosaccharide substituent could not be isolated. Collectively, our findings suggest that the *szM*, *sezV* locus defines a subtype of virulent streptococci and that an antibody to PNAG may have therapeutic applications in animal and human diseases caused by streptococci bearing SzM-like proteins.

Significance

M proteins are surface-anchored virulence factors in group A streptococci, human pathogens. Here we identified an M-like protein, SzM, and its positive regulator, SezV, in *Streptococcus equi* subspecies *zooepidemicus* (SEZ), an important group of pathogens for domesticated animals, including horses and pigs. SzM and SezV homologues were found in the genomes of all SEZ and *S. equi subspecies equi* and M18 GAS strains analyzed, but not in other streptococci. Mutant SEZ strains lacking either *sezV* or *szM* were highly attenuated in a mouse model of infection. Collectively, our findings suggest that SezV-related regulators and the linked SzM family of M-like proteins define a new subset of virulent streptococci.

Introduction

Streptococcus equi subspecies *zooepidemicus* (SEZ) is a Lancefield group C *Streptococcus* (GCS) that is a constituent of the normal upper respiratory tract flora of domesticated animals such as horses (1). SEZ is also an opportunistic pathogen, and infections in a wide range of animals including pigs, cows, dogs, and horses, have been reported (1, 2). Epizootic outbreaks in livestock (e.g. swine in China) can be widespread and severe, resulting in significant economic loss (3, 4). Infection of horses with another *S. equi* biovar, *S. equi* subspecies *equi* (SEE), causes the severe, contagious respiratory infection known as strangles (5). SEZ infection of humans, who typically acquire the pathogen through contact with infected animals or contaminated milk or cheese, typically causes meningitis and can be fatal (6, 7). Several virulence factors and protective antigens have been identified in SEZ/SEE, including two M-like proteins, referred to as SzM/SeM and SzP/SzPSe (8–10).

The M/M-like protein family is a class of surface-associated streptococcal proteins that includes M protein, a classical virulence factor of group A *Streptococcus* (GAS, *S. pyogenes*) (11). Proteins in the group are anchored to the cell-wall, share several features including a primarily alpha-helical/coiled-coil structure, polar cytoplasmic C-terminal tail, extracellular distal N-terminus, and function in modulating the host immune response to the pathogen. The diverse immunomodulatory functions of M/M-like proteins include but are not limited to fibrinogen binding, inhibition of complement activation, anti-phagocytic activity, and binding to Fc portions of immunoglobulins (11, 12). M/M-like proteins are typically key streptococcal virulence factors that can be targets of protective antibody responses (11–13). The flexible set of criteria for M/M-like protein classification has led to ambiguity in the field since various types of M-like proteins have different designations, functions, and sequences (11).

The SzM M-like protein in SEZ has been shown to bind fibrinogen, activate plasminogen, inhibit phagocytosis, and to serve as a protective antigen for vaccination (9, 14). However, the contribution of SzM to SEZ virulence in vivo has not been determined. SzP, a second SEZ M-like protein, has been shown to contribute to SEZ virulence in an animal model (8, 15). A SzM ortholog in SEE, SeM (also called FgBP (16)), has also been shown to be an important virulence factor when evaluated in mice (16, 17). Dale et al. (18) noted that certain strains of GAS, particularly of the M18 type, encode a SzM ortholog, called Spa (streptococcal protective antigen), which is linked to virulence (19). In the M18 strain studied, a Spa deletion mutant was more attenuated in mice than a mutant deleted for the canonical M18 protein and Spa was a protective antigen as well (19). Although homologs of SzM have been found in several SEZ, SEE, and GAS strains, the distribution and extent of conservation of SzM homologs in strains of these species and in other streptococcal species is unknown. Furthermore, factors governing the expression of SzM or its homologs have not been described. In contrast, the regulation of M protein and two classes of M-like proteins (M related protein [Mrp] and M-like protein [Enn]) in GAS by the central virulence regulator, Mga (multiple gene activator) is well-described (11, 20, 21).

In addition to M/M-like proteins, capsular polysaccharides and other surface polysaccharides are also important streptococcal virulence factors and protective antigens (22–26). For example, the group A *Streptococcus* carbohydrate is critical for GAS virulence (22, 23). Poly-*N*-acetyl-D-glucosamine (PNAG) is a surface polysaccharide found on numerous Gram-positive and -negative bacteria and fungi (27, 28), that can serve as a virulence factor and a protective antigen (27, 28). Surface PNAG has been identified in *S. pneumoniae*, *S. pyogenes*, and *S. dysgalactiae* (27), but the loci encoding PNAG biosynthesis have not been described in

streptococci. Also, the chemical bases of the association of PNAG with the microbial surface have not been elucidated and it is unknown whether PNAG is physically linked to the cell membrane or cell wall.

Here, we found that SEZ is bound by a monoclonal antibody to PNAG and performed a FACS-based transposon-insertion sequencing (Tn-seq) screen to identify genes required for surface production of PNAG by SEZ. Unexpectedly, the screen identified SzM and a conserved adjacent locus (termed here SezV, SEZ Virulence) that promotes *szM* expression. An anti-PNAG monoclonal antibody was found to directly interact with SzM, suggesting that this M-like protein may be decorated with a PNAG-like glycan. SEZ mutants lacking either *sezV* or *szM* were highly attenuated in a mouse model of infection. Comparative genomics revealed that *szM, seM,* and/or *spa* (here referred to as *szM/seM/spa*) is linked to *sezV* in all SEZ and SEE strains in the database as well as in all M18 GAS, but not in other streptococci. Thus, our findings suggest that SezV-related regulators and the linked SzM family of M-like proteins define a new subset of virulent streptococci.

Results

Tn-seq screen for loci required for surface PNAG expression

We observed that a porcine SEZ isolate (ATCC 35246) (referred to as SEZ) was bound by a human anti-PNAG monoclonal antibody (MAb F598) (Figure 2.1A) and we set out to identify loci that contribute to PNAG synthesis in SEZ. First, we developed a flow cytometry-based system for high-throughput analysis of the PNAG phenotype of single bacteria, using fluorescently conjugated F598 (AF488-F598). There was a considerable range of fluorescence

Figure 2.1. Tn-seq and WGS screens for identification of loci required for surface PNAG expression in SEZ.

A. (Left) Flow cytometry of SEZ stained with an AF488 conjugated monoclonal antibody to alginate (F429, negative control) or PNAG (F598). (Right) Microscopy of SEZ stained with DAPI and either AF488-F429 or AF488-F598. The white scale bar in micrographs represents 10 μm.

B. Schematic for Tn-seq screen for genes required for expression of surface PNAG.

C. AF488-F598 (anti-PNAG) binding to SEZ Tn library at different stages of the Tn-seq screen. Flow cytometry data are taken from screen B. Round 1 = after first round of selection, Round 3 = after third round of selection. Dotted lines are included to help guide the eye.

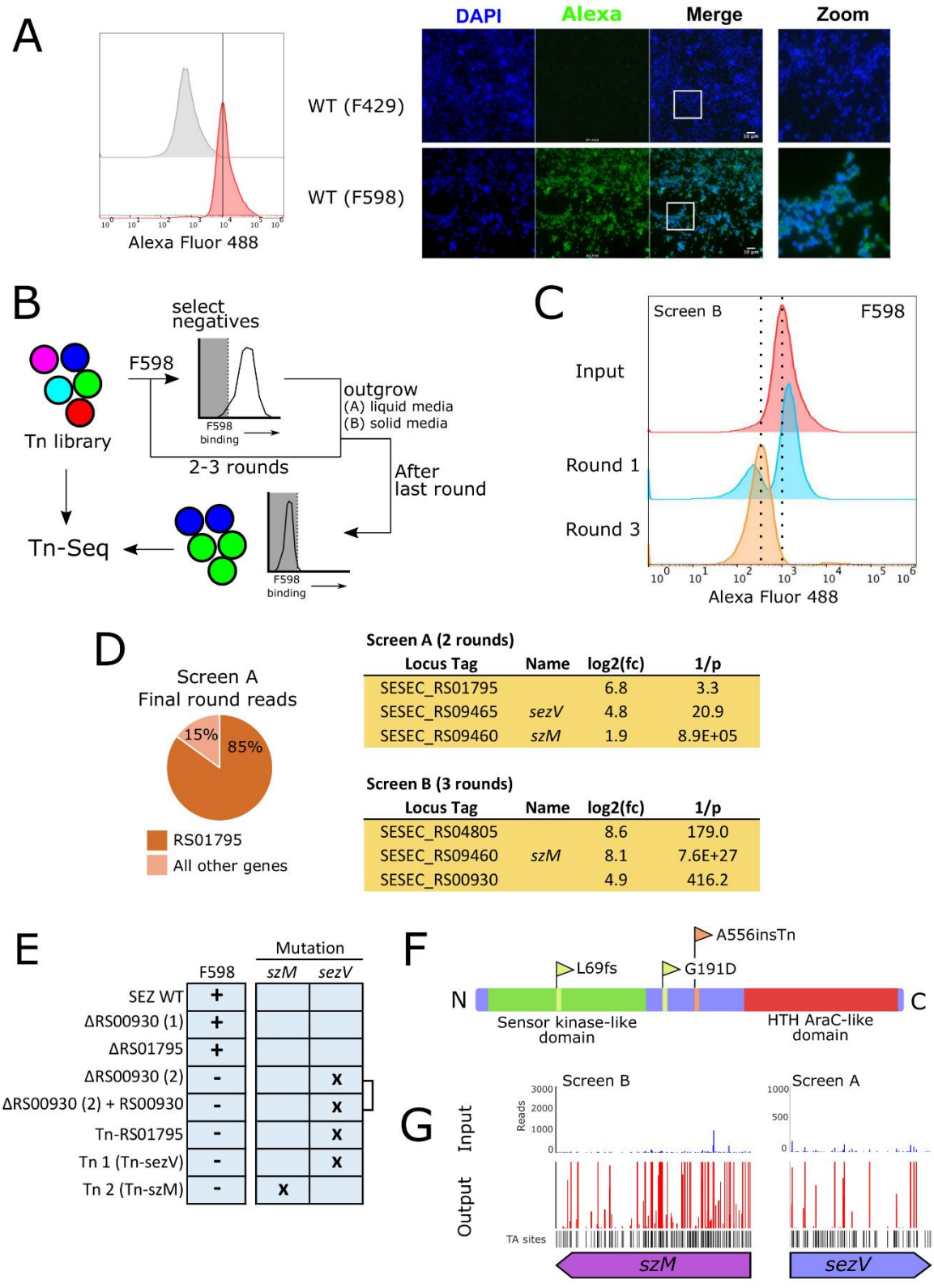
D. (Left) Pie chart of the percentage of all reads mapping to a single site in gene RS01795 in the last round of screen A. (Right) Table of selected potential candidate genes required for PNAG expression from Tn-seq screen.

E. AF488-F598 binding phenotype and gene mutated based on WGS of indicated strains. Bracket indicates that strain ΔRS00930 + RS00930 (2), which was derived from strain ΔRS00930 (2), has the exact same mutation in *sezV* as is found in ΔRS00930 (2). Tn1 contained a transposon insertion in *sezV*; Tn2 contained a transposon insertion in *szM*.

F. Location of mutants and domain organization of *sezV*. HTH = helix-turn-helix. Domains in SezV were predicted using HHphred (29, 30) and Phyre2 (31).

G. Comparison of the read count distributions of transposon insertions in *szM* and *sezV* in the input library and in the last round of the screen reveals uniform enrichment of insertions across these genes.

Figure 2.1 (Continued)



intensities among a population of AF598 bound stationary phase cells, but this range had minimal overlap with that of cells bound by an AF488-labeled isotype control antibody (F429) (Figure 2.1A). Notably, although F429 was raised against an unrelated carbohydrate, it was engineered to include IgG1 heavy and light chain constant region amino acid sequences identical to those of F598 (32). Fluorescence microscopy using these two labeled antibodies confirmed that F598, but not F429 bound to SEZ cells (Figure 2.1A), consistent with the idea that SEZ expresses surface PNAG.

To generate a genetically heterogeneous population of SEZ containing easily mappable mutations, we developed a method to generate a complex transposon-insertion (Tn) library in SEZ. Using the recently developed pMar4s mariner-based transposon delivery vector (33), a library containing transposons in >50% of potential insertion sites was created (Figure 2.2AB). A fluorescence-activated cell sorting (FACS) based screen of the Tn library with several rounds of selection for bacteria with low fluorescent intensity after staining with AF488-F598 was carried out (Figure 2.1BC). We performed the screen with outgrowth in either liquid (screen A) or solid (screen B) media; more modest gating thresholds were applied in the latter, with the aim of identifying genes with weaker phenotypes. Despite the low (background) fluorescence intensity of all bacterial cells after the final rounds of selection (Figure 2.1C), Tn-seq of the selected libraries revealed few genes (screen A: 2, screen B: 16; with average TA sites hit in the input $\geq$ 5, log 2 fold change [L2FC] $\geq$ 2, p-value $\leq$ 0.05) with statistically significant increased ratios of insertions (fold change values) in the sorted versus the input population (Figure 2.1D). Notably, after the final round of selection in screen A there was marked enrichment of only a single Tn insertion (in RS01795, Figure 2.1D), suggesting that a jackpot event occurred at some point in the screen. Single gene deletions in two of the top hits (RS01795, WP_014622330.1, L2FC =

Figure 2.2. Transposon library construction in SEZ and binding of anti-PNAG antibody to mutants identified in Tn-seq screen.

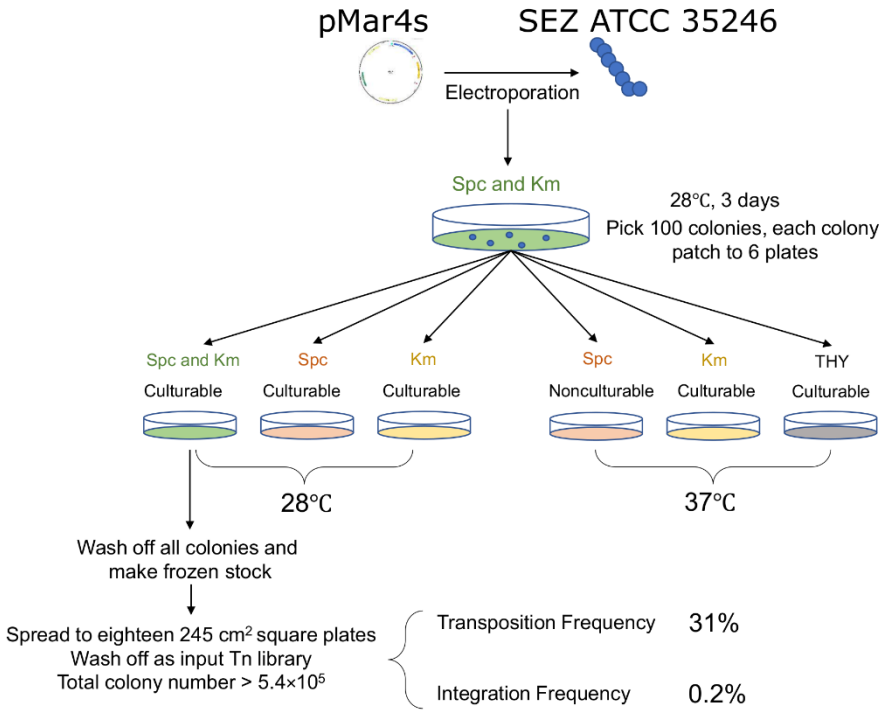
A. Schematic of construction of transposon insertion library in *S. equi* subspecies *zooepidemicus* strain ATCC 35246. pMar4s (33) is a transposon delivery vector that contains the Himar1 mariner C9 transposase, a transposon with a Kanamycin resistance cassette, and an additional Spectinomycin resistance cassette. pMar4s has a temperature sensitive origin of replication in SEZ, and will only be propagated at the permissive (cooler) temperature. pMar4s is introduced into WT SEZ via electroporation, and transformants are selected by growing on THY + Spc + Km for 3 days at the permissive temperature of 28°C. 100 colonies are then patched onto THY plates with antibiotics and temperature as indicated to identify colonies in which the transposon vector has not integrated into the genome and verify that there was still a replicating pMar4s plasmid. From the resistance profiles of the 100 colonies, the transposition and integration frequency can be estimated. As indicated, colonies with the desired resistances were picked from a second THY + Spc + Km plate grown at 28°C and frozen stocks were made from these colonies. The transposon library was generated by thawing a frozen stock and spreading it over several large square THY + Km plates that were subsequently grown at 37°C, yielding ~5.4 x 10^5 colonies. The colonies were scraped off the plate with THY media, and after adding glycerol were stored as frozen aliquots of the transposon library.

B. Genome and SEZ transposon library statistics. Insertions were identified in 75,610 (52%) of the 146,048 potential Tn insertion sites (i.e. TA dinucleotides).

C. Flow cytometry of indicated strains labeled with AF488-F598. ΔRS00930 (1) and ΔRS00930 (2) are independently derived strains containing deletions of RS00930; ΔRS00930 (2) + RS00930 is the deletion mutant complemented with RS00930.

Figure 2.2 (Continued)

A



B

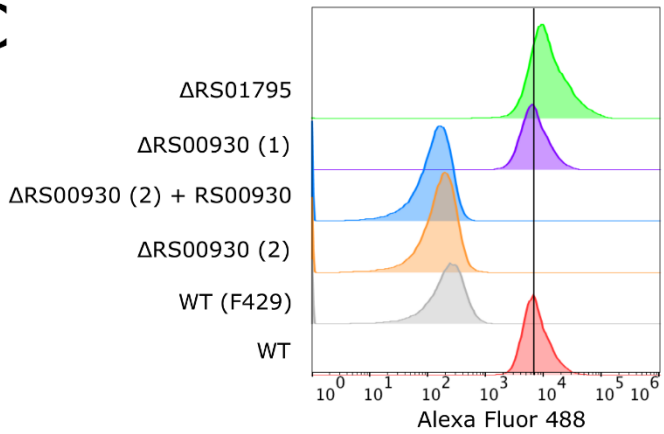
***Streptococcus equi* subspecies *zooepidemicus* ATCC 35246**

Genome size (Mb):	2.17
GC %:	41.6%
Protein coding genes:	1,894
Total TA sites:	146,048

Transposon Library (pMar4s)

Mapped reads:	1,988,331
TA sites hit:	75,610
Percentage of TA sites hit:	52%

C



6.8, 1/p-value = 3.3; and RS00930, WP_014622184.1, L2FC = 4.9, 1/p-value = 416.2) were created, but neither lost MAb F598 binding (Figure 2.2C). A second independently generated ΔRS00930 gene deletion strain lacked MAb F589 binding (Figure 2.2C), but complementation with RS00930 did not restore F598 binding, suggesting that RS00930 is not required for F598 binding. The preservation of F598 binding in the single gene deletion mutants suggested that the Tn insertions did not account for the absence of F598 binding in the screen hits.

Whole genome sequencing was used to identify mutations that were shared by SEZ mutants deficient in F598 binding. Several strains were sequenced, including a few Tn mutants and the RS00930 gene deletion strain that lacked F598 binding. Comparative analyses of these genome sequences revealed that 4 out of 5 F598 negative strains contained one of three distinct mutations in a single gene, RS09465 (WP_038674722.1, old locus tag SeseC_02416), which we have termed *sezV* (for SEZ Virulence, Figure 2.1EF). The amino acid sequence of SezV is predicted to contain two domains, an N-terminal sensor kinase-like domain and an AraC helix-turn-helix (HTH) DNA binding domain, suggesting that the protein is a transcriptional regulator. The remaining F598 negative strain sequenced contained a mutation in the gene coding SzM (RS09460, WP_014623570.1, old locus tag SeseC_02415), a gene adjacent to and divergently transcribed from *sezV* (Figure 2.1EG). None of the three F598-positive strains sequenced contained a mutation in either *sezV* or *szM*. Furthermore, Tn-insertions in both *sezV* and *szM* were hits in the Tn-seq screens (Figure 1.1D & G), suggesting that these linked loci play roles in presentation of PNAG related epitopes in SEZ, and thus facilitate F598 binding.

Figure 2.3. Inactivation of sezV or szM leads to loss of SEZ reactivity with anti-PNAG antibody and to reduction in surface material.

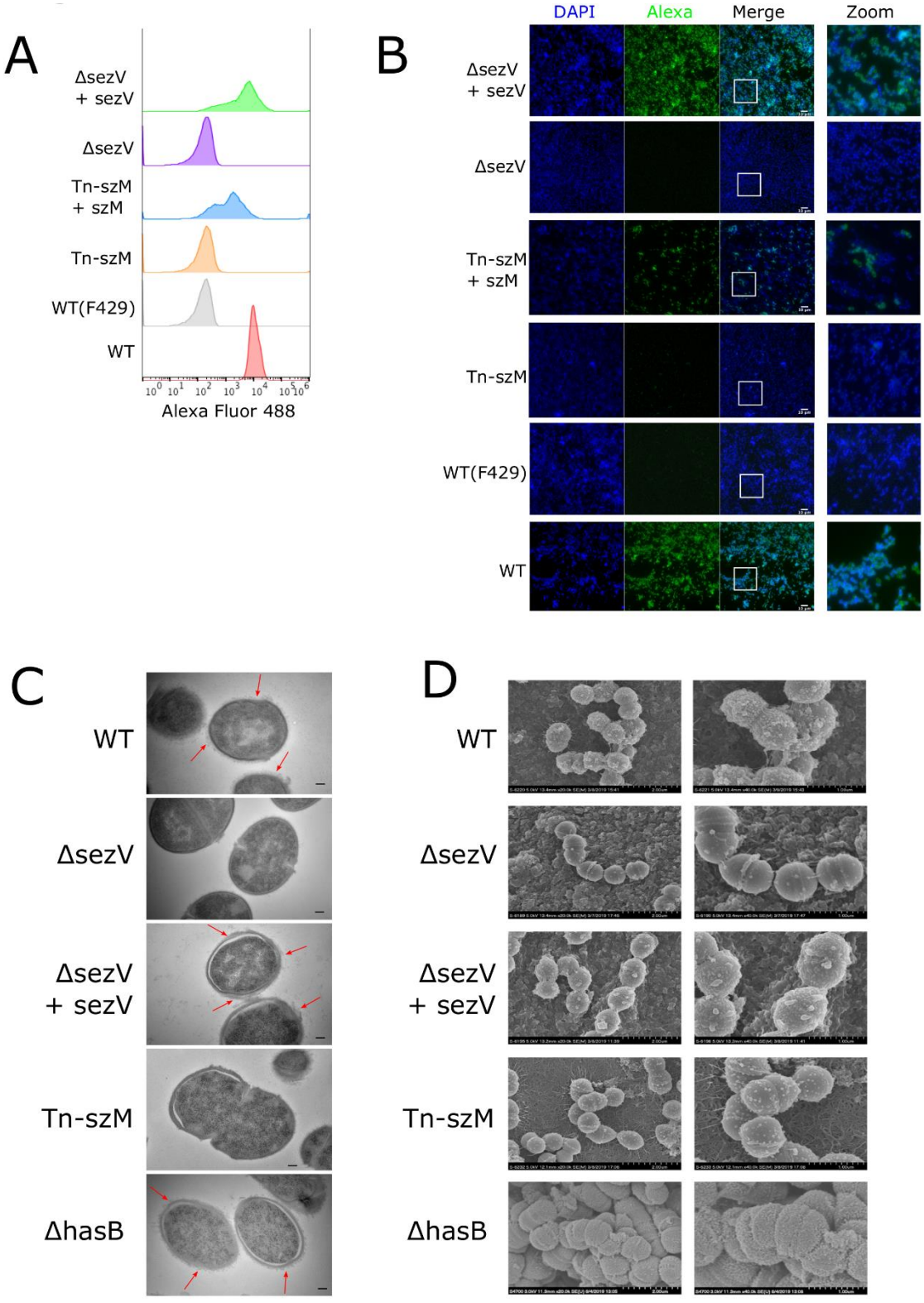
A. Flow cytometry of indicated strains with AF488-F598. The WT strain was also stained with the negative control antibody AF488-F429.

B. Microscopy of SEZ strains stained with DAPI and AF488-F598. The WT strain was also stained with AF488-F429. Rightmost panels provide zoomed view of the merged image. Scale bar in white represents 10 μm.

C. Transmission electron microscopy of SEZ strains. Red arrows indicate surface material present on select strains. Scale bar in black represents 500 nm.

D. Scanning electron microscopy of indicated SEZ strains at lower (20K, left) and higher (40K, right) resolutions.

Figure 2.3 (Continued)



Phenotypic characterization of strains lacking *sezV* or *szM*

A Δ*sezV* mutant was created and this strain did not bind AF488-F598 either in flow cytometry or fluorescence microscopy (Figure 2.3AB). Complementation with *sezV* restored the F598 binding activity, establishing that this putative transcription factor is required for binding of this anti-PNAG MAb. We were unable to construct a Δ*szM* mutant, but found that complementation of the Tn-*szM* mutant (a strain isolated with a transposon insertion in *szM*) with *szM* restored F598 binding activity (Figure 2.3AB).

Transmission electron microscopy (TEM) and scanning electron microscopy (SEM) were used to compare the surfaces of WT, Δ*sezV*, and Tn-*szM* mutant cells. In TEM, the Δ*sezV* and Tn-*szM* mutants had smoother outer surfaces than the WT, *sezV* complemented, and a Δ*hasB* strain, which does not produce the SEZ hyaluronic acid capsule (34) (Figure 2.3C).The surface structures in the WT, Δ*sezV* complemented strain, and Δ*hasB* strain resemble the electron dense surface "fuzzy coat" attributable to M protein fibers on the surface of GAS (35, 36). In SEM, the surfaces of the Δ*sezV* and Tn-*szM* mutants also appeared smoother than the other strains (Figure 2.3D). Together, these observations suggest that *sezV* and *szM* are required for the production of a major surface associated component in SEZ.

sezV activates expression of *szM*

Expression of M/M-like proteins in GAS is activated by *mga*, a gene that is generally found upstream of and transcribed in the same direction as the gene encoding the M/M-like protein. However, no regulator of *szM* (aka *seM*/*fgbp*, *spa*) has been described. Although *sezV* bears no similarity to *mga*, its predicted domain architecture suggests that it is a transcriptional regulator (Figure 2.1F) and the *sezV* gene is adjacent to the *szM* gene in the SEZ genome (Figure 2.4A).

Figure 2.4. *sezV* is required for expression of *szM* and SEZ virulence.

A. (Top) Genomic context of *szM* (RS09460) and *sezV* (RS09465) loci in the SEZ ATCC 35246 genome. (Bottom) Relative expression of *szM* and *nrdI*, in the Δ*sezV* and Δ*sezV* complemented (Δ*sezV* + *sezV*) strains compared to the WT strain.

B. Volcano plot of RNA-seq results comparing WT and Δ*sezV* strains. The most differentially expressed genes, *szM* and three genes from the same fimbriae operon (the *cne*-homolog-containing Fim1 pilus locus), are labeled. Genes whose expression requires *sezV* have a positive fold change.

C. RNA-seq data for a selected set of genes that includes those with the highest log2 fold change [log2(FC); WT / Δ*sezV*]. Adjusted p-value (adj p) from DESeq2. Genes found in gene clusters or operons are highlighted in the same color.

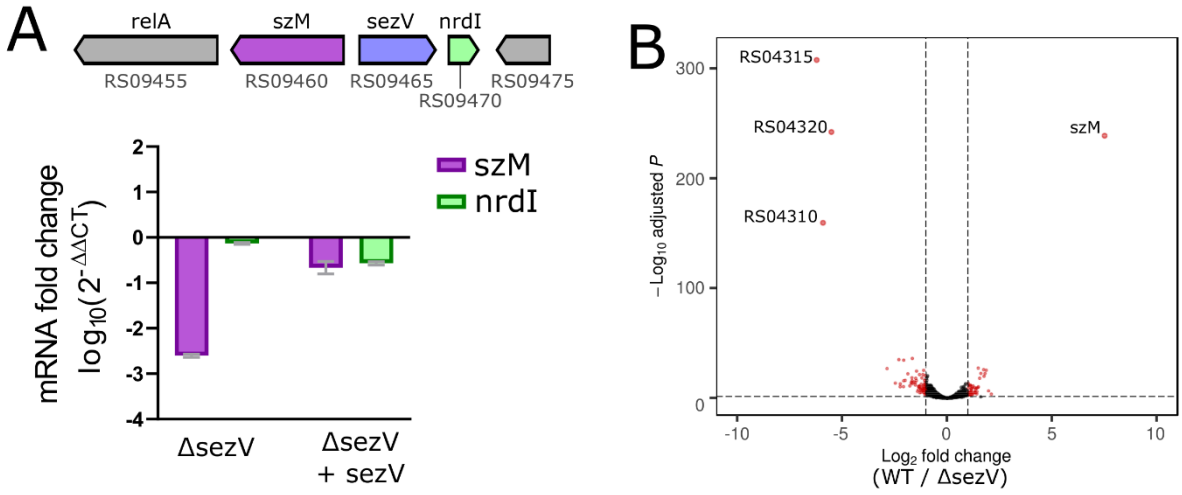
D. Kaplan-Meier survival curves of mice IV inoculated with indicated SEZ strains. Curves were compared using the log-rank (Mantel-Cox) test. Numbers in parentheses refer to animals that died and total animals inoculated.

**** $p < 0.0001$, ** $p<0.001$, * $p < 0.05$; Comparisons that are non-significant are not labeled.

E. SEZ Burden recovered from brains of infected animals 24 hours post infection. Colors of strains are the same as in D. Open circles represent animals for which no colony forming units were recovered. Groups were compared using a Kruskal-Wallis test with Dunn's multiple comparisons test.

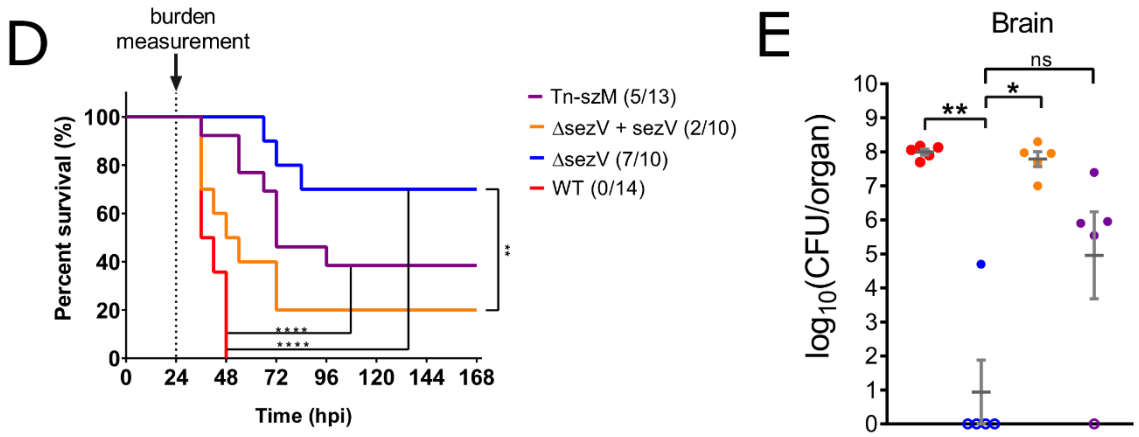
**** $p < 0.0001$, ** $p<0.001$, * $p < 0.05$; Comparisons that are non-significant are not labeled.

Figure 2.4 (Continued)



C

Locus tag	log_2(FC)	adj p	Description
SESEC_RS09460	**7.6**	**2.5E-241**	**(SzM) YSIRK-type signal peptide-containing protein**
SESEC_RS08250	1.9	4.5E-26	hypothetical protein
SESEC_RS05875	1.8	4.5E-23	TVP38/TMEM64 family protein
SESEC_RS05880	1.8	1.4E-20	ABC transporter permease
SESEC_RS08245	1.7	1.6E-26	transporter substrate-binding domain-containing protein
SESEC_RS05885	1.6	4.1E-23	ABC transporter ATP-binding protein
SESEC_RS01435	1.5	9.7E-19	5'-nucleotidase lipoprotein e(P4) family
SESEC_RS03220	1.5	6.3E-11	DNA/RNA non-specific endonuclease
SESEC_RS05890	1.5	5.6E-28	efflux RND transporter periplasmic adaptor subunit
SESEC_RS04365	-2.2	1.3E-13	membrane protein
SESEC_RS02260	-2.2	4.9E-11	tagatose-6-phosphate kinase
SESEC_RS04305	-2.3	1.1E-35	TetR/AcrR family transcriptional regulator
SESEC_RS02265	-2.5	2.6E-14	tagatose-bisphosphate aldolase
SESEC_RS02580	-2.9	1.5E-27	WXG100 family type VII secretion target
SESEC_RS04320	-5.5	7.6E-243	class C sortase
SESEC_RS04310	-5.9	3.3E-160	Cna B-type domain-containing protein
SESEC_RS04315	-6.2	0.0E+00	isopeptide-forming domain-containing fimbrial protein



Quantitative RT-PCR showed that there were reduced levels of *szM* transcripts in the Δ*sezV* background, and this deficiency was complemented by exogenous SezV, strongly suggesting that *sezV* promotes *szM* expression (Figure 2.4A). The transcriptomes of the WT and Δ*sezV* strains were compared to identify other loci regulated by *sezV* (Figure 2.4BC). This comparison revealed that *sezV* has a relatively small regulon; only one gene, sz*M*, had expression reduced >5-fold in the mutant background and only 3 genes, all within a fimbriae operon encoding a Fim1 pilus that includes a *cne* (collagen binding protein of *S. equi*) homolog (37, 38), had expression increased >10-fold in the absence of *sezV*. Expression of the gene encoding SzP, the other M-like protein in SEZ, was largely unaffected in the absence of *sezV* (log_2 fold change 0.8; WT/Δ*sezV*). Thus, *sezV* appears to primarily activate expression of *szM* and repress expression of a Fim1 pilus locus that encodes a *cne* homolog; notably, Cne has been explored as a candidate component of a vaccine for SEE (39–41).

sezV and *szM* both contribute to SEZ virulence

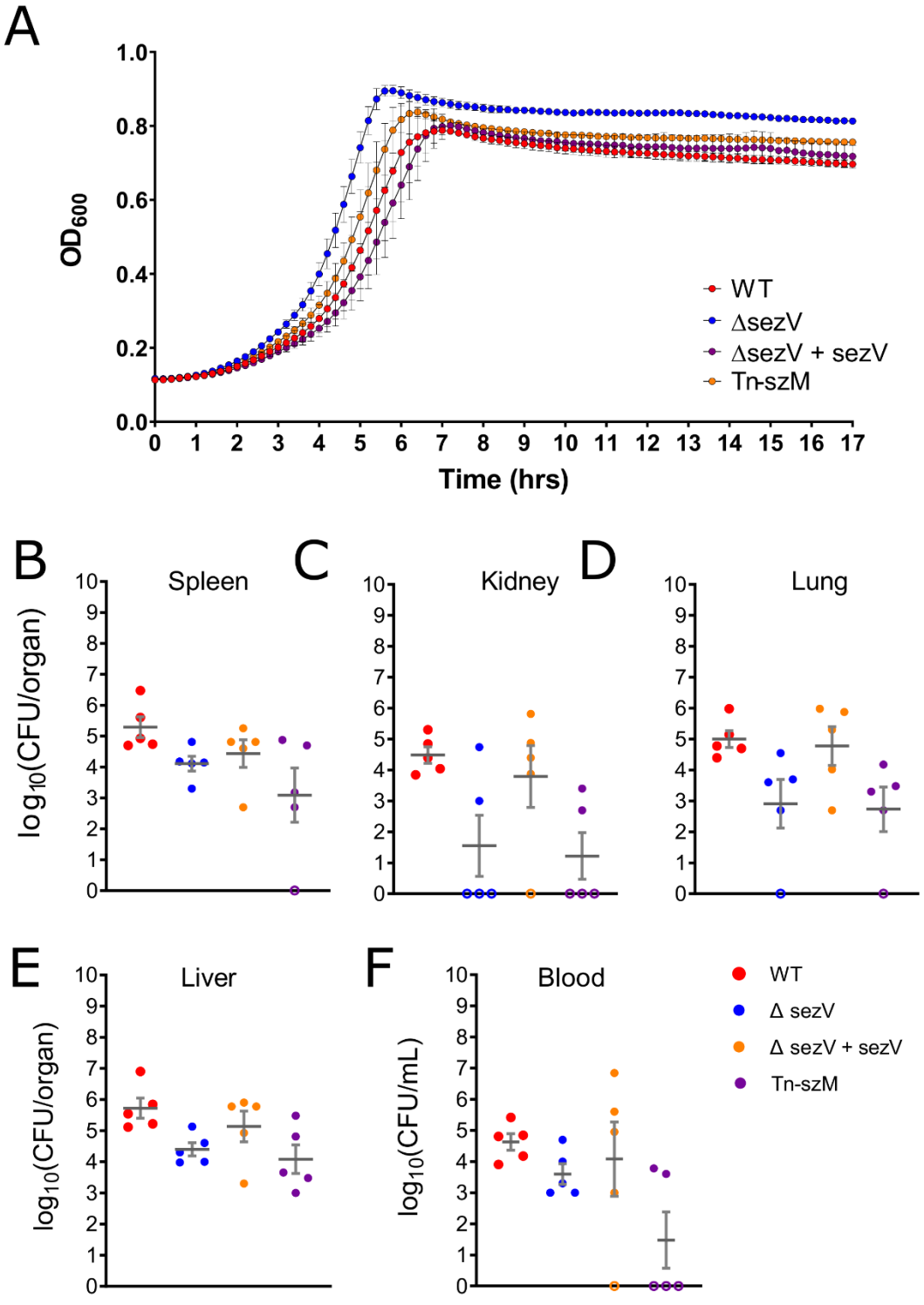
A murine model of SEZ virulence (42) was used to test whether *sezV* or *szM* contribute to SEZ pathogenicity. In this model, I.V. inoculation of female adult C57B6/J mice with WT SEZ results in 100% mortality by 48 hours post infection (hpi) (Figure 2.4D). The Δ*sezV* strain was highly attenuated and exhibited both delayed kinetics of mortality (0% death at 48 hpi) and diminished absolute lethality (30% dead one week after infection). Plasmid-based complementation of the Δ*sezV* mutant, largely restored virulence, indicating that *sezV* is a key regulator of SEZ virulence. The Tn-*szM* mutant was also attenuated in this model, revealing that *szM* contributes to SEZ pathogenicity. There was no difference between the Kaplan-Meier curves of the Δ*sezV* and Tn-sz*M* mutants, raising the possibility that the virulence defect of the

Figure 2.5. Growth of SEZ strains in culture and in vivo.

A. Growth curves of indicated SEZ strains.

B-F. Burden of indicated strains 24 hours post IV inoculation. Open circles represent animals for which no colony forming units were recovered. Results are from the same experiment displayed in figure 2.4E. Groups were compared using a Kruskal-Wallis test with Dunn's multiple comparisons test. Comparisons that are non-significant are not labeled.

Figure 2.5 (Continued)



sezV mutant is largely explained by defective *szM* expression in this strain. In addition, the WT, Δ*sezV* and Tn-*szM* mutant strains grew similarly in culture, suggesting that the attenuation of the mutant strains is not due to growth defects caused by these mutations (Figure 2.5A).

Given the rapid mortality induced by the WT SEZ strain, organs were collected 24 hpi to measure the bacterial burdens of the four strains in infected host tissues. As observed previously, the highest burden of the WT SEZ strain was found in the brain, reflecting the neurotropism of this pathogen (42). Strikingly, in most mice, no Δ*sezV* CFU were recovered from the brains of infected mice and this marked virulence defect was corrected by genetic complementation (Figure 2.4E). The burden of the Δ*sezV* mutant in the blood and other organs was also reduced compared to the wild type strain (Figure 2.5B-F), but the magnitude of the defect was not as pronounced as observed in the brain. The burden of the Tn-*szM* mutant in the brain was also reduced, but not as dramatically as the Δ*sezV* mutant; however, the burden of these two mutants was similarly reduced in other organs (Figure 2.4E, Figure 2.5B-F).

Monoclonal antibody to PNAG binds SzM

Our observations that the Tn-*szM* mutant was not recognized by MAb F598 in either flow cytometry or by fluorescence microscopy (Figure 2.3) and that F598 binding was restored by expression of *szM* strongly suggest that surface presentation of SzM and PNAG are linked. Immunoblots confirmed this linkage. In western blots of lysates from WT SEZ, a band of ~60 kD, slightly greater than the predicted molecular weight of the mature form of SzM (~57 kD, after N and C terminal processing) was detected with the F598 MAb to PNAG (Figure 2.6A). This band was not detected in lysates derived from strains with the transposon insertion in *szM* (Tn-*szM*) or the deletion of *sezV*, but was restored by complementation of the respective

Figure 2.6. SzM is directly bound by an anti-PNAG antibody.

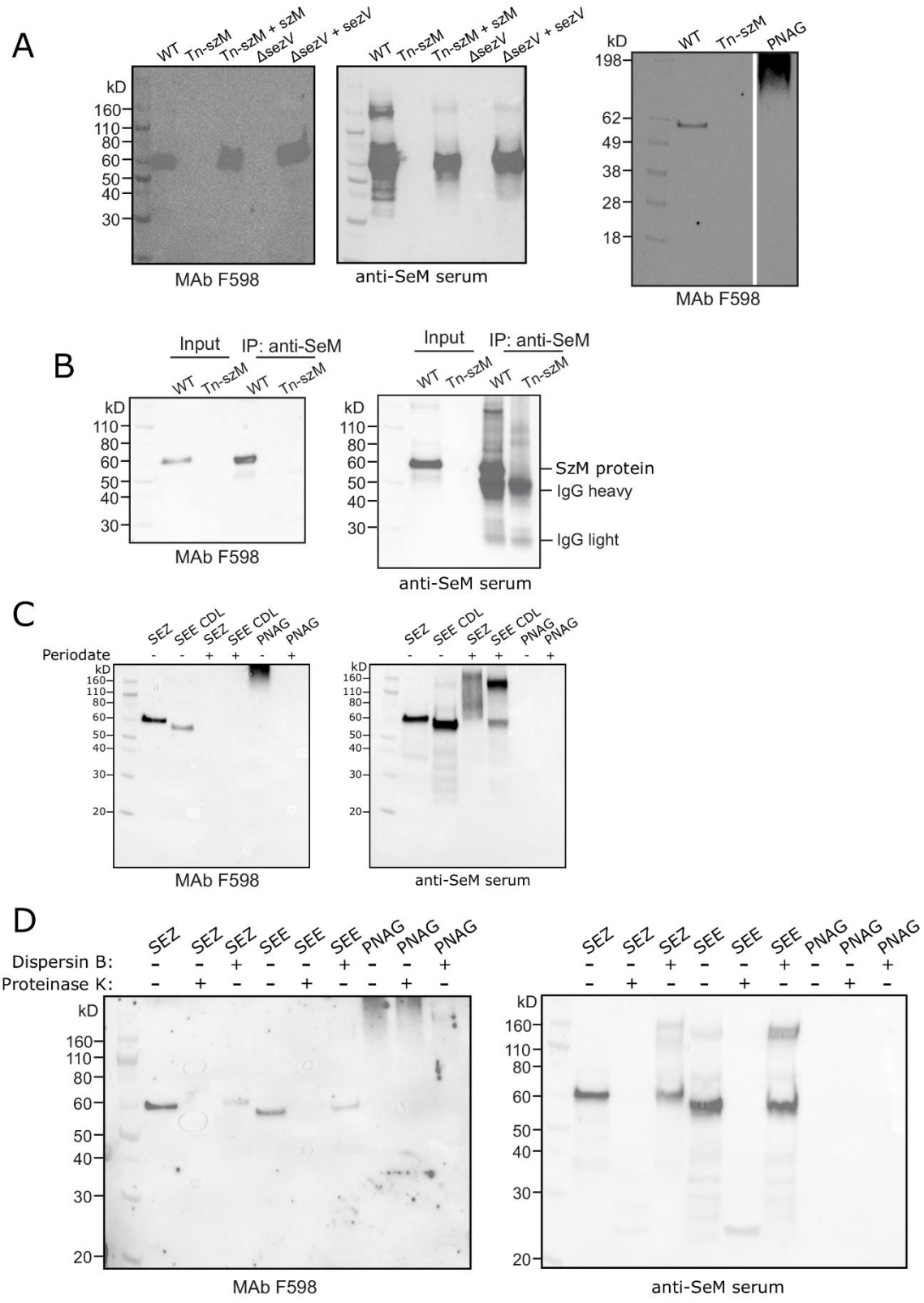
A. Western blot of bacterial lysates from SEZ strains. Electrophoresed lysates from the indicated strains were run in duplicate on a single SDS-PAGE gel, transferred to single membrane, which was cut into two parts and each half was blotted with the indicated antibody.

B. Western blot of immunoprecipitated SzM from SEZ WT and Tn-*szM* strains after SDS-Page.

C. Western blot of bacterial lysates of SEZ, SEE strain CDL, or purified PNAG carbohydrate with (+) or without (-) periodate treatment.

D. Western blot of bacterial lysates of SEZ, SEE strain CDL, or purified PNAG carbohydrate with (+) or without (-) treatment with either the dispersin B or proteinase K enzymes.

Figure 2.6 (Continued)



mutations (Figure 2.6A); moreover, a band of apparently the same molecular weight was observed in the same lysates (run on the same SDS-PAGE gel) when antisera raised against recombinant SEE M-like SeM protein (anti-SeM sera) was used (Figure 2.6A). When the blots were probed with MAb F429, an isotype control that was used at the same concentration as F598, bands were not detectable, though they became faintly detectable with longer exposure times (Figure 2.7A). Together, these observations suggest the possibility that SzM is decorated by some form of PNAG. However, PNAG is ordinarily an extended capsule-like polymer that barely enters polyacrylamide gels such as the purified material used in Figure 2.6A (and Figure 2.7B).

To confirm that the antibody to PNAG and the antibody to M-protein bound to the same protein, SzM was purified using anti-SeM sera; no protein was purified from lysates of Tn-*szM* using the same protocol (Figure 2.7C). The purified protein was bound by F598 as well as anti-SeM (Figure 2.6B). Mass spectrometry analysis of the peptides produced by trypsin digestion of the protein immunoprecipitated with the anti-SeM antisera confirmed that it was SzM; peptides were detected that span the majority of the predicted processed form of the protein (i.e., lacking the signal sequence and extreme C-terminus, which is presumably cleaved as a result of sortase mediated attachment of SzM to the cell wall) (Figure 2.7D). Together, these observations establish that the MAb F598 to PNAG binds to SzM. F598 also bound to a single band recognized by anti-SeM in lysates of SEE strain CDL (Figure 2.6C), suggesting that a PNAG-like molecule may decorate SzM homologs in other species.

To further investigate the chemical bases of the epitopes of SzM and SeM bound by F598, lysates of SEZ and SEE were treated with sodium periodate. This reagent opens the rings between vicinal hydroxyl groups generating aldehydes and is known to disrupt the structure of

Figure 2.7. Analysis of SzM binding to isotype control antibody F429, immunoprecipitation of SzM, and glycan mass spectrometry of purified SzM

A. Western blot of bacterial lysates from SEZ strains after SDS-PAGE with Mab F429.

B. Western blots of bacterial lysates from SEZ strains or purified PNAG carbohydrate after SDS-PAGE with Mab F429. Purified PNAG carbohydrate is not recognized by the isotype control antibody F429.

C. Coomassie blue stained SDS-PAGE gel of immunoprecipitated SzM. Anti-SeM sera was used to immunoprecipitate SzM from the indicated strains. IgG heavy and light refer to the antibody fragments of the anti-SeM sera. * refers to an unrelated protein present in WT and Tn-SzM strains that was enriched after immunoprecipitation.

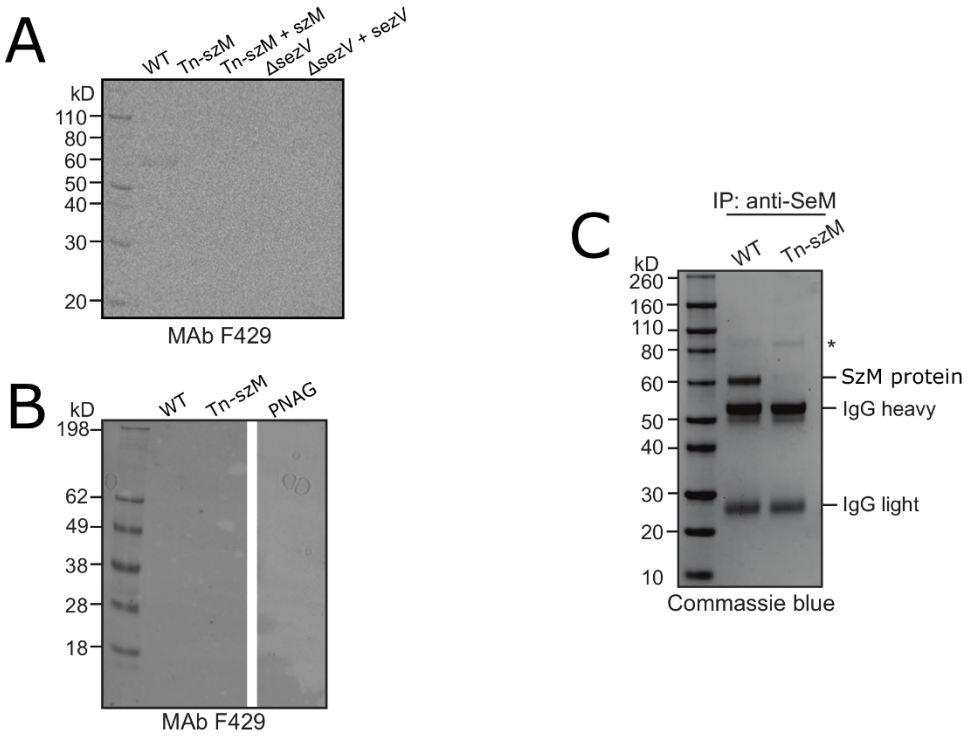
D. Mass spectrometry analysis of polypeptide sequence of SzM from SEZ. After immune-precipitation of SzM, the band corresponding to SzM was cut out and digested with trypsin prior to performing mass spectrometry. Amino acids underlined and in green correspond to regions to which a peptide mapped; no peptide mapped to amino acids in black. We did not expect to obtain coverage of the extreme N- and C-termini because these are predicted to be cleaved off in the processed, mature form of SzM. The predicted signal sequence [YF]SIRKxxxGxxS[VIA] and cell-wall anchoring motif LPxTG are underlined in black.

E. Glycan mass spectrometry plots of immunoprecipitated SzM (M protein) and control glycoprotein, fetuin (a eukaryotic glycoprotein). Proteins were analyzed for O-glycans (left) and N-glycans (right). No masses corresponding to carbohydrate modifications were found in the O-glycan fraction. In the N-glycan fraction, SzM (M protein) only contained background carbohydrate signals, which were also present in the control sample and do not represent glycosylation of a protein. Expected N-glycans were recovered from the fetuin sample. In both

Figure 2.7 (Continued)

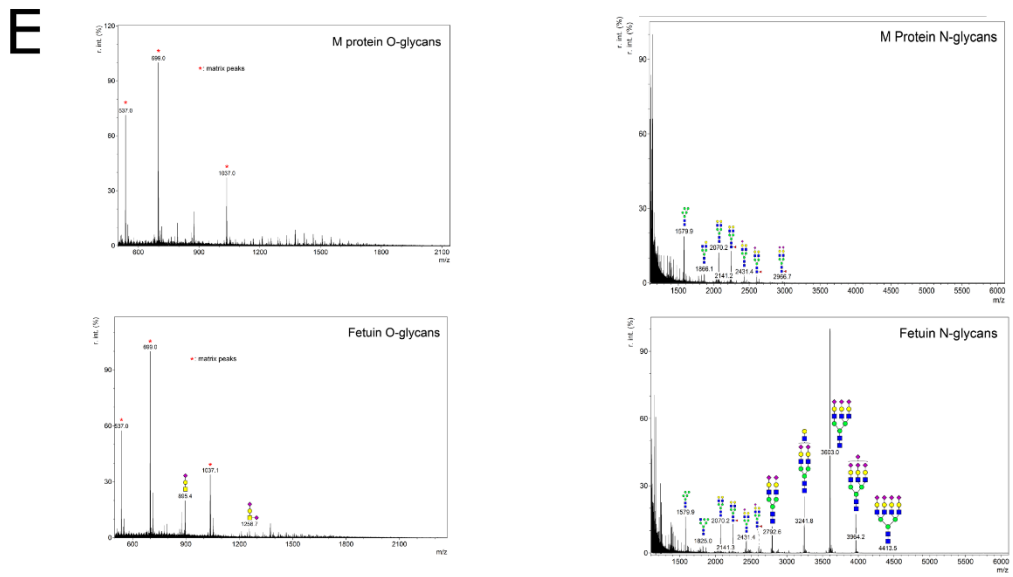
the O- and N-glycan samples of SzM (M protein), no unique carbohydrates were found that were not also found in fetuin.

Figure 2.7 (Continued)



D

```
MFLRNNKQKF SIRKLSAGAA SVLVAVSVLG GAAVKADTGT GVNRSFETLL KGKNVSNETS SRLVNVIAEL EATQWRDTDS ALKKLREILS DASVETIQAL
MRGLDSARTI TYGNDRANWF YRKLNTLGND RPNDVYRSRG LNLYSLLIDE VEQRLSKGES YAGVIKDKEI LIKLKDQWLK DKDTDLLKAK KEKEAILKHN
EENIKIKDQL LKEKDLAIVA LDKEKESITN ALKTSLKLTK EKEQELAKAR EANKSLSESV TKTLARSSKI ANELKDKLAA SEKDKDRAIQ ITTELANKLT
DAENSRDKAF AVSTELANKL AAKTAEAEKL MQNVGSLDRL VESAKREMAE KLAEIDQLTA DKAKADAELA TANDTIASLQ TELEKAKTEL AVSERLIESG
KREIAELQKQ KDASDKALAE SQANVAELEK QKAASDAKVA ELEKEVEAAK AEVADLKAQL AKKEEELEAV KKEKEALEAK IEELKKAHAE ELSKLKEMLE
KKDHANADLQ AEINRLKQEL ADRIKSLSQG GRASQTNPGT TTAKAGQLPS TGESANPFFT IAALTVIAGA GMAVVSPKRK EN
```



PNAG such that it is no longer recognized by F598 (27) (Figure 2.6C). Periodate treated bacterial lysates of SEZ and SEE were no longer bound by F598, but retained anti-SeM binding; however, the molecular weight of the proteins recognized by the anti SeM antisera shifted to slower migrating forms, suggesting that the periodate treatment may have led to the formation of higher-order oligomers of SzM and SeM (Figure 2.6C). These lysates were also treated with dispersin B, an enzyme that specifically cleaves the β-1,6 linkage between glucosamines in PNAG (43). After prolonged treatment with this enzyme, there was a marked reduction in F598 binding with only a minor reduction in binding of the anti-SeM antibody (Figure 2.6D). Protease treatment ablated reactivity with both F598 and anti-SeM sera, but had minimal effect on purified PNAG polysaccharide. Together, these immunochemical experiments with carbohydrate- and protein-specific degradative enzyme's effects on binding of F598 suggest that SzM and SeM are modified by a carbohydrate with a chemical composition similar to PNAG. Importantly however, we did not detect either O- or N- linked glycans in mass spectrometry analyses of purified SzM protein (Figure 2.7E). Thus, we cannot definitively conclude that these M-like proteins are glycosylated.

Conservation of and variation in *sezV* and *szM/seM/spa*

Given the importance of *sezV* and *szM* in SEZ virulence, we investigated whether orthologs of these virulence-linked genes exist in publicly available streptococcal genomes. Notably, homologs of *szM/seM/spa* or *sezV*, based on DNA sequence, were restricted to SEZ and SEE in GCS, and a subset of GAS strains. We did not find *szM* or *sezV* homologs in any other streptococcal species or Lancefield groups, nor in *Lactococcus* or *Enterococcus*, two related genera (Figure 2.8A). All strains with *szM* orthologs harbored an adjacent and divergently

Figure 2.8. Conservation of *szM/seM/spa* and *sezV*.

A. Conservation of *sezV* and *szM/seM/spa* across various streptococcocal species and Lancefield groups, and closely related genera. The cladogram is for illustrative purposes and show a qualitative representation of the relatedness of the species and genera, and was constructed based on Gao et al. (44) and Hug et al. (45). Branch lengths are not to scale nor a representation of the true genetic distances between species and strains. Genomes analyzed indicates the number of genomes queried in each species; +X represents number of strains newly sequenced in this study. Alignment to *sezV* and *szM* indicates species that have similarity to both genes. Lancefield groups are as in Facklam (46) and Okura et al. (47); **S. suis* was formerly classified as Group D (47).

B. Three classes of SzM/SeM/Spa-related proteins encoded adjacent to *sezV*. Visualization of multiple sequence alignment between representative members of each of the three protein clusters demonstrates large differences between clusters and similarity within clusters. Lines represent gaps in the multiple sequence alignment, blocks represent continuous sequences.

C. Amino acid sequence variation in SzM/SeM/Spa proteins. (Top) Similarity plot of amino acids at each position in the protein across all variants of the protein. Similarity (y-axis) represents the relative sequence conservation over a set of adjacent residues; a higher value indicates more conservation. (Middle) Schematic of SzM/SeM/Spa with labeled and shaded domains and regions. SS = Signal sequence, which contains YSIRK motif; Rep A/B = repeat A/B; CC = predicted coiled-coil domain; W = cell wall and cell membrane spanning region, which contains LPXTG motif. Delineations of regions taken from refs (16, 48, 49); coiled-coil domain determined by MARCOIL (50). (Bottom) Multiple sequence alignment of representative SzM/SeM/Spa variants demonstrates various domain/region architectures present across the

Figure 2.8 (Continued)

variants. Lines represent gaps in the multiple sequence alignment, blocks represent continuous sequences.

Figure 2.8 (Continued)

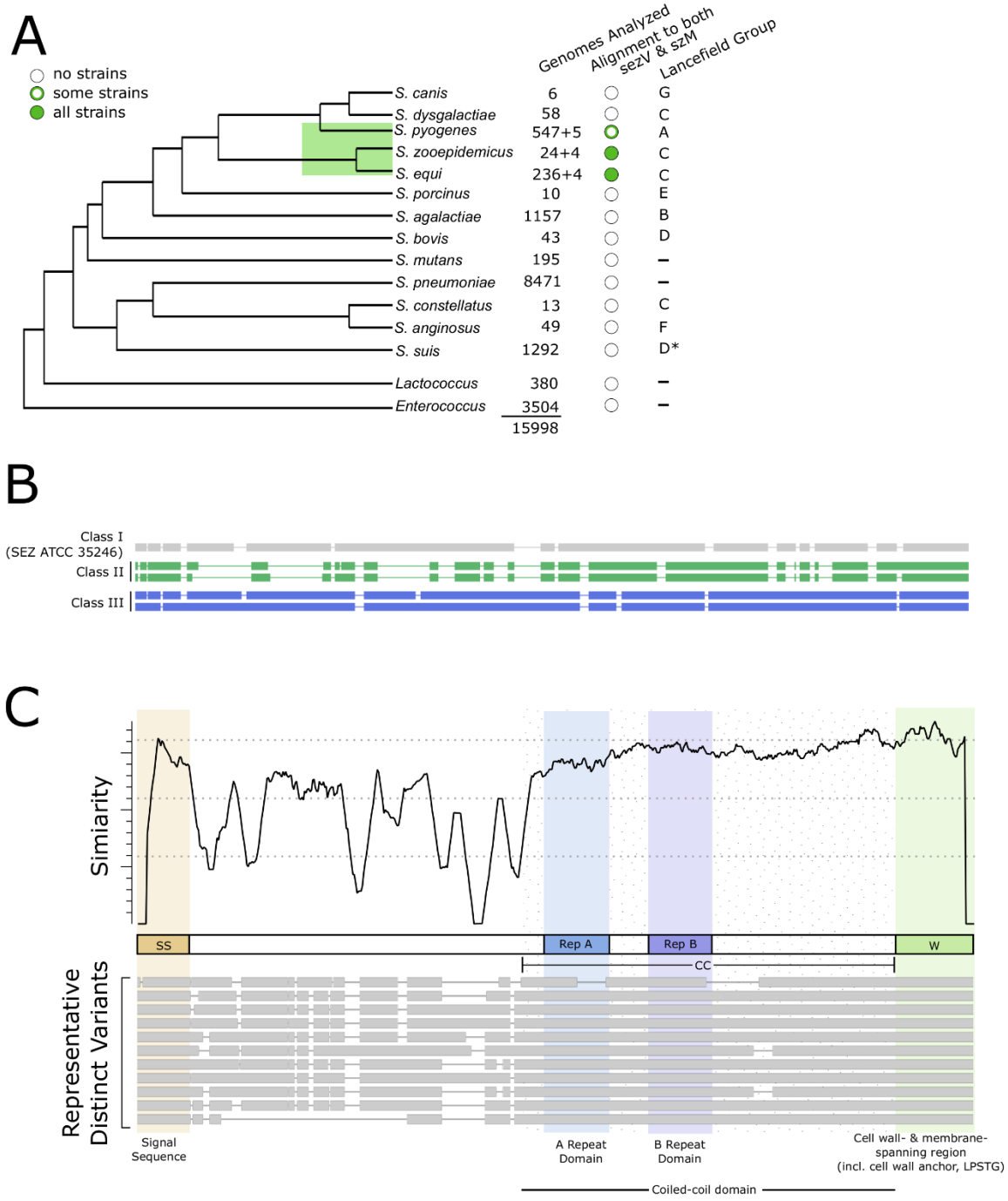


Figure 2.9. Analyses of SzM/SeM/Spa conservation and variants

A. Distance between start sites of *szM/seM/spa* and *sezV* in strains containing the two genes. In 6 strains with draft genomes, the two genes were located on different contigs. We excluded these genes from the distance and orientation analyses.

B. Relative orientation of *szM/seM/spa* and *sezV* in strains containing the two genes. The 6 strains in which the two genes were not found on the same contig were excluded from this analysis.

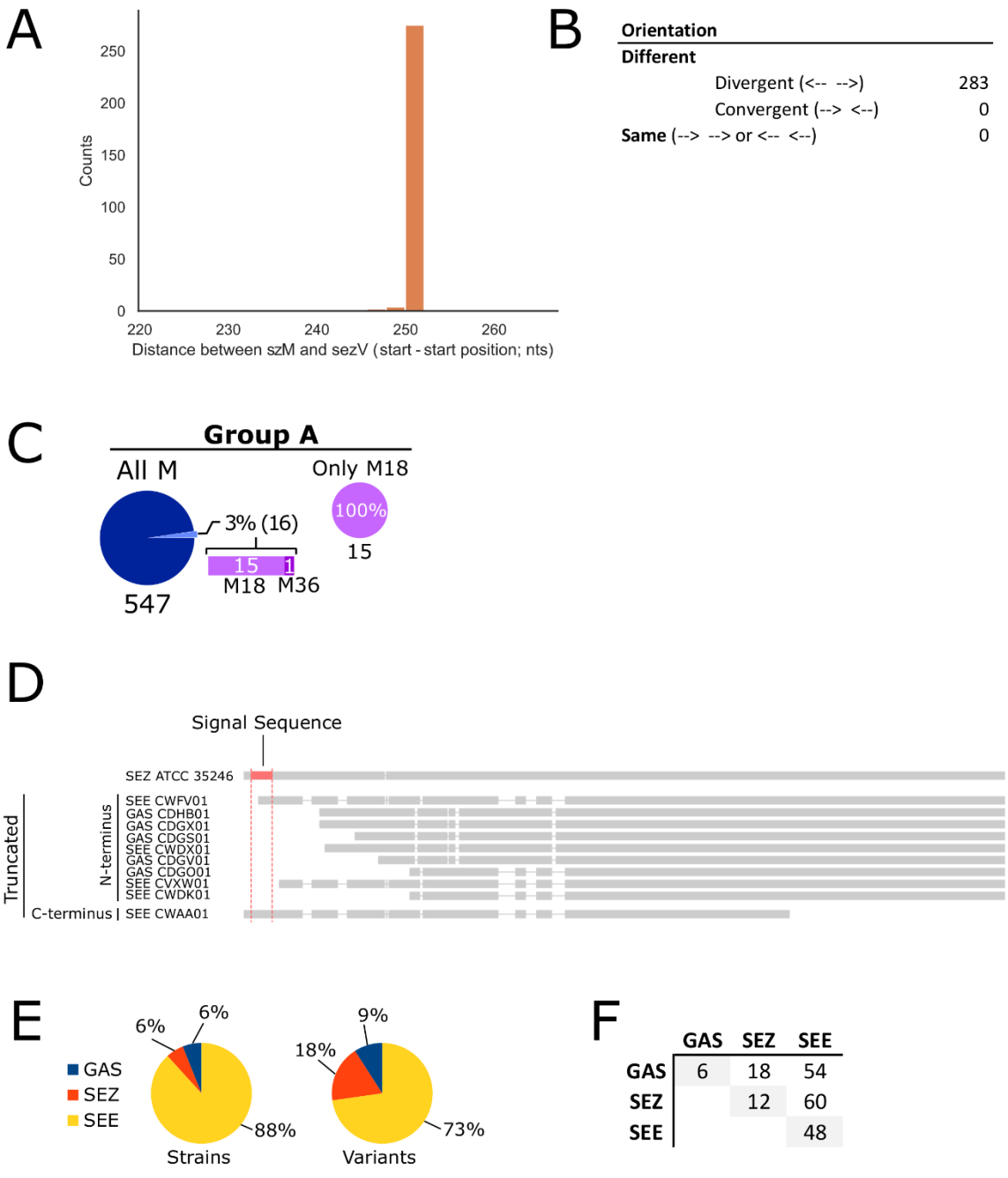
C. Strains in group A *Streptococcus* (GAS) that contain both *spa* (*szM* ortholog) and *sezV*.

D. N- and C-terminal truncated SzM/SeM/Spa variants discovered in several strains. Translated protein sequences of these genes were not included in additional analyses of *szM/seM/spa* due to the deletion of the signal sequence or cell wall anchoring motif, which will likely render them non-functional or not expressed on the cell surface. Species (SEZ, SEE, or GAS) and strain name or WGS identifier is indicated on the left.

E. Distribution of total numbers of strains containing *szM/seM/spa* included in our analyses, and distribution of number of SeM/Spa variants per species. There were a total of 66 SzM/SeM/Spa variants identified. Although the largest percentage of variants belong to SEE, there were ~10x fewer SEZ strains than SEE strains and the percentage of variants within SEZ is higher than the percentage of SEZ strains in the database analyzed. Hence, the results suggest that SEZ is a more diverse subspecies than SEE.

F. Number of SzM/SeM/Spa variants in each species. The diagonal indicates the number of SzM/SeM/Spa variants in each species. The off-diagonal indicates the number of SzM/SeM/spa variants in two species, specified by the row and column header. No variants are shared across a pair of species, or across all three species.

Figure 2.9 (Continued)



oriented *sezV* (Figure 2.9AB), suggesting that SezV regulation of SzM is conserved. Among GAS strains, we found that all 15 M18 strains in the database and 1 M36 strain contained the szM ortholog (*spa*) and *sezV* (Figure 2.9C).

The amino acid sequences of SzM/SeM/Spa and SezV from all strains in the database as well as from 13 newly sequenced strains (see below) were compared. There was considerable variation in the 289 SzM/SeM/Spa amino acid sequences analyzed. These M-like sequences could be divided into 3 classes (Figure 2.8B). Most of the sequences (274) were similar to SzM from SEZ ATCC 35246; the 2 other classes, of 7 and 8 proteins respectively, were primarily found in SEZ strains and also encoded adjacent to a *sezV* ortholog. Some strains with truncated SeM/Spa variants were identified (Figure 2.9D), which may contribute to persistence of these strains in their hosts (51), but these were not included in the analyses below.

There were 66 variants found among the 264 full-length class I SzM/SeM/Spa proteins (Figure 2.9EF). No variants were found in common between species (Figure 2.9F). Multiple sequence alignment of the amino acid sequences of the SzM/SeM/Spa variants revealed that most of the variation was found in their respective N-termini beginning after their signal sequences (SS) (Figure 2.8C), concordant with observations regarding SeM proteins in several SEE strains (49, 52). The highest similarity in the SzM/SeM/Spa proteins was found in their signal sequences, coiled-coil domains, and C-terminal cell-wall and cell-membrane spanning regions (W) (Figure 2.8C). The coiled-coil domains contain recognizable A and B repeats, regions of short repeated sequences previously identified in SeM/SzM proteins found in SEE and SEZ (9, 16).

In general, the 289 SezV proteins analyzed exhibited more conservation than the SzM/SeM/Spa proteins (Figure 2.10A), and have similar domain organization and lengths as

Figure 2.10. Analysis of SezV conservation

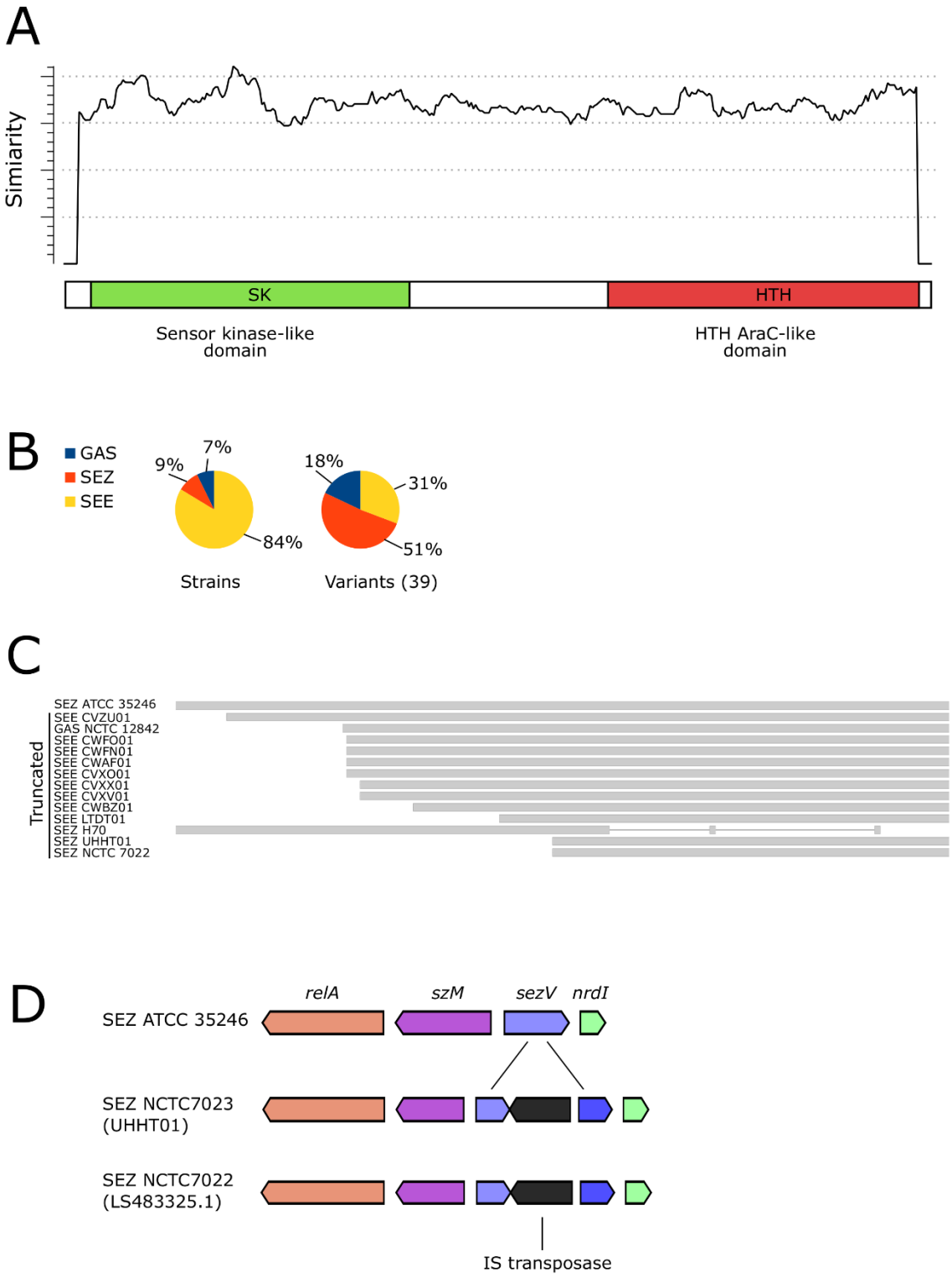
A. Variation in SezV polypeptide sequences. (Top) A similarity plot of amino acids at each position in the protein across all variants. (Bottom) Schematic of SezV with labeled domains. HTH = helix-turn-helix.

B. Distribution of total numbers of strains containing *sezV* included in our analyses, and distribution of number of SezV variants per species. There were a total of 39 SezV variants identified. The greatest number of variants were found among SEZ strains. The percentage of variants belonging to SEZ or GAS was greater than the percentage of the representation of these strains in the database, suggesting that these species have more SezV diversity than found in SEE.

C. N- and C-terminal truncated *sezV* variants identified in several strains. Translated protein sequences of these genes were not included in additional analyses of SezV. Species (SEZ, SEE, or GAS) and strain name or WGS identifier is indicated on the left.

D. *sezV* variants interrupted by an insertion sequence transposase.

Figure 2.10 (Continued)



SezV from SEZ ATCC 35246. Thus, *sezV* appears to be a conserved transcriptional regulator. However, 39 variants with single amino acid substitutions or small deletions were identified (Figure 2.10B) along with several loci that contained more severe changes (Figure 2.10CD).

SzM/SeM/Spa is recognized by the antibody to PNAG in SEZ, SEE, and M18 GAS

Given the presence of *szM* and *sezV* orthologs in SEZ, SEE, and M18 GAS strains, we investigated whether MAb F598 to PNAG recognized SzM-related proteins from additional strains. Western blots of bacterial lysates derived from eight SEZ and SEE strains isolated from horses with anti-SeM sera detected proteins of ~50 to 60 kD in all strains but longer exposures were needed to detect the reactive band in some SEZ isolates (e.g. SEZ 14102 and 17006, Figure 2.11A, Figure 2.12A), whose SzM/SeM amino acid sequences differed more from that of SEZ ATCC 35246 (WT) than those from the other SEE isolates. Notably, F598 recognized a protein of size similar to that of the SzM reactive band in all strains tested except for SEZ 14102, an isolate whose SzM sequence contained a 59 aa deletion (Figure 2.11A, Figure 2.12B). Thus, the anti-PNAG F598 antibody is broadly reactive with diverse SzM/SeM proteins from equine SEZ and SEE isolates.

We also investigated the expression of M and Spa proteins in M18 GAS isolates and whether MAb F598 bound to any of these proteins. The genomes of five clinical M18 isolates were sequenced and assembled (Figure 2.13A). All five of these M18 strains contained linked and divergently oriented *sezV* and *szM* orthologs (*spa*). The locus containing *sezV* and *spa* was not linked to the locus encoding the canonical M18 M protein (*emm*) or its associated regulator, *mga* (Figure 2.11B). Western blotting of lysates from these five strains with the F598 MAb revealed an ~60 kD band, whereas blots with the isotype control MAb F429 were negative

Figure 2.11. Anti-PNAG antibody binds SzM/SeM/Spa in several SEZ, SEE, and GAS strains.

A. Western blots of bacterial lysates of indicated SEZ and SEE strains after SDS-PAGE with anti-SeM or F598 antibodies.

B. Genomic context of the *sezV, spa* (the *szM* ortholog) and mga loci in a canonical M18 GAS strain, MGAS8232 (53).

C. Western blots of bacterial lysates of indicated SEZ and newly sequenced M18 GAS strains with F598, anti-SeM or anti-GAS M antibodies. The anti-GAS M serum was sera raised against a purified, canonical GAS M protein. Lower bands in GAS strains may be indicative of M protein degradation products, as seen in (18).

Figure 2.11 (Continued)

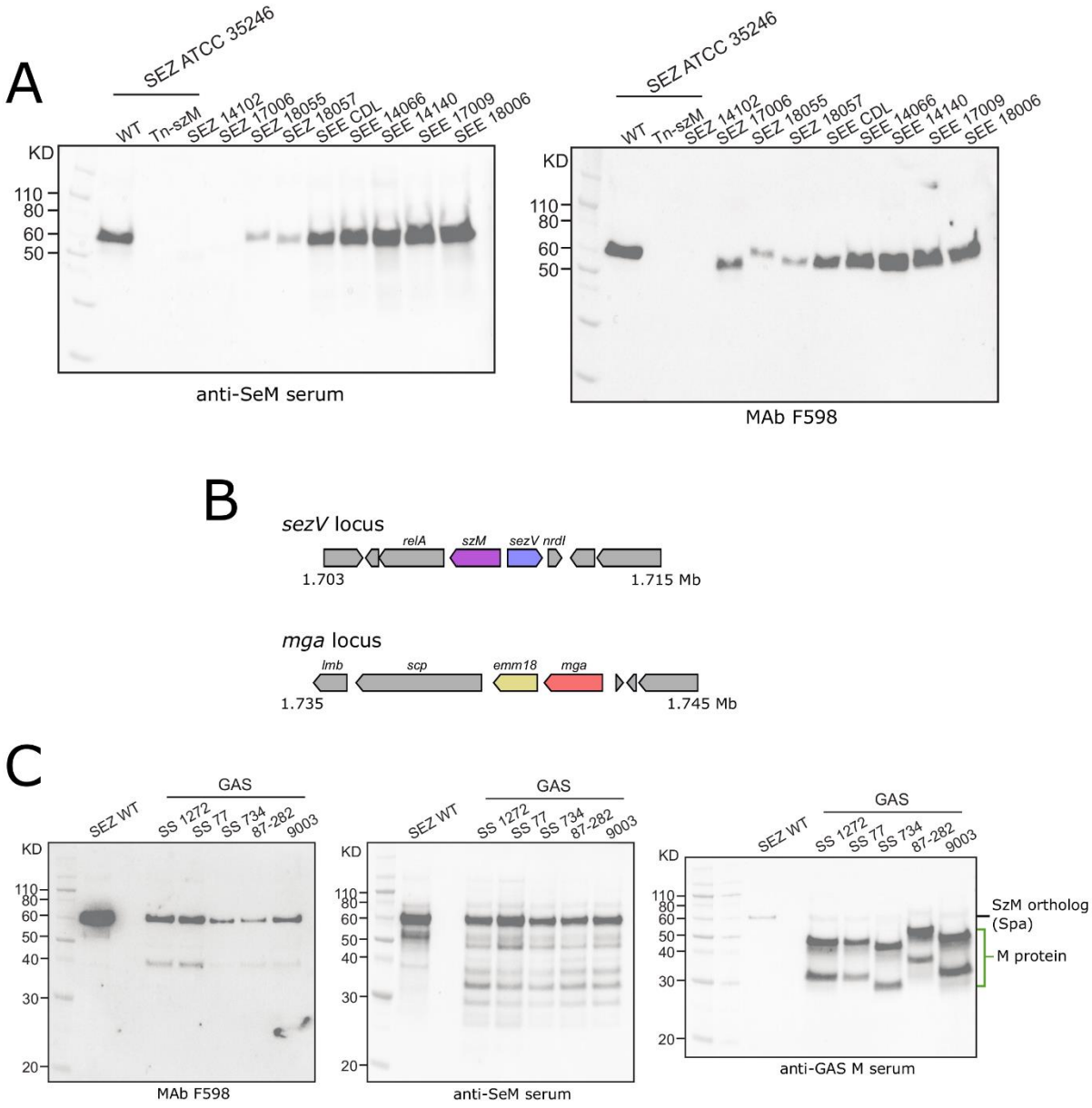
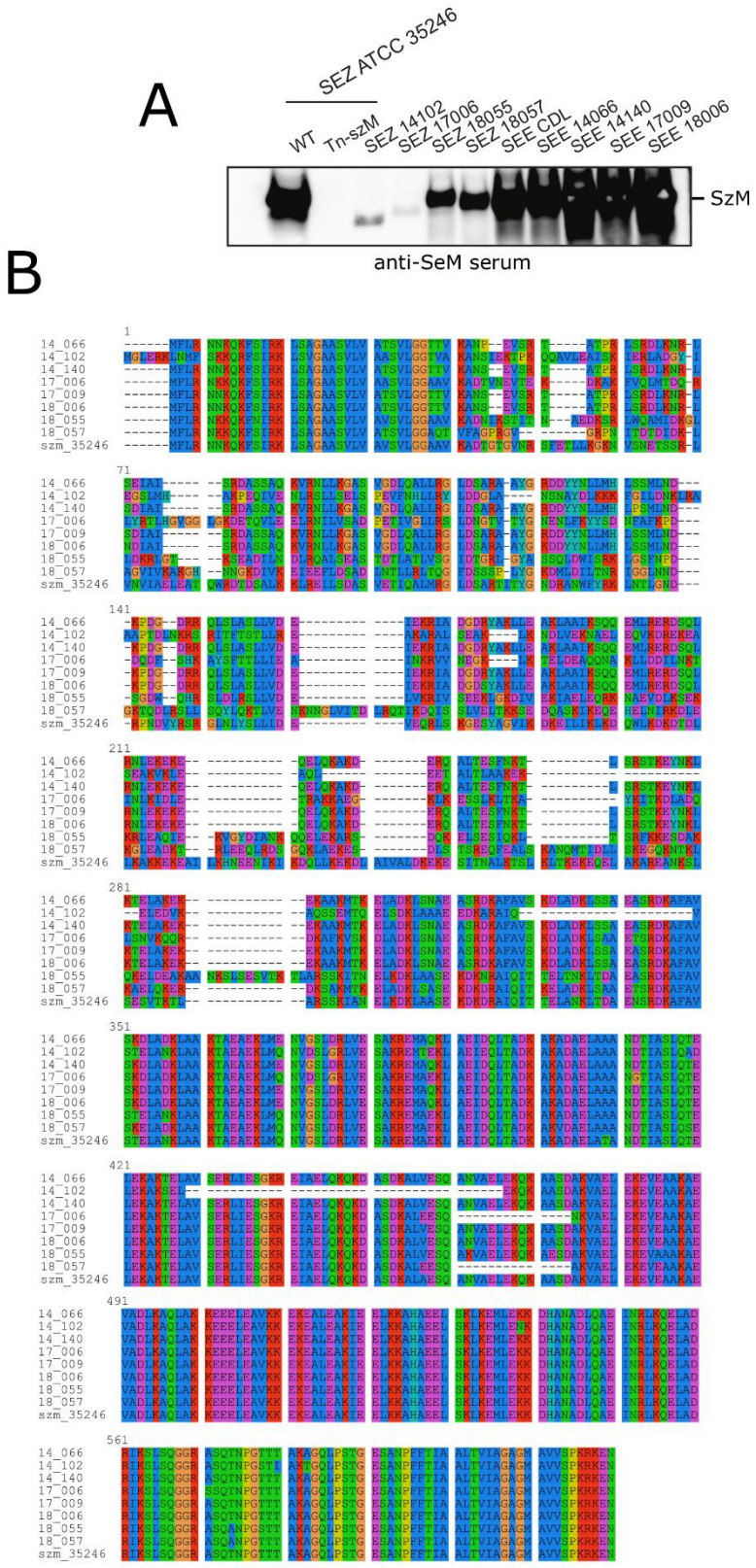


Figure 2.12. SzM/SeM from several SEZ and SEE strains

A. Western blot of bacterial lysates from SEZ and SEE strains. Longer exposure of the membrane demonstrates expression of SzM in SEZ 14102 and 17006. This is the same membrane shown in Figure 2.11A (left).

B. Multiple sequence alignment of SzM/SeM from the indicated strains demonstrates that the 14102 strain has two large deletions in the relatively well conserved C-terminal region of the protein.

Figure 2.12 (Continued)



(Figure 2.11C, Figure 2.13B). A very similar sized band was also detected in blots using anti-SeM serum; this serum also weakly reacted with bands of ~46 and 34 kD, which likely correspond to the M18 protein because bands of these sizes were seen in blots using anti M18 antisera (Figure 2.11C). Notably, however, these two bands (likely corresponding to the M18 protein) were not detected with the anti-PNAG antibody. Thus, these M18 GAS strains appear to express at least two M-like proteins, the canonical M18 M protein, which is not recognized by F598, and Spa, which is bound by this anti-PNAG monoclonal antibody.

Discussion

Our investigation of genes required for surface PNAG reactivity in a porcine SEZ isolate led to the identification of *szM*, which encodes an M-like protein and its linked activator *sezV*. SzM/SeM/Spa (also called FgBP), is representative of a class of M-like proteins distinct from M, Mrp, and Enn. SzM/SeM/Spa proteins appear to be decorated with a PNAG-like oligosaccharide, although this conclusion is based on immunochemical analysis rather than definitive chemical isolation of an oligosaccharide associated with these M-like proteins. Both *szM* and *sezV* are required for robust SEZ virulence and homologues of these linked virulence genes were identified in all SEZ, SEE and M18 GAS genomes in the database. Thus, the *szM/seM/spa*, *sezV* locus appears to define a subtype of virulent streptococci.

We used labeled MAb F598 to PNAG to carry out a FACS-based screen of a Tn library in SEZ strain ATCC 35246 to identify insertion mutants that were deficient in surface expression of PNAG. The screen appeared to work well, and even after one round of selection, mutants lacking F598 binding were identified (Figure 2.1C); after three rounds of selection, the population was nearly uniformly PNAG negative. However, the library diversity was compromised by Tn

A

		Emm type (or subtype)
GAS M18 87-282		emm 18.1
# contigs	59	
Total length	1850575	
N50	71739	
GAS M18 9003		emm 18.1
# contigs	95	
Total length	1888727	
N50	38025	
GAS M18 SS-77		emm 18
# contigs	64	
Total length	1835639	
N50	74541	
GAS M18 SS-734		emm 18.17
# contigs	51	
Total length	1849878	
N50	85574	
GAS M18 SS-1272		emm 18.1
# contigs	49	
Total length	1850497	
N50	74220	

contigs refers to contigs >= 500 bp

B

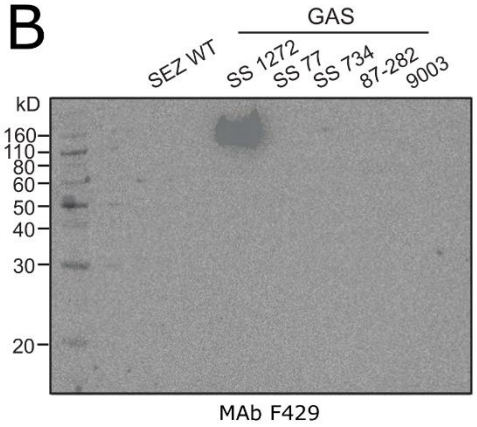


Figure 2.13. Assembly statistics for newly sequenced GAS strains and lack of binding of the isotype control antibody F429.

A. Assembly statistics for newly sequenced GAS strains. *emm* type and/or subtype indicated on right determining via blastp for the M protein and typing via the CDC *emm* blast server. Subtypes are indicated only if the queried M protein had 100% identity over 100% of the subject typing sequence.

B. Western blot of bacterial lysates from SEZ and newly sequenced M18 GAS strains.

jackpots. Linkage between the phenotype (the absence of F598 binding to the Tn insertion mutant) and the genotype (the site of transposon insertion) was not established for several insertions. Whole genome sequencing led to the identification of mutations in the *sezV, szM* locus that were shared among independently derived PNAG-negative SEZ mutants. We were unable to enrich for PNAG negative cells when a WT (non-mutagenized) SEZ culture was FACS sorted and it is not clear why PNAG negative mutants were easily detected in the Tn library.

Although SezV lacks sequence similarity to Mga, it appears to be a functional analogue of this central GAS virulence activator. In GAS, Mga activates expression of virulence-associated and linked M and M-like proteins (Mrp, Enn) (54, 55). Like Mga, SezV proved to be a critical activator of SzM expression and to be required for SEZ virulence in mice. The SezV regulon was relatively small as *szM* was the only locus with >5-fold reduction in transcript levels in the absence of *sezV*. SezV does not appear to regulate expression of SzP, the other M-like protein in SEZ/SEE, suggesting that there are additional SEZ virulence regulators that have yet to be identified. SezV also appears to down regulate (directly or indirectly) the expression of several genes, including an operon coding for the synthesis of a *cne*-containing fimbrial Fim 1 pilus. It is tempting to speculate that the SezV coordinated induction of SzM expression and repression of pilus expression corresponds to a SEZ virulence program that facilitates the organism's dissemination from sites, such as the oropharynx, where it is a commensal, to sites such as the blood stream, lungs, and brain, where it is pathogenic. The presence of the predicted sensor kinase-like domain in SezV suggests the possibility that there may be host-derived stimuli that trigger SezV activity.

The key role of SzM in SEZ virulence was not known, but a SzM homologue in SEE (SeM) has been linked to virulence (16, 17). In addition, McLellan et al. (19) found that Spa, a

SzM homolog in M18 GAS, appears to play a more substantial role in pathogenicity than the canonical M18 protein. Similar to canonical M proteins, SzM proteins likely promote virulence through their immunomodulatory functions. Timoney and colleagues have demonstrated that SzM from SEZ strain NC78 can inhibit phagocytosis and bind to and modify the activity of components of the coagulation cascade such as plasminogen (9, 14). In addition to roles in pathogenicity, SzM/SeM/Spa proteins from SEZ, SEE and M18 GAS have found to be effective immunogens eliciting protective immune responses (9, 14, 16, 18). Our analysis of SzM/SeM/Spa sequences from ~300 strains revealed that the highest variability is present in the N-terminal regions, whereas the C-terminal coiled-coil domains are more conserved. There appears to be greater variability in SEZ vs SEE, consistent with the idea that SEZ as a whole is more genetically variable than SEE (Figure 2.9E).

Unexpectedly, SzM and its activator SezV answered the screen for mutants that were not bound by a MAb (F598) to PNAG. Two observations lend support to the idea that SzM is decorated by a PNAG-like carbohydrate. First, purified SzM was bound by F598, whereas F429, an isotype matched control antibody that includes identical constant regions as F598, bound with much lower affinity. Second periodate and dispersin B, reagents that cleave carbohydrates, ablated SzM's reactivity with MAb F598, but not to anti-SeM sera. Specific glycosylation of proteins with oligomers of PNAG (or a related oligosaccharide) has not been described, but O-glycosylation of a streptococcal adhesin has been reported (56). However, it is premature to conclude that SzM is decorated by a PNAG-related oligosaccharide. We did not identify N- or O-linked oligosaccharides bound to SzM by mass spectrometry and the transposon screen did not yield genes obviously implicated in synthesis of a polysaccharide. Regardless of whether F598 is recognizing an oligosaccharide-linked to SzM, our observation that this antibody binds to a

wide-range of SEZ, SEE and M18 GAS strains raises the possibility that immunity to PNAG may have therapeutic applications in animal and human diseases caused by streptococci bearing SzM-like M proteins.

Acknowledgements

We gratefully acknowledge Gunnar Lindahl for providing the anti-SeM sera and anti-M sera, and Michael Wessels for providing the GAS M18 strains (87-282, 9003, SS-77, SS-734, and SS-1272). We thank Gunnar Lindahl and Brigid Davis for insightful comments on the manuscript.

This study was supported by the National Natural Science Foundation of China (31772746, 31302093); National Key Research and Development Program (2016YFD0501607, 2018YFD0500506); Young Backbone Teachers Study Abroad Program of China Scholarship Council (201706855039); Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) (M. Z.); the National Institute of General Medical Sciences (NIGMS) grant T32GM007753 (J.D.D.); National Natural Science Foundation of China (31272581); National Key Research and Development Program (2017YFD0500203) (H. F.); and the National Institute of Allergy and Infectious Diseases (NIAID) grant R01-AI-043247 and HHMI (M.K.W).

We thank the Harvard Medical School Biopolymers Facility for assistance with whole genome sequencing, Maria Ericsson and the Harvard Medical School Electron Microscopy Facility for assistance with electron microscopy sample preparation and imaging, the Taplin Biological Mass Spectrometry Facility for mass spectrometry, and the National Center for Research Resources (a part of the NIH) for the support of the National Center for Functional Glycomics (NCFG) at Beth Israel Deaconess Medical Center (P41GM103694), which assisted with glycan mass spectrometry.

Materials and Methods

Bacterial strains and growth

Streptococcal bacteria were routinely grown aerobically in THY (Todd Hewitt Broth + 0.5% Yeast extract) media with shaking or THY agar at 37°C. For GAS strains, liquid cultures were grown aerobically without agitation at 37°C. Antibiotics, when used, were included at the following concentrations: Spectinomycin (Spc) 100 µg/mL, Kanamycin (Km) 300 µg/mL. *E. coli* bacteria were routinely grown in lysogeny broth (LB) media or agar. Antibiotics were used at the following concentrations: Carbenicillin (Carb) 100 µg/mL, Km 50 µg/mL.

For growth curves, a fresh colony was picked, inoculated in THY media, and grown to log-phase (4-6 h) at 37ºC. After harvesting by centrifugation, the supernatant was discarded and each pellet was resuspended in fresh THY to OD600 = 0.50, and then diluted 10x with the same medium to an OD600 ~ 0.05. 200 µL of culture was then added per well of a honeycomb plate, with 5 replicates for each sample. A Bioscreen C (Growth Curves USA) growth curve machine was used to measure the OD600 over 24h with measurements every 10 min.

Immunofluorescence Microscopy

Stationary phase bacteria were labeled with F598 or F429 conjugated to AF488 at room temperature for 4 h after being fixed in ice cold methanol for 1 min. Widefield microscopy was performed on a Nikon Eclipse Ti microscope. Slides were fixed in ProLong Diamond Antifade Mountant with DAPI (Invitrogen).

Transposon library construction

A transposon library was constructed in SEZ ATCC 35246 using the pMar4s transposon plasmids by modifying a method used for Tn library construction in GAS (33, 57). A detailed graphical depiction of the method can be found in Figure 2.2A. Briefly, 1 μg of purified pMar4s plasmid was electroporated into SEZ. Transformants were resuspended in 5 ml THY+10% sucrose recovery broth and cultured at 28°C for 4 h. Bacteria were spread on a THY plate with Km and Spc after dilution and incubated at 28°C for 3 days. More than 100 individual colonies were screened to verify that pMar4s was present as an independent plasmid (i.e. had not yet undergone integration or transposition). Each colony was streaked onto 6 different plates with various antibiotics and grown at 28°C for 3 days or overnight at 37°C as shown in Figure 2.2A. Colonies that were unable to grow on a THY+ Spc plate at 37°C were collected from Spc and Km THY plates, and frozen stocks were made (25% glycerol). After re-verifying the antibiotic resistance profile of the frozen stock strains, a final Tn library was constructed by spreading > 5.4×10^5 colonies from a frozen stock onto 245 cm^2 square plates (Corning) of THY media, and incubating at 37°C overnight. Colonies were scraped from the plate with THY broth and mixed, and after adding glycerol (25% final concentration) the library was frozen at -80°C.

FACS-based PNAG screen

Flow cytometry and FACS was performed on a Sony SH800S Cell Sorter. For routine measurement of antibody binding, live bacteria were labeled with MAb F598 or F429 conjugated to Alexa Fluor 488 (AF488). FACS enrichment of the SEZ transposon library was conducted as follows. An aliquot of the library was stained with AF488-F598 prior to loading on the FACS. Bacteria whose fluorescence intensity was below a specified threshold, the lowest ~10% (screen

A) and ~20% (screen B) of AF488-F598 labeled bacteria were collected. After enrichment, bacteria were outgrown overnight in either (A) liquid THY + 1% dextrose media or (B) solid THY plates prior to re-enrichment on the FACS using a similar threshold as the first round. For bacteria grown on solid THY, prior to FACS sorting, the bacteria were scraped from the plate and resuspended in THY media. With each round, the percentage of bacteria that were below the threshold increased. There were 2 rounds of selection for screen A and 3 for screen B. Bacteria from each round and the input were collected and genomic DNA was extracted for Tn-seq. Several individual bacterial colonies from the final round of each screen were also picked and stored as frozen stocks. Some of these non F598 binding strains were subjected to WGS, leading to bacterial strains with transposon insertions in RS01795, *sezV* (RS09465), and *szM* (RS09460).

Tn-seq analysis

Tn-seq library construction and data analysis was performed as previously described (58–60); briefly, genomic DNA was extracted, transposon junctions amplified (a custom primer, Himmer3outMar, was designed for the pMar4s transposon plasmid), sequencing was performed on an Illumina MiSeq, and data was analyzed using a modified ARTIST pipeline. Sequence reads were mapped onto the SEZ ATCC 35246 genome. Reads at each TA site were tallied and assigned to annotated genes. We sequenced the original Tn library as well as the input Tn library used for the FACS screen (a thawed aliquot of the original Tn library) and the output Tn libraries after FACS selection for absence of F598 binding. For each output sample, a multinomial distribution based on the distribution in the output sample was used to re-sample reads in the corresponding input sample multiple times and total reads in the simulated inputs was adjusted to match that of the respective output sample. A Mann-Whitney U (MWU) test was used to identify

genes that have significantly different insertion profiles and read distributions between the output and simulated input samples; note that only sites in which there were nonzero reads in either the simulated input and/or output were used, as has been previously described (58). The log2 fold change was also calculated for each gene based on the reads in the output and simulated input library. Genes were filtered based on their containing at least 5 TA sites with a representative mutant in the input library, a log2 fold change of at least 1, and MWU p-value of at least 0.05.

Whole Genome Sequencing

Genomic DNA was purified from stationary phase cultures of bacteria using the Masterpure Gram Positive DNA Purification Kit (Epicentre cat. MGP04100). DNA concentration was adjusted prior to library preparation and sequencing at the Biopolymers Facility at Harvard Medical School. Libraries were constructed using the Nextera XT kit (Illumina) and paired-end sequencing was performed on a MiSeq (Illumina) with dual-indexing and 2x75 bp reads. Reads were mapped to the SEZ ATCC 35246 reference genome using bwa (61). Nucleotide variants were called as previously described (60), using GATK (ver 3.5) (62), annotated using ANNOVAR (63), and filtered and compared using vcftools (64). For identification of transposon insertions in WGS reads, reads were processed as for Tn-seq analysis. Putative mutations that result in loss of MAb F598 binding were found by identifying mutations present in the strains that could not bind F598 and absent from strains that could bind F598. For newly sequenced GAS strains, libraries were prepared and sequenced as for SEZ strains. However, in this case, Illumina reads were trimmed with sickle (ver 1.33) (65) and assembled using SPADES (ver 3.11.1) (66). Quast (ver 5.0.2) (67) was used to determine the assembly statistics.

Strain construction

Single gene deletions in SEZ ATCC 35246 were constructed as previously described (42) using the temperature sensitive shuttle plasmid pSET4s (68). Upstream and downstream regions of each gene were cloned into pSET4s, which confers Spc resistance. The constructed plasmids were introduced into competent SEZ cells via electroporation using the following settings: voltage: 2500V, Capacitance: 25μF, Resistance: 200Ω. The transformed bacteria were grown at 37°C in THY broth + Spc to generate, single-crossovers. Double-crossover mutants were generated by repeatedly passaging the single-crossover strains at 28°C on THY without Spc. The gene deletions were verified by PCR and Sanger sequencing.

To construct complementation plasmids, the *sezV* and *szM* genes were amplified by PCR from SEZ ATCC 35246 genomic DNA and inserted into the pSET2 plasmid (69). For the *szM* gene, the constructed plasmid was amplified in *E. coli* DH5α before electroporation into SEZ. For the *sezV* gene, a dialyzed Gibson assembly production of pSET2 and *sezV* was used to directly transform SEZ. For complementation with RS00930, the RS00930 gene, containing a synonymous amino acid mutation, was cloned into pSET4s and used to restore the RS00930 gene in its native chromosomal locus.

PCR was routinely performed using Phusion High-Fidelity DNA Polymerase (NEB) and plasmids were constructed via Gibson assembly using the NEBuilder HiFi DNA Assembly Master Mix (NEB).

Electron Microscopy

Bacterial samples were submitted to the Electron Microscopy Facility at Harvard Medical School as colonies grown on solid media THY plates. Scanning EM (SEM) was performed on bacterial samples fixed via glutaraldehyde.

Quantitative Reverse Transcriptase PCR

RNA was isolated from 3 ml log-phase cultures of bacteria. After harvesting bacteria via centrifugation, 50 mg of 0.1 mm glass beads were added to the tube along with 1 ml TRIzol (Invitrogen). After homogenizing the mixture vigorously for 2 min, chloroform extraction was performed followed by purification using the RNEasy Kit (Qiagen). The RNA was subject to DNAase treatment and then reverse transcribed using SuperScript II RT (Invitrogen). Fast SYBR Green Master Mix was used for qPCR reactions, and all reactions were conducted with an ABI StepOne real-time PCR system.

RNA sequencing

RNA was isolated from log-phase cultures of bacteria as described above for qPCR.
RNA-seq library construction and sequencing was performed by Genewiz, Inc (www.genewiz.com). Briefly, RNA was depleted of rRNA using the Illumina RiboZero kit and sequencing libraries were prepared using the NEBNext Ultra kit. Paired-end sequencing of libraries was performed on a single lane of a HiSeq 4000 (Illumina) with single-indexing and 2x150 bp reads. Total reads per sample were ~40+ million, well above the number recommended for bacterial differential gene analysis given the genome size of SEZ (70). Reads were mapped to the SEZ ATCC 35246 genome using bwa-mem (61), and read counts per gene were generated

using the Rsubread package (ver. 1.34.6) (71) in the R computing environment. Differential gene analysis was performed using the DESeq2 package (ver. 1.24.0) (72) in R. Genes were filtered based on outputs of DESeq2, using a log2 fold change magnitude of 1 and an adjusted p-value of 0.05. A volcano plot was generated using the EnhancedVolcano package in R.

Animal Experiments

Female C57Bl/6J mice aged 6-8 weeks obtained from the Jackson Laboratory were used for all experiments. Animal experiments were performed as previously described (42), using protocols approved by the BWH IACUC. Mouse experiments were conducted according to the recommendations of the National Institutes of Health Guide for the Care and Use of Laboratory Animals. Adult mice were euthanized via isoflurane inhalation and subsequent cervical dislocation. IV infections were performed via tail-vein injection of $1x10^7$ cfu of log-phase bacteria. For determination of tissue burdens, 24 hours after inoculation, the animals were sacrificed, dissected, and tissues were homogenized, and serial dilutions of the homogenates were plated on THY media to enumerate bacterial cfu. For bacterial burden in the blood, prior to sacking, 20 uL of blood was drawn and serial dilutions were plated on THY media.

Data from animal experiments was analyzed in Prism (ver 8) (GraphPad). The Mantel-Cox was used for statistical analyses of the Kaplan-Meier survival curves and the Kruskal-Wallis was with Dunn's multiple-comparison test used comparing the tissue cfu burdens.

Western Blots

LDS loading dye (NEB) was added to samples, which were then boiled at 95°C for 10 min prior to running on SDS-PAGE gels. Either Novex Sharp Pre-stained Protein Standard (Invitrogen) or

SeeBlue Pre-stained Protein Standard (Invitrogen) were used as molecular weight markers. Transfer was performed using the iBlot2 (Thermo Fisher Scientific) system. Membranes were blocked in 1% BSA, 5% powdered milk prior to blotting with primary antibodies. Antibodies were used at the following concentrations: F598 (13 μg/ml), F429 (13 μg/ml)), anti-SeM sera (1:1000), anti-M sera (1:1000), HRP conjugated goat anti-rabbit IgG antibody (1:10000, ab6858, Abcam), HRP conjugated goat anti-human IgG antibody (1:10000, A4914, Sigma). The anti-SeM and anti-M sera were both gifts from Gunnar Lindahl. The anti-SeM sera was raised against the full length SeM protein. The anti-M sera (13) was raised against the C-terminal BCW-region, which contains the B repeats, the conserved C repeat, and the conserved W (wall-spanning) region of the *S. pyogenes* M5 protein. SuperSignal Chemiluminescent Substrates (Thermo Fisher Scientific) was used as the HRP substrates. Gels and western blots were imaged with a ChemiDoc Imaging Systems (Bio-Rad).

Immunoprecipitation

A 50 mL of overnight cultured bacteria was harvested by centrifugation at 16,000g for 2 min. After discarding the supernatant and re-suspending the pellet in 1ml 0.5 M EDTA, the samples were boiled for 5 min to lyse the cells and then centrifuged at ~ 20,000 g for 5 min and the resulting supernatants was used as the bacterial lysates. The lysates were filtered through an Amico Ultra-4 10K centrifugal filter device to remove EDTA, and mixed with PBS-T (PBS pH7.4 with 0.01% Tween 20). Dynabeads Protein G (Invitrogen) were used for immunoprecipitation. Briefly, 5 μg of F598 or 5 μl anti-M serum were diluted in 200 μl PBS-T and rotated for 30 min with 50 μl magnetic beads at room temperature. The beads were washed once with 200 μl PBS-T. After removing the supernatant with a magnet rack and adding 1ml of

bacterial lysate with protease inhibitors, the mixture was incubated with rotation for 1 h at room temperature or overnight at 4°C to allow the antigen to bind to the magnetic bead-Ab complex. Beads were washed 3 times with PBS-T, and then eluted by adding 20 µl of 50 mM glycine pH2.8 and incubating at 70°C for 10 min.

Periodate & Dispersin B Treatments

Bacterial lysates or purified PNAG polysaccharide were mixed with 0.4 M periodate dissolved in PBS at 1:1 (v/v) and incubated at 37°C for 1 h. The final concentration of Dispersin B used for bacterial lysate or PNAG polysaccharide treatment was 500 µg/ml, and samples were incubated at 37°C for 24 h.

Computational analyses

The nucleotide sequence of the *sezV-szM* locus from ATCC 35246 plus additional flanking sequence was used as a query in blastn searches of the NCBI complete and WGS prokaryotic genomes to identify strains of streptococci and related genera that contain the two genes *szM/seM/spa* and *sezV*. Thresholds for identifying homologs were an alignment of >200 nucleotides for *szM* and an alignment of >300 nucleotides for *sezV*. These settings yielded only one homolog in each strain that contained a homolog. SzM homologs were not identified in some strains using more stringent settings; these strains were subsequently identified as containing class II or III *szM* homologs. Protein sequences were aligned using MUSCLE (ver 3.8.31) (73). Sequence alignments were visualized in Seaview (ver 4.7) (74) and Genome Workbench (ver 3.0.0) (NCBI).

Protein sequences were clustered with MMSeqs2 (29, 75), using the normal/slow sensitive settings. For identification of representative alleles, 80% minimum sequence identity and 80% minimum alignment coverage was used; for identification of distinct classes of *szM/seM/spa*, the thresholds were both adjusted to 50%. Similarity of amino acids across a multiple sequence alignment was conducted using plotcon (76). Similarity (y-axis) represents the sequence conservation over a set of adjacent residues. The similarity at one position is calculated as the average of all possible pairwise substitution scores (taken from a similarity matrix) of the residues at the position. The average of position similarities over a set of adjacent residues is graphed in Figure 2.8C and 2.10A.

Distance between genes *szM/seM/spa* and *sezV* was determined as the distance between the first base of either gene that occurred first in the strain's genome and the first base of the second gene. Orientation of transcription of the *szM/seM/spa* and *sezV* genes was determined based on genomic annotations and subsequently compared, taking into account the order in which the genes occurred in the genome.

Identification of the loci in newly sequenced and assembled streptoccocal strains was performed using blastn from blast+ (ver 2.7.1) (NCBI). Open reading frames were identified using ORFfinder (NCBI). *emm* (M protein) typing was performed by identifying the annotated M protein in the strain and the CDC *emm* blast server (Streptococci Group A Subtyping Request Form Blast 2.0 Server; https://www2a.cdc.gov/ncidod/biotech/strepblast.asp). For newly sequenced GAS strains, the M protein was identified using blast+ (tblastn), the top hit was chosen, and the sequence was then input into the CDC *emm* blast server to type the strain.

M18 strains were identified among the GAS strains using tblastn. The M protein from MGAS 8232 was used as the query sequence to identify additional M18 strains. Thresholds for

e-value and percent identity were lowered until the aligned M protein sequence began to be mapped to a different *emm* type, which was determined manually by inputting the M protein sequences into the CDC *emm* blast server. Based on the final thresholds, a total of 15 M18 strains were identified among the GAS strains analyzed.

Mass spectrometry

Mass spectrometry of immunoprecipitated and gel-purified SzM was carried out at the Taplin Biological Mass Spectrometry Facility, Harvard Medical School. In-gel trypsin digestion was performed prior to mass spectrometry analysis. Analysis of O-linked and N-linked glycans on SzM was performed at the National Center for Functional Glycomics. Briefly, N- and O- linked glycans were obtain from SzM and fetuin, a control glycosylated protein, by treating proteins with PNGaseF, to obtain the N-link glycans, and then subsequently with sodium borohydride ($NaBH_4$), to obtain the O-linked glycans.

Data availability

Illumina reads for the WGS and RNA-seq experiments have been uploaded to the NCBI Short Read Archive (SRA) with the following BioSample accession numbers: SAMN12770224-25, SAMN12772038-42, and SAMN12784884-89.

References

1. **Fulde M, Valentin-Weigand P**. 2013. Epidemiology and pathogenicity of zoonotic streptococci. *Curr Top Microbiol Immunol* **368**:49–81.

2. **Lindahl S, Aspán A, Båverud V, Ljung H, Paillot R, Pringle J, Söderlund R, Wright NL, Waller AS**. 2012. A clonal outbreak of upper respiratory disease in horses caused by Streptococcus equi subsp. zooepidemicus. *J Equine Vet Sci* **32**:S24.

3. **Feng Z, Hu J, Feng, Z.G. HJS**. 1977. Outbreak of swine streptococcosis in Sichan province and identification of pathogen. *Anim Husb Vet Med Lett* **2**:7–12.

4. **Liu PH, Shun SF, WAng YK, Zhang SH**. 2001. Identification of swine Streptococcus isolates in Shanghai. *Chin J Vet Med* **21**:42–46.

5. **Boyle AG, Timoney JF, Newton JR, Hines MT, Waller AS, Buchanan BR**. 2018. Streptococcus equi Infections in Horses: Guidelines for Treatment, Control, and Prevention of Strangles-Revised Consensus Statement. *J Vet Intern Med* **32**:633–647.

6. **Mori N, Guevara JM, Tilley DH, Briceno JA, Zunt JR, Montano SM**. 2013. Streptococcus equi subsp. zooepidemicus meningitis in Peru. *J Med Microbiol* **62**:335–7.

7. **Eyre DW, Kenkre JS, Bowler ICJW, McBride SJ**. 2010. Streptococcus equi subspecies zooepidemicus meningitis--a case report and review of the literature. *Eur J Clin Microbiol Infect Dis* **29**:1459–63.

8. **Boschwitz JS, Timoney JF**. 1994. Characterization of the antiphagocytic activity of equine fibrinogen for Streptococcus equi subsp. Equi. *Microb Pathog* **17**:121–129.

9. **Velineni S, Timoney JF**. 2013. Characterization and protective immunogenicity of the szm protein of Streptococcus zooepidemicus NC78 from a clonal outbreak of equine respiratory disease. *Clin Vaccine Immunol* **20**:1181–1188.

10. **Timoney JF, Artiushin SC, Boschwitz JS**. 1997. Comparison of the sequences and functions of Streptococcus equi M-like proteins SeM and SzPSe. *Infect Immun* **65**:3600–3605.

11. **Frost HR, Sanderson-Smith M, Walker M, Botteaux A, Smeesters PR**. 2018. Group A streptococcal M-like proteins: From pathogenesis to vaccine potential. *FEMS Microbiol Rev* **42**:193–204.

12. **Fischetti VA**. 1989. Streptococcal M protein: molecular design and biological behavior. *Clin Microbiol Rev* **2**:285–314.

13. **Lannergård J, Gustafsson MCU, Waldemarsson J, Norrby-Teglund A, Stålhammar-Carlemalm M, Lindahl G**. 2011. The Hypervariable Region of Streptococcus pyogenes M Protein Escapes Antibody Attack by Antigenic Variation and Weak Immunogenicity. *Cell Host Microbe* **10**:147–157.

14. **Velineni S, Timoney JF**. 2013. Identification of novel immunoreactive proteins of streptococcus zooepidemicus with potential as vaccine components. *Vaccine* **31**:4129–4135.

15. **Hong-Jie F, Fu-yu T, Ying M, Cheng-ping L**. 2009. Virulence and antigenicity of the szp-gene deleted Streptococcus equi ssp. zooepidemicus mutant in mice. *Vaccine* **27**:56–61.

16. **Meehan M, Nowlan P, Owen P**. 1998. Affinity purification and characterization of a fibrinogen-binding protein complex which protects mice against lethal challenge with Streptococcus equi subsp. equi. *Microbiology* **144**:993–1003.

17. **Meehan M, Owen P, Lynagh Y, Woods C**. 2001. The fibrinogen-binding protein (FgBP) of Streptococcus equi subsp. equi additionally binds IgG and contributes to virulence in a mouse model. *Microbiology* **147**:3311–3322.

18. **Dale JB, Chiang EY, Liu S, Courtney HS, Hasty DL**. 1999. New protective antigen of group A streptococci. *J Clin Invest* **103**:1261–1268.

19. **McLellan DGJ, Chiang EY, Courtney HS, Hasty DL, Wei SC, Hu MC, Walls MA, Bloom JJ, Dale JB**. 2001. Spa Contributes to the Virulence of Type 18 Group A Streptococci. *Infect Immun* **69**:2943–2949.

20. **Simpson WJ, Cleary PP**. 1987. Expression of M type 12 protein by a group A streptococcus exhibits phaselike variation: evidence for coregulation of colony opacity determinants and M protein. *Infect Immun* **55**:2448–55.

21. **Simpson WJ, LaPenta D, Chen C, Cleary PP**. 1990. Coregulation of type 12 M protein and streptococcal C5a peptidase genes in group A streptococci: evidence for a virulence regulon controlled by the virR locus. *J Bacteriol* **172**:696–700.

22. **Sabharwal H, Michon F, Nelson D, Dong W, Fuchs K, Manjarrez RC, Sarkar A, Uitz C, Viteri-Jackson A, Suarez RSR, Blake M, Zabriskie JB**. 2006. Group A Streptococcus (GAS) Carbohydrate as an Immunogen for Protection against GAS Infection. *J Infect Dis* **193**:129–135.

23. **van Sorge NM, Cole JN, Kuipers K, Henningham A, Aziz RK, Kasirer-Friede A, Lin L, Berends ETM, Davies MR, Dougan G, Zhang F, Dahesh S, Shaw L, Gin J, Cunningham M, Merriman JA, Hütter J, Lepenies B, Rooijakkers SHM, Malley R, Walker MJ, Shattil SJ, Schlievert PM, Choudhury B, Nizet V**. 2014. The Classical Lancefield Antigen of Group A Streptococcus Is a Virulence Determinant with Implications for Vaccine Design. *Cell Host Microbe* **15**:729–740.

24. **Henningham A, Davies MR, Uchiyama S, van Sorge NM, Lund S, Chen KT, Walker MJ, Cole JN, Nizet V**. 2018. Virulence Role of the GlcNAc Side Chain of the Lancefield Cell Wall Carbohydrate Antigen in Non-M1-Serotype Group A Streptococcus. *MBio* **9**:e02294-17.

25. **Wei Z, Fu Q, Chen Y, Cong P, Xiao S, Mo D, He Z, Liu X**. 2012. The capsule of Streptococcus equi ssp. zooepidemicus is a target for attenuation in vaccine development. *Vaccine* **30**:4670–5.

26. **Stollerman GH, Dale JB**. 2008. The Importance of the Group A Streptococcus Capsule in the Pathogenesis of Human Infections: A Historical Perspective. *Clin Infect Dis* **46**:1038–1045.

27. **Cywes-Bentley C, Skurnik D, Zaidi T, Roux D, DeOliveira RB, Garrett WS, Lu X, O'Malley J, Kinzel K, Zaidi T, Rey A, Perrin C, Fichorova RN, Kayatani AKK, Maira-Litran T, Gening ML, Tsvetkov YE, Nifantiev NE, Bakaletz LO, Pelton SI, Golenbock DT, Pier GB**. 2013. Antibody to a conserved antigenic target is protective against diverse prokaryotic and eukaryotic pathogens. *Proc Natl Acad Sci* **110**:E2209–E2218.

28. **Lu X, Skurnik D, Pozzi C, Roux D, Cywes-Bentley C, Ritchie JM, Munera D, Gening ML, Tsvetkov YE, Nifantiev NE, Waldor MK, Pier GB**. 2014. A Poly-N-Acetylglucosamine-Shiga Toxin Broad-Spectrum Conjugate Vaccine for Shiga Toxin-Producing Escherichia coli. *MBio* **5**:1–9.

29. **Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V**. 2018. A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J Mol Biol* **430**:2237–2243.

30. **Söding J**. 2005. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**:951–60.

31. **Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE**. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* **10**:845–858.

32. **Pier GB, Boyer D, Preston M, Coleman FT, Llosa N, Mueschenborn-Koglin S, Theilacker C, Goldenberg H, Uchin J, Priebe GP, Grout M, Posner M, Cavacini L**. 2004. Human Monoclonal Antibodies to Pseudomonas aeruginosa Alginate That Protect against Infection by Both Mucoid and Nonmucoid Strains. *J Immunol* **173**:5671–5678.

33. **Liu R, Zhang P, Su Y, Lin H, Zhang H, Yu L, Ma Z, Fan H**. 2016. A novel suicide shuttle plasmid for Streptococcus suis serotype 2 and Streptococcus equi ssp. zooepidemicus gene mutation. *Sci Rep* **6**:1–9.

34. **Xu B, Pei X, Su Y, Ma Z, Fan H**. 2016. Capsule of Streptococcus equi subsp. zooepidemicus hampers the adherence and invasion of epithelial and endothelial cells and is attenuated during internalization. *FEMS Microbiol Lett* **363**:fnw164.

35. **Swanson J, Hsu KC, Gotschlich EC**. 1969. Electron microscopic studies on streptococci: I. M Antigen. *J Exp Med* **130**:1063–1091.

36. **Phillips GN, Flicker PF, Cohen C, Manjula BN, Fischetti VA**. 1981. Streptococcal M protein: alpha-helical coiled-coil structure and arrangement on the cell surface. *Proc Natl Acad Sci* **78**:4689–4693.

37. **Lannergård J, Frykberg L, Guss B**. 2003. CNE, a collagen-binding protein of Streptococcus equi. *FEMS Microbiol Lett* **222**:69–74.

38. **Steward KF, Robinson C, Maskell DJ, Nenci C, Waller AS**. 2017. Investigation of the Fim1 putative pilus locus of Streptococcus equi subspecies equi. *Microbiology* **163**:1217–1228.

39. **Flock M, Karlström A, Lannergård J, Guss B, Flock J-I**. 2006. Protective effect of vaccination with recombinant proteins from Streptococcus equi subspecies equi in a strangles model in the mouse. *Vaccine* **24**:4144–51.

40. **Timoney JF, Qin A, Muthupalani S, Artiushin S**. 2007. Vaccine potential of novel surface exposed and secreted proteins of Streptococcus equi. *Vaccine* **25**:5583–90.

41. **Guss B, Flock M, Frykberg L, Waller AS, Robinson C, Smith KC, Flock J-I**. 2009. Getting to grips with strangles: an effective multi-component recombinant vaccine for the protection of horses from Streptococcus equi infection. *PLoS Pathog* **5**:e1000584.

42. **Ma Z, Peng J, Yu D, Park JS, Lin H, Xu B, Lu C, Fan H, Waldor MK**. 2019. A streptococcal Fic domain-containing protein disrupts blood-brain barrier integrity by activating moesin in endothelial cells. *PLOS Pathog* **15**:e1007737.

43. **Ramasubbu N, Thomas LM, Ragunath C, Kaplan JB**. 2005. Structural analysis of dispersin B, a biofilm-releasing glycoside hydrolase from the periodontopathogen Actinobacillus actinomycetemcomitans. *J Mol Biol* **349**:475–86.

44. **Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, Li W-J**. 2014. Comparative Genomics of the Bacterial Genus Streptococcus Illuminates Evolutionary Implications of Species Groups. *PLoS One* **9**:e101229.

45. **Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF**. 2016. A new view of the tree of life. *Nat Microbiol* **1**:16048.

46. **Facklam R**. 2002. What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. *Clin Microbiol Rev* **15**:613–630.

47. **Okura M, Osaki M, Nomoto R, Arai S, Osawa R, Sekizaki T, Takamatsu D**. 2016. Current Taxonomical Situation of Streptococcus suis. *Pathogens* **5**:45.

48. **Meehan M, Lewis MJ, Byrne C, O'Hare D, Woof JM, Owen P**. 2009. Localization of the equine IgG-binding domain in the fibrinogen-binding protein (FgBP) of Streptococcus equi subsp. equi. *Microbiology* **155**:2583–2592.

49. **Kelly C, Bugg M, Robinson C, Mitchell Z, Davis-Poynter N, Newton JR, Jolley KA, Maiden MCJ, Waller AS**. 2006. Sequence variation of the SeM gene of Streptococcus equi allows discrimination of the source of strangles outbreaks. *J Clin Microbiol* **44**:480–486.

50. **Delorenzi M, Speed T**. 2002. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* **18**:617–625.

51. **Harris SR, Parkhill J, Holden MTG, Robinson C, Steward KF, Webb KS, Paillot R, Waller AS**. 2015. Genome specialization and decay of the strangles pathogen, Streptococcus equi, is driven by persistent infection. *Genome Res* **25**:1360–1371.

52. **Anzai T, Kuwamoto Y, Wada R, Sugita S, Kakuda T, Takai S, Higuchi T, Timoney JF**. 2005. Variation in the N-terminal region of an M-like protein of Streptococcus equi and evaluation of its potential as a tool in epidemiologic studies. *Am J Vet Res* **66**:2167–2171.

53. **Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS, Sylva GL,**

Sturdevant DE, Ricklefs SM, Porcella SF, Parkins LD, Beres SB, Campbell DS, Smith TM, Zhang Q, Kapur V, Daly JA, Veasy LG, Musser JM. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. *Proc Natl Acad Sci* **99**:4668–4673.

54. **Hondorp ER, McIver KS**. 2007. The Mga virulence regulon: Infection where the grass is greener. *Mol Microbiol* **66**:1056–1065.

55. **Hondorp ER, Hou SC, Hause LL, Gera K, Lee CE, Mciver KS**. 2013. PTS phosphorylation of Mga modulates regulon expression and virulence in the group A streptococcus. *Mol Microbiol* **88**:1176–1193.

56. **Zhou M, Wu H**. 2009. Glycosylation and biogenesis of a family of serine-rich bacterial adhesins. *Microbiology* **155**:317–27.

57. **Le Breton Y, McIver KS**. 2013. Genetic manipulation of Streptococcus pyogenes (the Group A Streptococcus, GAS). *Curr Protoc Microbiol* **30**:9D.3.1-9D.3.29.

58. **Hubbard TP, Chao MC, Abel S, Blondel CJ, Abel zur Wiesch P, Zhou X, Davis BM, Waldor MK**. 2016. Genetic analysis of *Vibrio parahaemolyticus* intestinal colonization. *Proc Natl Acad Sci* **113**:6283–6288.

59. **Kimura S, Waldor MK**. 2019. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *Proc Natl Acad Sci* **116**:1394–1403.

60. **Hubbard TP, Billings G, Dörr T, Sit B, Warr AR, Kuehl CJ, Kim M, Delgado F, Mekalanos JJ, Lewnard JA, Waldor MK**. 2018. A live vaccine rapidly protects against cholera in an infant rabbit model. *Sci Transl Med* **10**:1–11.

61. **Li H, Durbin R**. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.

62. **DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ**. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**:491–498.

63. **Wang K, Li M, Hakonarson H**. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**:e164–e164.

64. **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R**. 2011. The variant call format and VCFtools. *Bioinformatics* **27**:2156–2158.

65. **Joshi N, Fass J**. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.

66. **Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Prjibelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, Stepanauskas R, Clingenpeel SR, Woyke T, Mclean JS, Lasken R, Tesler G, Alekseyev MA, Pevzner PA**. 2013.

Assembling Single-Cell Genomes and Mini-Metagenomes From Chimeric MDA Products. *J Comput Biol* **20**:714–737.

67. **Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A**. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **34**:i142–i150.

68. **Takamatsu D, Osaki M, Sekizaki T**. 2001. Thermosensitive Suicide Vectors for Gene Replacement in Streptococcus suis. *Plasmid* **46**:140–148.

69. **Takamatsu D, Osaki M, Sekizaki T**. 2001. Construction and Characterization of Streptococcus suis–Escherichia coli Shuttle Cloning Vectors. *Plasmid* **45**:101–113.

70. **Haas BJ, Chin M, Nusbaum C, Birren BW, Livny J**. 2012. How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes? *BMC Genomics* **13**:734.

71. **Liao Y, Smyth GK, Shi W**. 2019. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**:e47–e47.

72. **Love MI, Huber W, Anders S**. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550.

73. **Edgar RC**. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**:1792–1797.

74. **Gouy M, Guindon S, Gascuel O**. 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**:221–224.

75. **Steinegger M, Söding J**. 2017. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* **35**:1026–1028.

76. **Rice P, Longden I, Bleasby A**. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**:276–7.

Chapter 3

An oral inoculation infant rabbit model for *Shigella* infection

Adapted from a manuscript in mBio (January 21, 2020)

An oral inoculation infant rabbit model for *Shigella* infection

Carole J. Kuehl[*1,2], Jonathan D. D'Gama[*1,2], Alyson R. Warr[1,2], and Matthew K. Waldor[#1,2,3]

1.Division of Infectious Diseases, Brigham & Women's Hospital, Boston MA, 02115

2. Department of Microbiology, Harvard Medical School, Boston, MA, 02115

3. Howard Hughes Medical Institute, Boston, MA 02115

* equal contribution

for correspondence: mwaldor@research.bwh.harvard.edu

Running title: Oral inoculation model of shigellosis

Author Contributions

This chapter was adapted from a publication in *mBio* (January 2020). Jonathan D'Gama contributed to designing the study, performing experiments, analyzing the data, and wrote the manuscript. Carole Kuehl contributed to designing the study, performing experiments, analyzing the data, and edited the manuscript. Alyson Warr contributed to performing experiments and edited the manuscript. Matthew Waldor contributed to designing the study and edited the manuscript. Matthew Waldor supervised the study.

Abstract

Shigella species cause diarrheal disease globally. Shigellosis is typically characterized by bloody stools and colitis with mucosal damage and is the leading bacterial cause of diarrheal death worldwide. Following oral ingestion, the pathogen invades and replicates within the colonic epithelium through mechanisms that rely on its type III secretion system (T3SS). Currently, oral infection-based small animal models to study the pathogenesis of shigellosis are lacking. Here, we found that oro-gastric inoculation of infant rabbits with *S. flexneri* resulted in diarrhea and colonic pathology resembling that found in human shigellosis. Fasting animals prior to *S. flexneri* inoculation increased the frequency of disease. The pathogen colonized the colon, where both luminal and intraepithelial foci were observed. The intraepithelial foci likely arise through *S. flexneri* spreading from cell-to-cell. Robust *S. flexneri* intestinal colonization, invasion of the colonic epithelium, and epithelial sloughing all required the T3SS as well as IcsA, a factor required for bacterial spreading and adhesion in vitro. Expression of the proinflammatory chemokine IL-8, detected with in situ mRNA labeling, was higher in animals infected with wild-type *S. flexneri* versus mutant strains deficient in *icsA* or T3SS, suggesting that epithelial invasion promotes expression of this chemokine. Collectively, our findings suggest that oral infection of infant rabbits offers a useful experimental model for studies of the pathogenesis of shigellosis and for testing of new therapeutics.

Importance

Shigella species are the leading bacterial cause of diarrheal death globally. The pathogen causes bacillary dysentery, a bloody diarrheal disease characterized by damage to the colonic mucosa and is usually spread through the fecal-oral route. Small animal models of shigellosis that rely on the oral route of infection are lacking. Here, we found that oro-gastric inoculation of infant rabbits with *S. flexneri* led to a diarrheal disease and colonic pathology reminiscent of human shigellosis. Diarrhea, intestinal colonization and pathology in this model were dependent on the *S. flexneri* type III secretion system and IcsA, canonical *Shigella* virulence factors. Thus, oral infection of infant rabbits offers a feasible model to study the pathogenesis of shigellosis and to develop and test new therapeutics.

Introduction

Shigella species are Gram-negative, rod-shaped bacteria that cause bacillary dysentery, a severe and often bloody diarrheal disease characterized by inflammatory colitis that can be life-threatening (1). This enteric pathogen, which is spread by the fecal-oral route between humans, does not have an animal reservoir or vector (1). Annually, *Shigella* infections cause tens of millions of diarrhea cases and ~200,000 deaths (2, 3). It is likely the leading cause of diarrheal mortality worldwide in individuals older than 5 years (2, 3). Most *Shigella* infections are attributable to *S. flexneri*, one of the four *Shigella* species, although in developed nations the prevalence of *S. sonnei* is higher (4–7).

The pathogen primarily causes colonic pathology that usually includes mucosal ulceration and erosion due to sloughing of epithelial cells, and is typically characterized by acute inflammation, with recruitment of neutrophils and plasma cells, congestion of blood vessels, distorted crypt architecture, and hemorrhage (8, 9). While inflammatory responses to *Shigella* invasion of colonic epithelial cells were thought to be the underlying cause of epithelial cell destruction and hemorrhage, recent evidence suggests that pathogen-mediated destruction of epithelial cells also plays a role in the development of pathology (10).

Shigella pathogenesis is attributable to a multifaceted set of virulence factors that enable the pathogen to invade and proliferate within the cytoplasm of colonic epithelial cells and evade host immune responses. The pathogen can also infect and rapidly kill macrophages (11). Most known virulence factors are encoded on a large (>200 kbp) virulence plasmid, which is required for *Shigella* pathogenicity (12–14). Key virulence determinants include a type III secretion system (T3SS) and its suite of protein effectors that are injected into host cells (15), and the cell surface protein IcsA, which directs polymerization of host actin and enables intracellular

movement (16, 17). The force generated by intracellular actin-based motility allows the pathogen to form membrane protrusions into neighboring uninfected cells, which the pathogen subsequently enters. Cell-to-cell spread is thought to promote pathogen proliferation in the intestine and evasion of immune cells (11). The ~30 T3SS effector proteins encoded by *Shigella* strains have varied functions, but primary roles include facilitating invasion of epithelial cells and suppression of host immune responses including cytokine production.

Among animals used to model infection, only non-human primates develop shigellosis from oral inoculation (18); however, the expense of this model limits its utility. Several small animal models of *Shigella* infection have been developed, yet none capture all the features of natural human infection. Historically, the Sereny test was used to identify *Shigella* virulence factors required for induction of an inflammatory response (19); however, this ocular model bears little resemblance to natural infection. The adult rabbit ligated ileal loop model has proven useful for the study of *Shigella* virulence factors (20). However, this model bypasses the normal route of infection and challenges the small intestine, which is not the primary site of pathology in human infections. Intra-rectal guinea pig infection induces colonic pathology and bloody diarrhea (21), and has been used to dissect the contribution of *Shigella* and host factors in several aspects of pathogenesis (22–24). Adult mice, the most genetically tractable mammalian model organism, are recalcitrant to developing disease when inoculated orally (25). As an alternative to oral inoculation, an adult mouse pulmonary model of *Shigella* infection involving intranasal inoculation of mice with *Shigella* has been developed (26); this model provides a platform to investigate host immune responses and vaccine candidates (27, 28), and has improved understanding of the innate immune response to *Shigella* infection (29). In contrast to adult mice, infant mice are susceptible to oral inoculation within a narrow window of time after birth, and

inoculation with a high dose of the pathogen leads to mortality within a few hours; however, pathology is evident in the proximal small intestine rather than the distal small intestine or colon, and infected suckling mice do not develop diarrhea or intestinal fluid accumulation (30, 31). A zebrafish larvae model, in which the *Shigella* T3SS is required for pathogen virulence, has been useful for characterizing cell mediated innate immune responses to *Shigella* due to the ability to image infection in vivo (32, 33). Recently, an infant rabbit intra-rectal inoculation model was described in which animals develop disease and rectal pathology reminiscent of natural infections (10). The lack of a robust, oral inoculation-based, small animal model of shigellosis has limited understanding of the role of virulence factors in pathogenesis, particularly of the importance of such factors for enabling intestinal colonization and for generating pathology and clinical signs.

Here we found that oro-gastric inoculation of infant rabbits with *S. flexneri* results in severe disease resembling human shigellosis. Orally infected animals develop diarrhea and colonic pathology marked by damage to the epithelial cell layer and edema. Furthermore, the pathogen invaded and appeared to spread between colonic epithelial cells. We found that both the T3SS and IcsA were required for signs of disease, intestinal colonization and pathology. In addition, invasion of the pathogen into the epithelial cell layer was required for induction of host IL-8 expression. In situ mRNA labeling revealed that induction of IL-8 transcripts occurs primarily in cells adjacent to invaded epithelial cells, and not in the infected cells. Thus, our findings suggest that the oro-gastric infant rabbit model provides a powerful and accessible small animal model for further investigation of factors contributing to *Shigella* pathogenesis and for testing new therapeutics.

Figure 3.1. Clinical signs and gross pathology of infant rabbits following oro-gastric inoculation of *S. flexneri*.

A. Clinical signs in infant rabbits infected with *S. flexneri* or isogenic mutant strains. Statistical significance for development of diarrhea between the animals in the WT group and in each of the other groups was determined using a Fisher's exact test.

B-C. Hind regions of animals inoculated with the WT strain (B) or an uninfected animal (C). Arrows indicate liquid feces stuck on anus and hind paws.

D. Colons from animals inoculated with the WT strain (left) or of an uninfected animal (right). Arrowheads point to regions of liquid feces and arrows indicate solid fecal pellets.

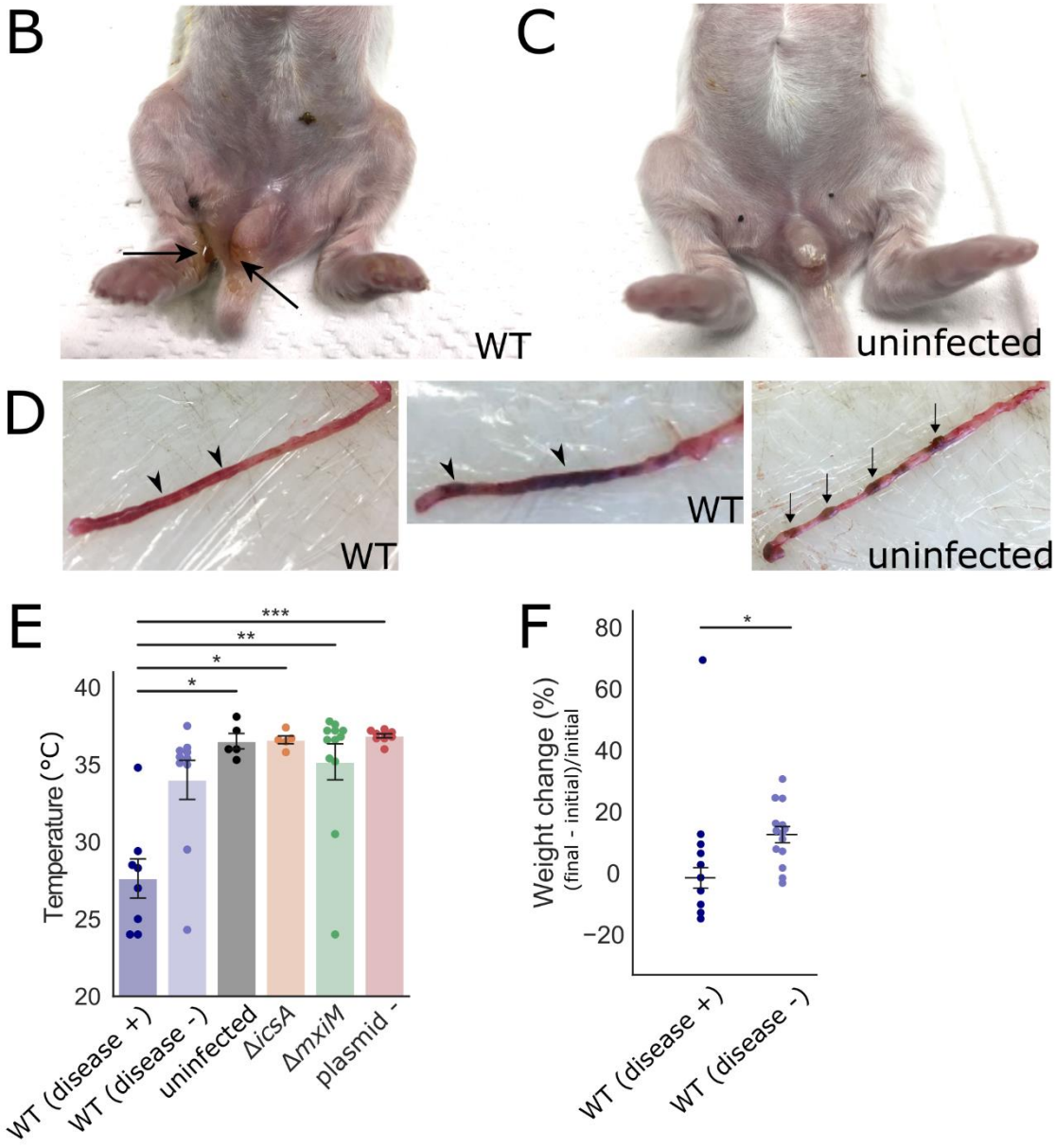
E. Body temperature of animals inoculated with the indicated strains 36 hpi or when they became moribund. Standard error of the mean values are superimposed. Disease +/- indicates whether or not animals developed diarrhea or became moribund early; all groups were compared to the WT (diarrhea +) group using a Kruskal-Wallis test with Dunn's multiple-comparison. p-values: <0.05, *; <0.01, **; <0.001, ***.

F. Percentage change in weight of infant rabbits infected with the WT strain, grouped by whether or not they developed disease (+/-). Percentage change in weight is calculated as difference between the final weight of the animal at 36 hpi or the last weight measurement taken when they became moribund (final) and the initial weight of the animal upon arrival in the animal facility (initial). Means and standard error of the mean values are superimposed. Groups were compared with a Mann-Whitney U test. p-values: <0.05, *.

Figure 3.1 (Continued)

A

Strain	WT	uninfected	Δ*icsA*	Δ*mxiM*	Plasmid -
Diarrhea	13/22 (59%)	0/8 (0%)	0/10 (0%)	0/12 (0%)	0/12 (0%)
Death	6/22 (27%)	0/8 (0%)	0/10 (0%)	0/12 (0%)	0/12 (0%)
P-value for diarrhea (vs. WT)	-	4.40E-03	1.60E-03	6.00E-04	6.00E-04



Results

Infant rabbits develop diarrhea following oro-gastric inoculation with *S. flexneri*

In previous work, we found that oro-gastric inoculation of infant rabbits with Enterohemorrhagic *Escherichia coli* (EHEC), *Vibrio cholerae*, and *V. parahaemolyticus* (34–36) leads to diarrheal diseases and pathologies that mimic their respective human counterparts. Here, we explored the suitability of oro-gastric inoculation of infant rabbits to model *Shigella* infection. *S. flexneri* 2a strain 2457T, a human isolate that is widely used in the research community as well as in challenge studies in humans (37), was used in this work. We utilized a streptomycin resistant derivative of this strain for infections to facilitate enumeration of pathogen colony forming units (cfu) in samples from the rabbit intestine. This strain, which contains a point mutation resulting in a K43R mutation in the small (30S) ribosomal subunit protein RpsL, retains the full virulence plasmid and grows as well as the parent strain.

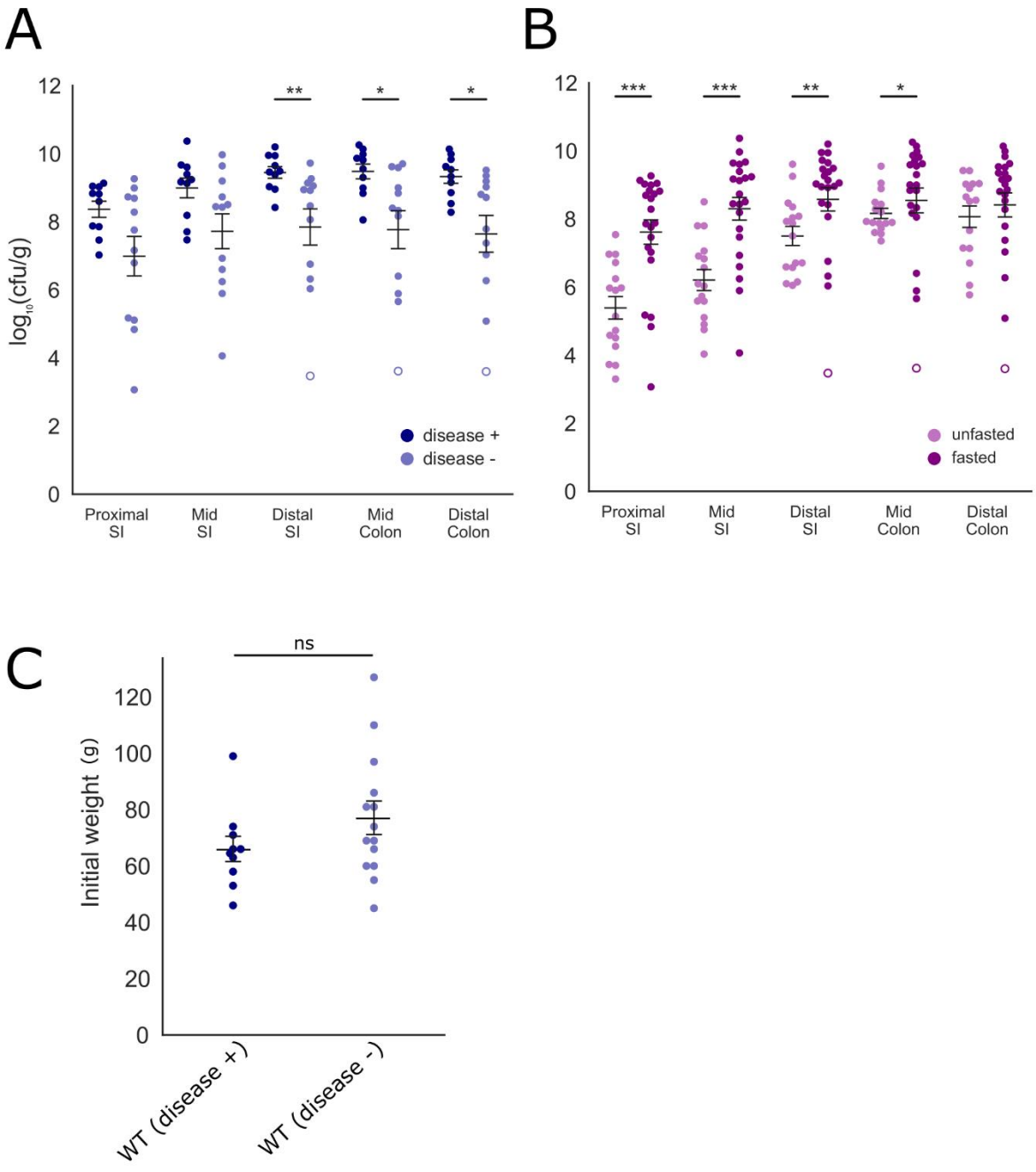
In order to investigate infant rabbits as a potential *Shigella* host, we orally inoculated two to three-day-old rabbits that were co-housed with their dam and then monitored for signs of disease. There was considerable variability in the development of diarrhea and colonization in initial studies using suckling rabbits fed ad libitum. Previous work using four-week-old rabbits suggested that a milk component could protect animals from disease by degrading the *Shigella* T3SS components (38, 39); consequently, additional experiments were performed with infant rabbits separated from their lactating dam for 24 hours prior to inoculation. Using this protocol, we obtained more reliable clinical disease and robust intestinal colonization. By 36 hours post infection (hpi), the majority (59%) of animals developed diarrhea, which was grossly visible as liquid fecal material adhering to the fur of the hind region of the rabbits (Figure 3.1A-C), and high levels of intestinal colonization; occasionally the diarrhea was frankly bloody. We chose the

Figure 3.2. Factors influencing development of diarrheal disease in infant rabbits after oro-gastric inoculation of *S. flexneri*.

A-B. Bacterial burden of *S. flexneri* in the indicated intestinal sections 36 hpi; SI = small intestine. Each point represents measurement from one rabbit. Data plotted as log transformed colony forming units (cfu) per gram of tissue. Means and standard error of the mean values are superimposed. Open circles represent the limit of detection of the assay and are shown for animals where no cfu were recovered. For each anatomical section, groups were compared with a Mann-Whitney U test. p-values: <0.05, *; <0.01, **; <0.001, ***. (A) 'disease +' refers to animals infected with the WT *S. flexneri* strain that developed disease (diarrhea or became moribund early), 'disease -' refers to animals infected with the WT *S. flexneri* strain that did not develop disease. (B) 'unfasted' refers to animals fed ad libitum prior to inoculation, 'fasted' refers to animals separated from dams for 24 hours prior to inoculation.

C. Initial body weights of infant rabbits inoculated with WT *S. flexneri*, grouped based on whether animal developed disease (diarrhea or early mortality). Means and standard error of the mean values are superimposed. Groups were compared with a Mann-Whitney U test.

Figure 3.2 (Continued)



36 hpi timepoint because from preliminary timecourse experiments we observed that all animals that were going to develop diarrhea developed disease by this timepoint and there was significant intestinal pathology at this time. Upon necropsy, the colon of infected animals was often bloody and contained liquid fecal material, in contrast to that of uninfected animals, which contained solid fecal pellets (Figure 3.1D). Furthermore, some infected rabbits (27%) succumbed to infection rapidly and became moribund prior to 36 hpi, though not all of these animals developed diarrhea (Figure 3.1A). Infected animals had highest bacterial burdens in the colon as well as the mid and distal small intestine (Figure 3.2AB & 3.3A). The development of disease was associated with higher pathogen burdens in the colon (Figure 3.2A). Separation of kits from the dam prior to inoculation led to a statistically significant elevation in intestinal colonization (Figure 3.2B).

Although not all *S. flexneri* inoculated animals developed signs of disease, infected rabbits that developed diarrhea or died early displayed additional disease signs. The animals that developed disease had significantly lower body temperatures than uninfected animals (8-9°C lower than uninfected, Figure 3.1E), and they had significantly smaller gains in body weight than infected animals without disease (-2% vs 13%) (Figure 3.1F) over the course of the experiment. Despite the relatively large intra- and inter-litter variation in body weight, with a constant pathogen dose per animal ($1x10^9$ colony forming units, cfu), a lower initial body weight did not appear to be a risk factor for the development of disease (Figure 3.2C).

Histopathologic examination of the intestines from infected rabbits revealed colonic pathology reminiscent of some of the features observed in infected human tissue, including substantial edema (Figure 3.3B) as well as sloughing of colonic epithelial cells (Figure 3.3C). In unusual cases, there was massive hemorrhage in the colonic tissue of infected rabbits (Figure

Figure 3.3. Intestinal colonization and colonic pathology in infant rabbits infected with *S. flexneri*.

A. Bacterial burden of *S. flexneri* in the indicated intestinal sections 36 hpi; SI = small intestine. Each point represents measurement from one rabbit. Data plotted as log transformed colony forming units (cfu) per gram of tissue; mean values are indicated with bars. Open circles represent the limit of detection of the assay and are shown for animals where no cfu were recovered.

B-D. Representative haematoxylin and eosin-stained colonic sections from infected animals (B,C) 36 hpi or uninfected animals (D). Arrowheads in (B) indicate areas of edema in the lamina propria. Arrowheads in (C) indicate areas where the epithelial cell layer is absent. Arrows in (D) point to the intact layer of epithelial cells seen in the colon. The dashed lines indicate the presence (inset, D) or absence (inset, C) of the epithelial cell layer. Scale bar is 100 μm.

Figure 3.3 (Continued)

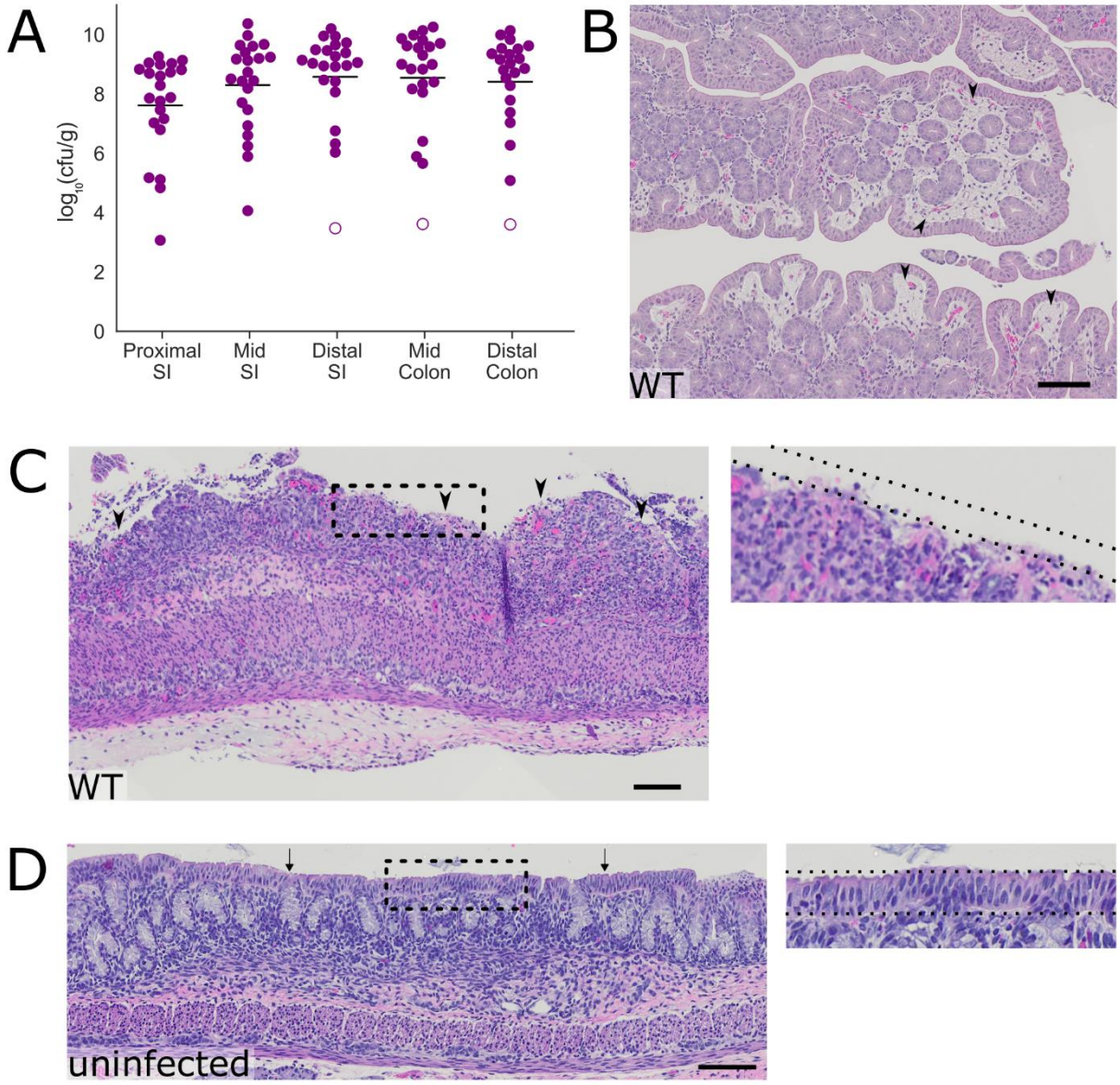
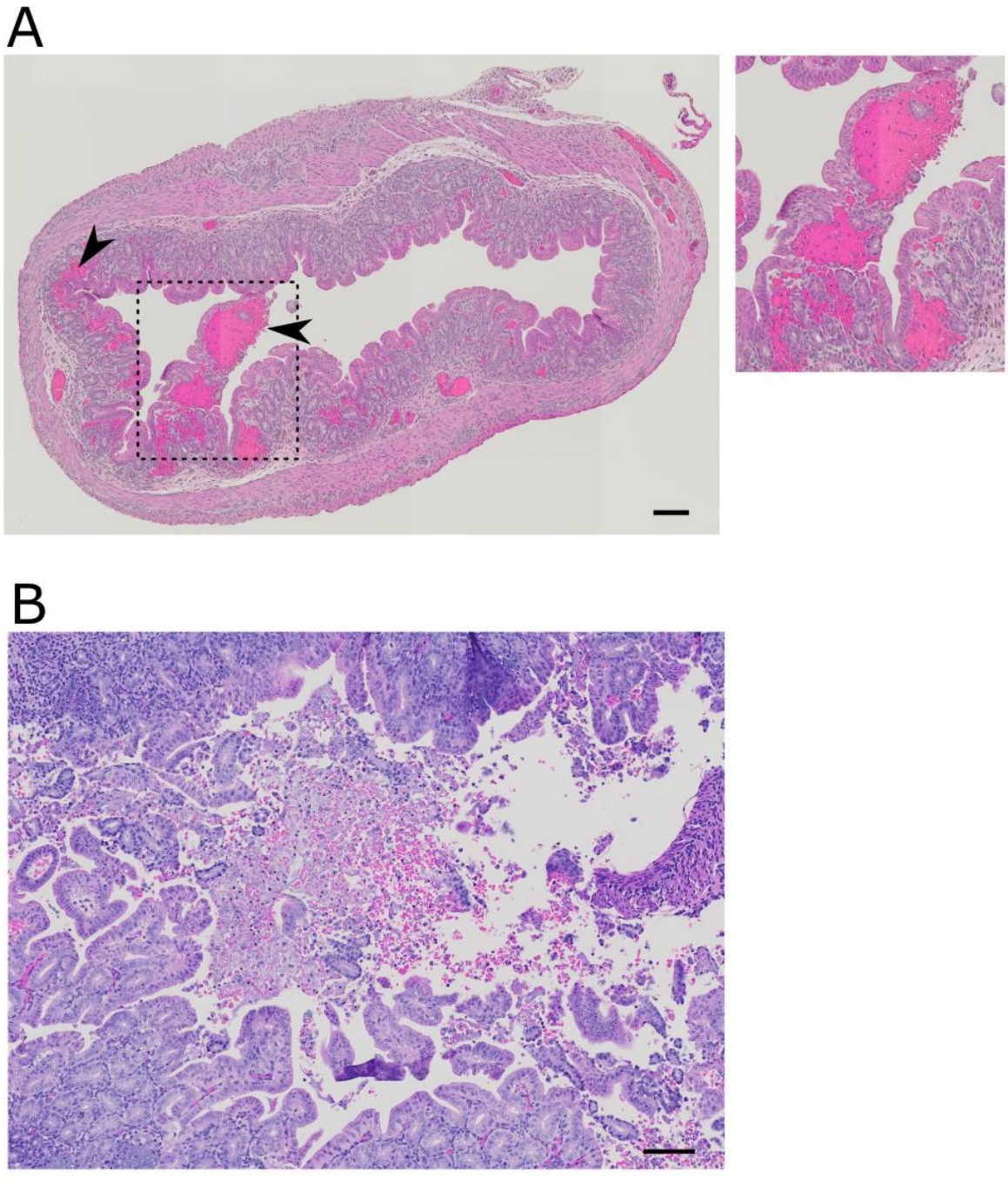


Figure 3.4. Examples of severe colonic pathology in infant rabbits inoculated with *S. flexneri* infection.

A-B. Haematoxylin and eosin-stained colonic sections of severe hemorrhage in lamina propria and colonic lumen from animals infected with the WT strain at 36 hpi. Arrowheads in (A) indicate either an area of hemorrhage in the lamina propria or hemorrhage spreading to the lumen (inset, A). (B) Hemorrhage and epithelial cell sloughing in colonic lumen. Scale bar is 100 μm.

Figure 3.4 (Continued)



3.4). Uninfected rabbits, which were similarly fasted, did not display colonic pathology and had no edema or disruption of the surface layer of epithelial cells (Figure 3.3D). Notably, although the bacterial burden in the colon was similar to that of the distal small intestine (Figure 3.3A), substantial pathology was not observed in the distal small intestine, suggesting organ-specific host factors influence the development of intestinal pathology.

S. flexneri invades colonic epithelial cells after oro-gastric infection

Tissue sections from the colons of infected rabbits were examined with immunofluorescence microscopy to determine the spatial distribution of *S. flexneri* in this organ. The pathogen, which was labeled with an anti-*Shigella* antibody, was detected in the intestinal lumen and in many scattered foci within the epithelium (Figure 3.5AB). At low magnification, the signal from the immunostained pathogen appeared to overlap with epithelial cells (Figure 3.5A-C). At high magnification, immunostained *S. flexneri* was clearly evident within the boundaries of epithelial cells, which were visualized with phalloidin staining of actin and an antibody against E-cadherin (Figure 3.5D). Several *S. flexneri* cells were frequently observed within an infected epithelial cell. In some infected epithelial cells, we observed *S. flexneri* cells associated with phalloidin stained actin tails (Figure 3.5E), and in other foci, we observed *S. flexneri* in protrusions emanating from a primary infected cell with many cytosolic bacteria (Figure 3.5F, asterisk), similar to structures seen in *Shigella* infections of tissue cultured cells (40, 41). The detection of actin tails and protrusions supports the hypothesis that the pathogen is actively spreading within the epithelial cell layer in the colon. *S. flexneri* cells were primarily localized to the epithelial cell layer and were infrequently observed in the lamina propria, the region of the intestinal wall directly below the epithelial cell layer. We did not find bacteria in

Figure 3.5. Localization of *S. flexneri* in the colons of infected infant rabbits.

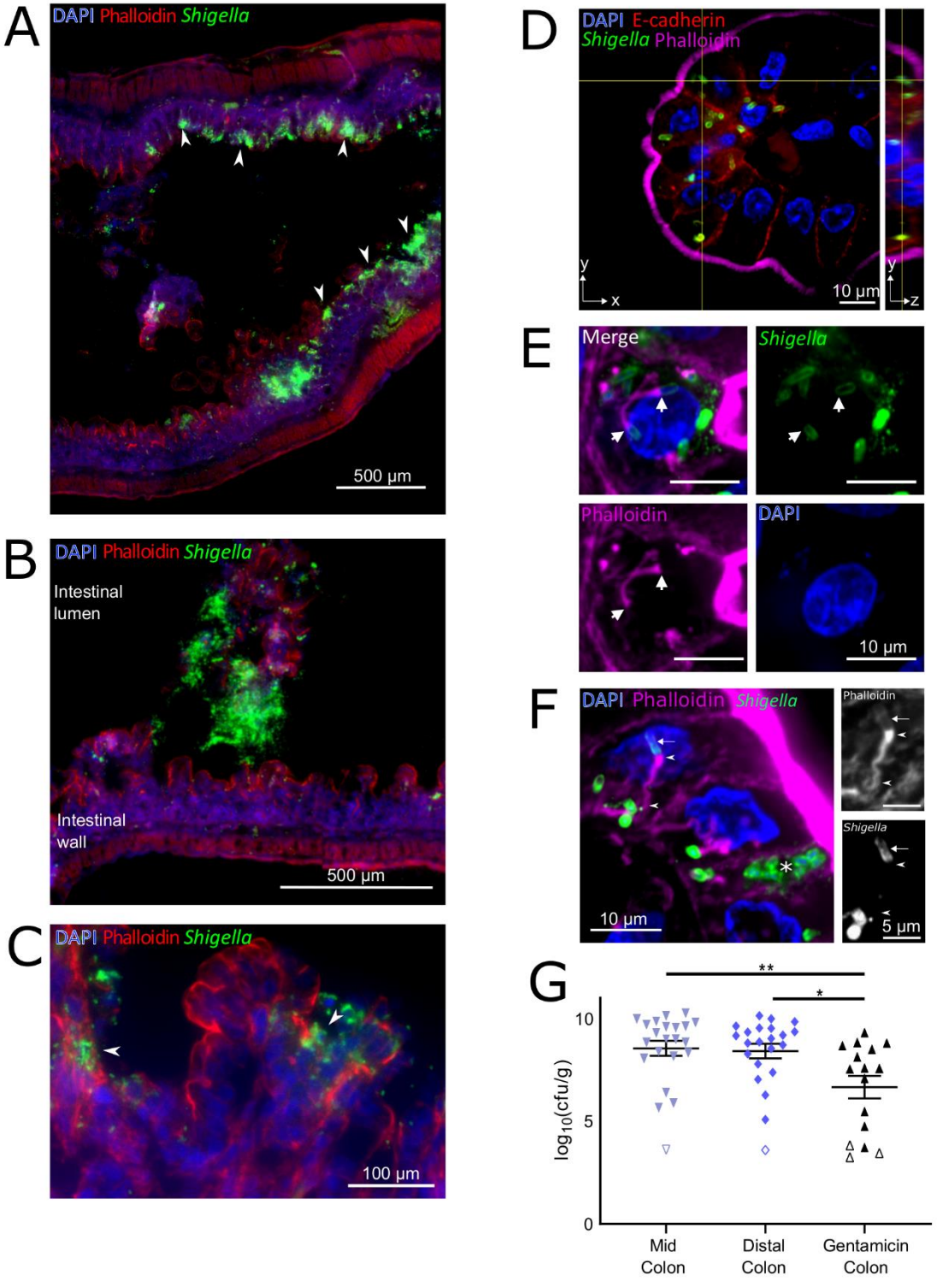
A-F. Immunofluorescence micrographs of *S. flexneri* in colonic tissue of infected rabbits 36 hpi. (A) *S. flexneri* bacteria were found in large numbers in epithelial foci (A, arrowheads point to selected foci). Scale bar is 500 μm. (B) *S. flexneri* bacteria in the lumen of the colon; the intestinal lumen and intestinal wall are labeled. Scale bar is 500 μm. (C) Arrowheads show infection foci where multiple neighboring cells contain intracellular *S. flexneri*. Scale bar is 100 μm. (D) Immunofluorescence z-stack micrograph of *S. flexneri* within colonic epithelial cells. Scale bar is 10 μm. Left (square) panel shows xy plane at a single z position, indicated by the horizontal axis of the cross-hairs in the yz projection. Right (rectangular) panel shows yz projection along the plane indicated by the vertical axis of the cross-hairs in the xy plane. Scale bar is 10 μm. (E) Immunofluorescence micrograph of *S. flexneri* associated with actin tails within colonic epithelial cells. Arrows point to poles of *S. flexneri* bacterial cells from which the actin tail is formed. Scale bar is 10 μm. (F) Immunofluorescence micrograph of *S. flexneri* forming protrusions during cell-to-cell spread between colonic epithelial cells. Asterisk marks a likely primary infected cell. Scale bar is 10 μm. Panels show zoomed region of phalloidin or anti-*Shigella* channels. Scale bar is 5 μm. Arrow points to actin surrounding the bacterial cell in a protrusion, arrowheads indicate the actin tail and actin cytoskeleton inside the protrusion at the pole of the bacterial cell and at the base of the protrusion. Blue, DAPI; green, FITC-conjugated anti-*Shigella* antibody; red (A-C or magenta in D-F), phalloidin-Alexa 568; and when present, red (D), anti-E-cadherin.

G. Bacterial burden of *S. flexneri* WT strain in the indicated intestinal sections 36 hpi. Each point represents measurement from one rabbit. Data plotted as log transformed colony forming units (cfu) per gram of tissue; means and standard error of the mean values are superimposed. Open

Figure 3.5 (Continued)

symbols represent the limit of detection of the assay and are shown for animals where no cfu were recovered. Statistical significance was determined with a Kruskal-Wallis test with Dunn's multiple comparison. p-values: <0.05, *; <0.01, **.

Figure 3.5 (Continued)



the deeper layers of the intestine (Figure 3.5A). Hence, following oro-gastric inoculation of infant rabbits with *S. flexneri*, the pathogen appears to proliferate both within the colonic lumen and in epithelial cells without penetration into deeper tissues.

We also measured the burden of intracellular *S. flexneri* in the colon using a modified gentamicin protection assay previously used to study the intracellular burden of *Listeria monocytogenes* and *Salmonella enterica* serovar Typhimurium in murine intestinal tissues (42–46). After dissecting intestines from infected infant rabbits, colonic tissue was incubated with gentamicin, an antibiotic that selectively kills extracellular (i.e. luminal) bacteria. We observed an ~2 log decrease in bacterial burden after gentamicin treatment (Figure 3.5G), suggesting that only a small portion of *S. flexneri* in the colon are intracellular.

IL-8 transcripts are often observed in epithelial cells near infected cells

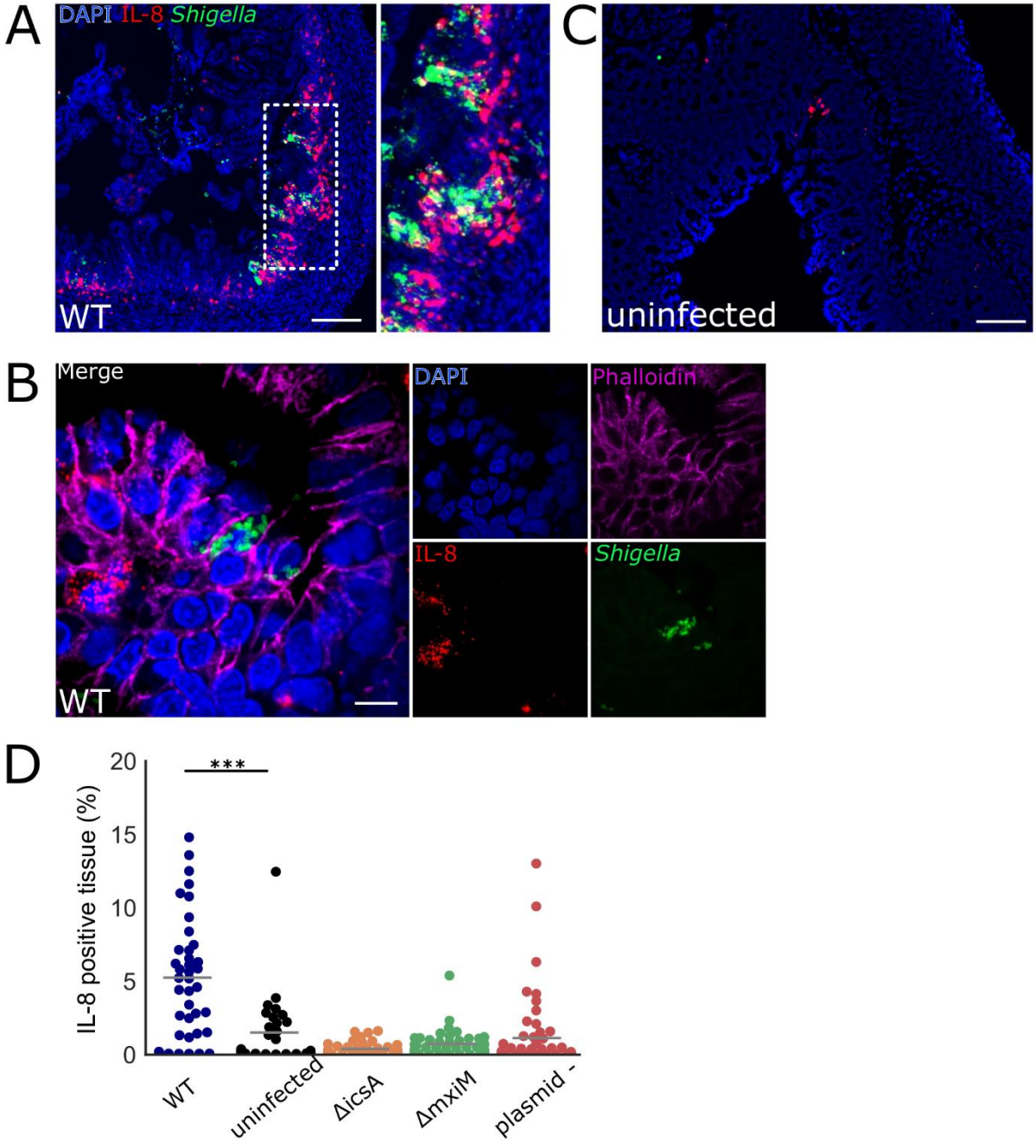
We next investigated aspects of the infant rabbit host innate immune response to *S. flexneri* infection. IL-8, a proinflammatory CXC family chemokine that recruits neutrophils (47), has been shown to be elevated during *Shigella* infection in animal models (10, 21, 48) and in humans (49, 50). However, in preliminary experiments it was difficult to detect significant elevations of IL-8 transcripts in bulk colonic tissue using a qPCR-based assay. Due to the patchiness of the infection foci observed through immunofluorescence imaging of colonic tissue, we wondered whether a localized response to infection might be masked when analyzing bulk intestinal tissue specimens. Local expression of IL-8 mRNA in *S. flexneri*-infected tissue was assessed using RNAscope technology, a sensitive, high-resolution in situ mRNA imaging platform that permits spatial analysis of mRNA expression. In the colon, we detected localized expression of IL-8 mRNA in colonic epithelial cells near infection foci (Figure 3.6AB). In

Figure 3.6. Colonic IL-8 mRNA in rabbits infected with *S. flexneri*.

A-C. Immunofluorescence micrographs of colonic sections from infant rabbits infected with WT *S. flexneri* (A, B) or uninfected control (C). Sections were stained with an RNAscope probe to rabbit IL-8 (red), an antibody to *Shigella* (FITC-conjugated anti-*Shigella* green), and with DAPI (blue). (A) Colon section infected with WT *S. flexneri.* Inset on right of A depicts magnified view of boxed area on left image. Scale bar is 200 μm. (B) High magnification of colonic epithelium infected with WT *S. flexneri.* Sections were also stained with anti-E-cadherin antibody (magenta). Scale bar is 10 μm. (C) Uninfected colon section. Scale bar is 100 μm.

D. Percentage of IL-8 expressing cells in each field of view from colonic tissue sections stained with probe to rabbit IL-8 from rabbits infected with the indicated strain. See methods for additional information regarding the determination of these measurements. Mean values are indicated with bars. All groups were compared to the sections from the uninfected animals. Statistical significance was determined using a Kruskall-Wallis test with Dunn's multiple comparison. P-values: <0.001, ***.

Figure 3.6 (Continued)



contrast, very few IL-8 transcripts were detected in the colons of uninfected kits (Figure 3.6CD). Combined detection of IL-8 and *S. flexneri* demonstrated that IL-8 expressing cells were typically near cells containing *S. flexneri*, but not themselves infected with the pathogen (Figure 3.6A, B & D, & 3.7). The majority (>90%) of infected epithelial cells did not express IL-8 mRNA, while >40% of these infected cells were adjacent to uninfected cells that did express IL-8 mRNA. Several T3SS effectors from *S. flexneri*, e.g. IpgD (51) and OspF (52), have been shown to reduce IL-8 expression in infected cells, which may explain the weak or absent IL-8 production in infected cells. There was a wide range in the prevalence of IL-8 producing cells in infected animals (Figure 3.6D & 3.7). The variability of IL-8 expression after infection may reflect the patchiness of *S. flexneri* invasion along the colon (Figure 3.5A). Together, these observations suggest that *S. flexneri* infection induces IL-8 mRNA expression (and perhaps additional cytokines as well) in infant rabbits.

Narrow bottleneck to *Shigella* infection of the infant rabbit colon

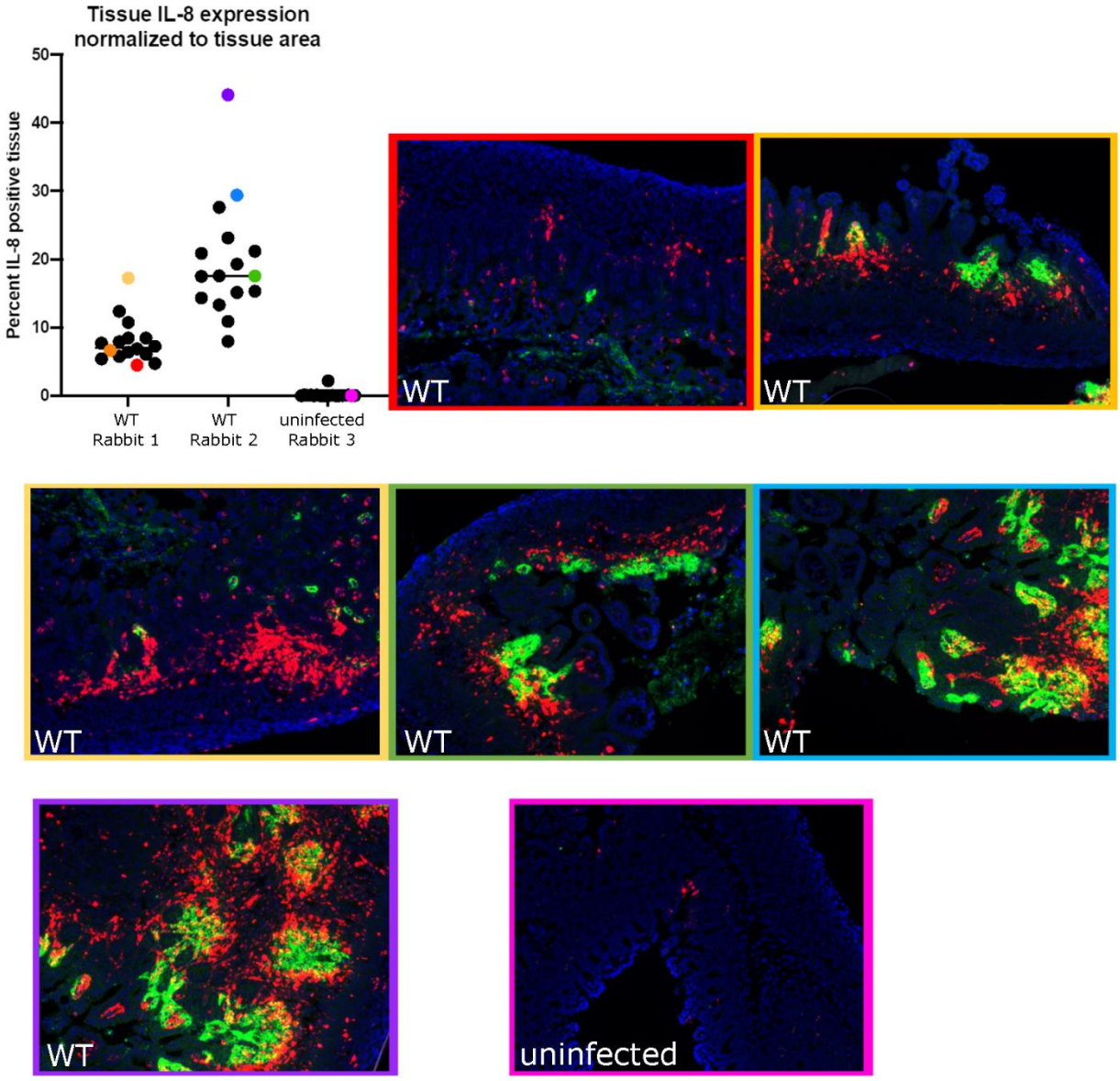
We attempted to use transposon-insertion sequencing (TIS) to identify genetic loci contributing to *S. flexneri* colonization and pathogenesis, as we have done with *V. cholerae* (53, 54), *V. parahaemolyticus* (55), and EHEC (56). Initially, a high-density transposon mutant library in *S. flexneri* was created using a *mariner*-based transposon that inserts at TA dinucleotide sites in the genome. The library included insertions across the entirety of the genome, including the virulence plasmid. Infant rabbits were inoculated with the transposon library and transposon mutants that persisted after 36 hpi were recovered from the colon. Comparison of the frequencies of insertions in the input and output libraries revealed that the output transposon libraries recovered from rabbit colons only contained ~20% of the

Figure 3.7. Range of IL-8 expression in colons of infected infant rabbits.

(Top left, plot) Percentage of IL-8 expressing cells in each field of view from colonic tissue sections stained with probe to rabbit IL-8 from individual rabbits infected with the WT strain or from uninfected rabbits. Colored dots correspond to micrographs with similar colored borders. Mean values are indicated with bars.

(Micrographs) Immunofluorescence micrographs of colonic sections from uninfected animals or infant rabbits infected with WT *S. flexneri* strain. Sections were stained with a RNAscope probe to rabbit IL-8 (red), and an antibody to *Shigella* (green), and with DAPI (blue).

Figure 3.7 (Continued)



transposon mutants that were present in the input library. These observations suggest that there is a very narrow bottleneck for *S. flexneri* infection in rabbits, leading to large, random losses of diversity in the input library. These random losses of mutants confound interpretation of these experiments and precludes accurate identification of genes subject to in vivo selection. Modifications to the in vivo TIS protocol will be necessary to apply TIS to identify additional *S. flexneri* colonization factors.

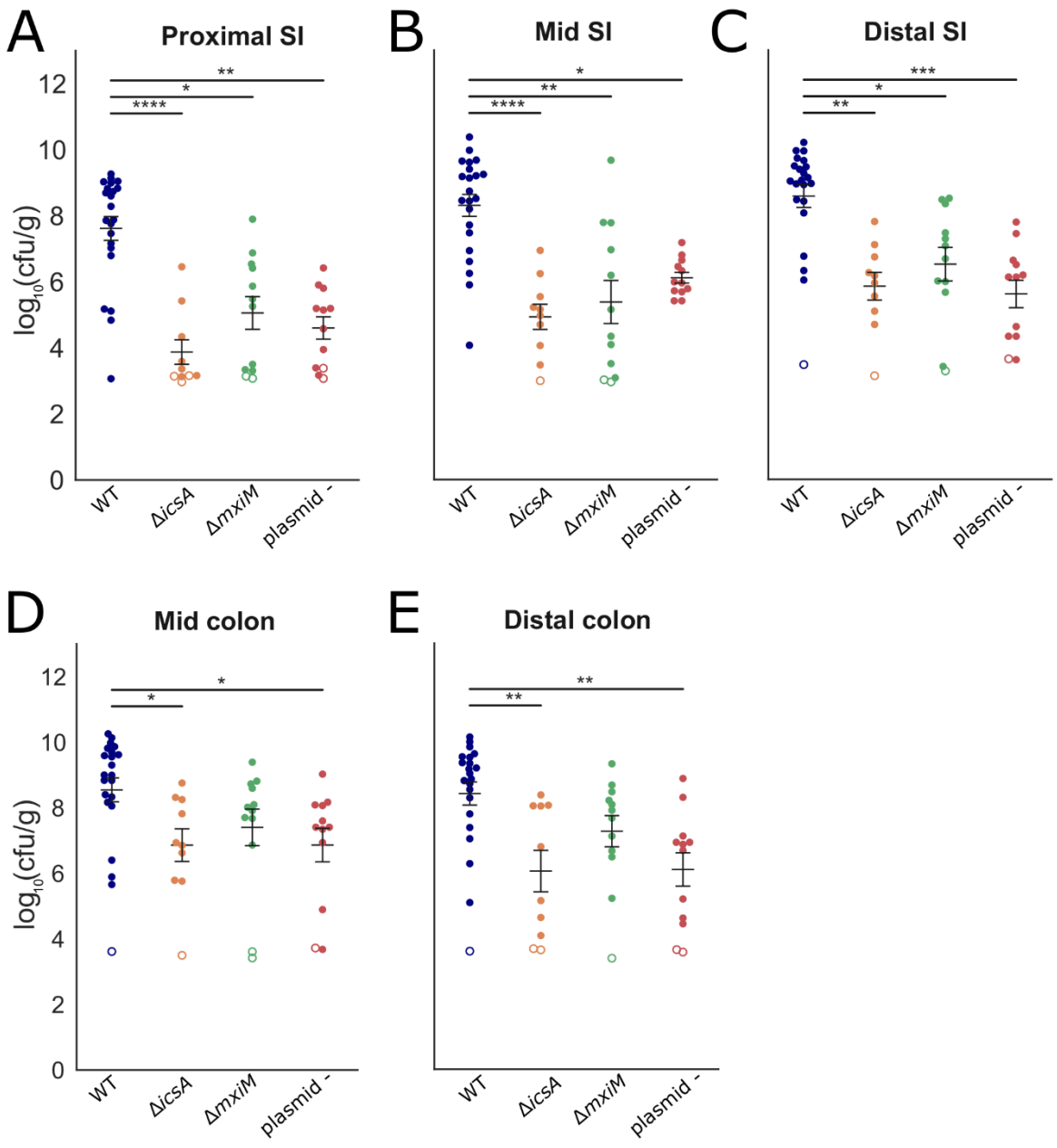
Canonical *S. flexneri* virulence factors are required for intestinal colonization and pathogenesis

Next, we investigated the requirement for canonical *Shigella* virulence factors in intestinal colonization and disease pathogenesis. First, we tested a strain that lacked the entire virulence plasmid, which contains most of the known virulence factors encoded in the *S. flexneri* genome, including the T3SS. As anticipated, this strain was avirulent; animals inoculated with the plasmidless (plasmid -) *S. flexneri* strain did not die or develop diarrhea or reduced temperature (Figure 3.1A & E). We also tested isogenic mutants that lack one of two key virulence factors: IcsA (Δ*icsA* strain), which is required for intracellular actin-based motility and cell-to-cell spreading, and MxiM (Δ*mxiM* strain), which is a T3SS structural component (57). *MxiM* deletion mutants do not assemble a functional T3SS, do not secrete T3SS effectors, and do not invade tissue-cultured epithelial cells (57–59). Like the plasmidless strain, the Δ*icsA* and Δ*mxiM* strains did not cause disease; none of the rabbits infected with either of these two mutant strains developed diarrhea, succumbed to infection, or had a reduction in body temperature (Figure 3.1A & E). Additionally, none of the mutants induced colonic edema or epithelial cell sloughing, pathologic features that characterized WT infection (Figure 3.3). Collectively, these

Figure 3.8. Intestinal colonization of WT and mutant *S. flexneri*.

A-E. Bacterial burden of the indicated strains in the indicated intestinal sections 36 hpi. SI = small intestine. Each point represents measurement from one rabbit. Data plotted as log transformed colony forming units (cfu) per gram of tissue. Means and standard error of the mean values are superimposed. Open symbols represent the limit of detection of the assay and are shown for animals where no cfu were recovered. For each section, burdens from all strains were compared to each other; statistical significance was determined using a Kruskal-Wallis test with Dunn's multiple comparison. P-values: <0.05, *; <0.01, **; <0.001, ***; <0.0001, ****.

Figure 3.8 (Continued)



data indicate that both IcsA and the T3SS are required for *Shigella* pathogenesis in the infant rabbit model.

All three of the mutant strains had reduced capacities to colonize the infant rabbit intestine (Figure 3.8). Notably, the reduction in the colonization of the *icsA* mutant was at least as great as the other two mutant strains, suggesting that cell-to-cell spreading or the adhesin function of IcsA is critical for intestinal colonization. The colonization defects were most pronounced in the small intestine, where up to 10^4-fold reductions in recoverable *S. flexneri* cfu were observed (Figure 3.8). Reductions in the colon were less marked and did not reach statistical significance for the Δ*mxiM* strain (Figure 3.8).

Interestingly, the *icsA* mutant led to an accumulation of heterophils (innate immune cells that are the rabbit equivalent of neutrophils) in the colon that was not observed in animals infected with the WT strain (Figure 3.9). Thus, IcsA may contribute to immune evasion by limiting the recruitment of innate immune cells. The *mxiM* mutant also recruited more heterophils to the lamina propria and epithelial cell layer than the WT strain (Figure 3.9A & C). Unlike the Δ*icsA* and Δ*mxiM* strains, the plasmidless strain did not recruit heterophils in the colon. Thus, both IcsA and the T3SS appear to antagonize heterophil recruitment, perhaps by facilitating pathogen invasion. However, the absence of heterophil influx in the plasmidless strain challenges this hypothesis and suggests that another plasmid-encoded factor can counteract the actions of IcsA and/or the T3SS in blocking heterophil infiltration.

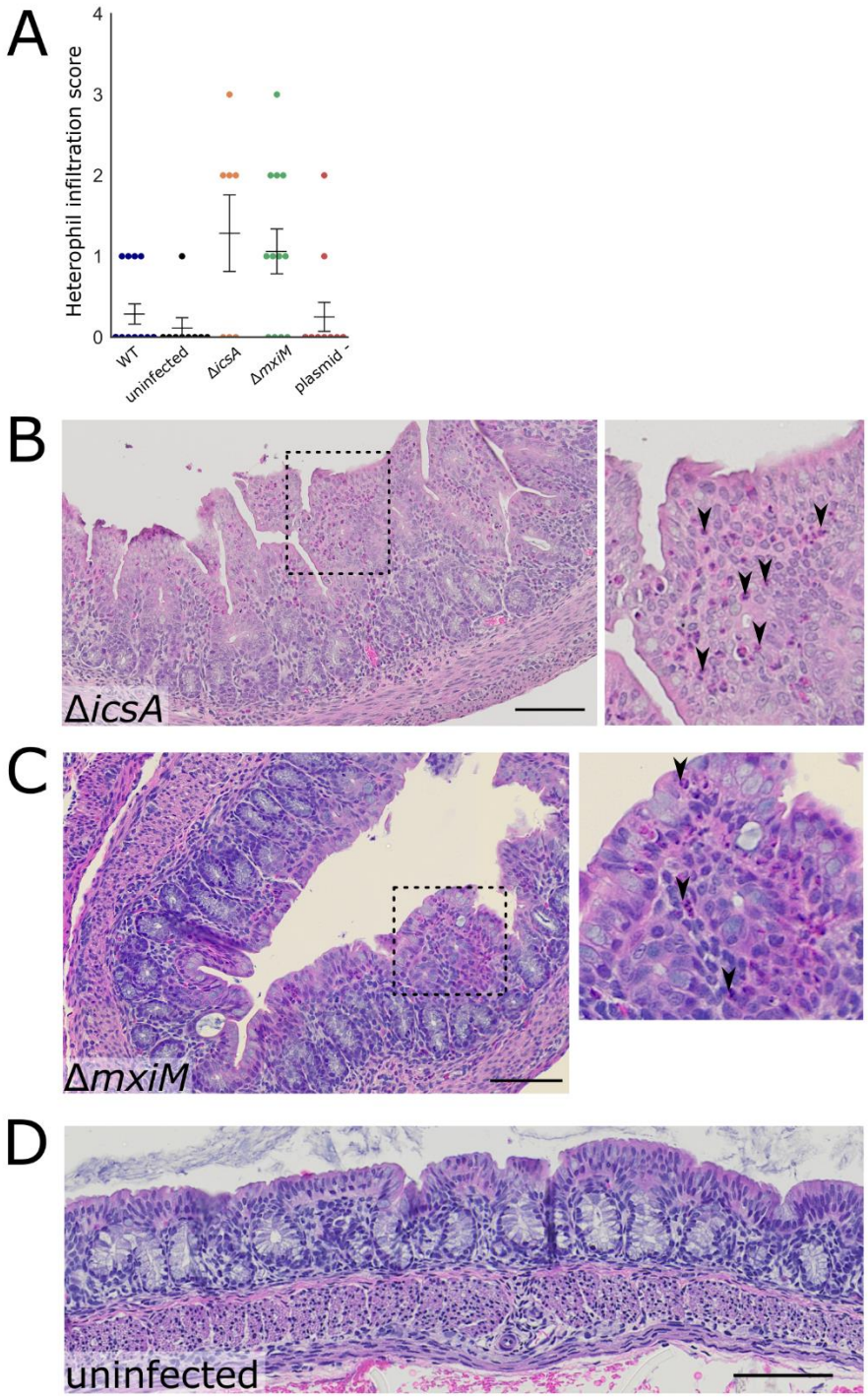
Since colonic pathology was altered in the mutant strains, we investigated the intestinal localization and IL-8 production induced by the mutants. All three of the mutant strains were found almost exclusively in the lumen of the colon (Figure 3.10A & 3.11); in contrast to the WT strain (Figure 3.5), it was difficult to detect infection foci in the epithelial cell layer in animals

Figure 3.9. Colonic pathology in rabbits infected with WT or mutant *S. flexneri*.

A. Histopathological scores of heterophil infiltration in colonic sections of animals infected with indicated strains of *S. flexneri*. Means and standard error of the mean values are superimposed. Statistical significance was determined using a Kruskal-Wallis test with Dunn's multiple comparison; comparisons that are non-significant are not labeled.

B-D. Representative haematoxylin and eosin-stained colonic sections from rabbits infected with the indicated strains 36 hpi. In B, the inset on the right displays the magnified version of the boxed region of the larger micrograph. Arrowheads point to heterophils (pink cytoplasm, multi-lobular darkly stained nucleus). Scale bar is 100 μm. In C (MxiM mutant), the inset on the right displays a magnified version of the boxed region of the larger micrograph. Arrowheads point to heterophils. Scale bar is 100 μm.

Figure 3.9 (Continued)



infected with mutant strains (Figure 3.10). The *icsA* mutant was occasionally observed inside epithelial cells (Figure 3.10B), but larger foci were not detected. As expected, we observed very few cells expressing IL-8 mRNA in the colons of rabbits infected with any of the three mutant *S. flexneri* strains (Figure 3.6D & 3.10B), supporting the idea that induction of IL-8 expression requires *S. flexneri* invasion of the epithelial cell layer in this model.

Discussion

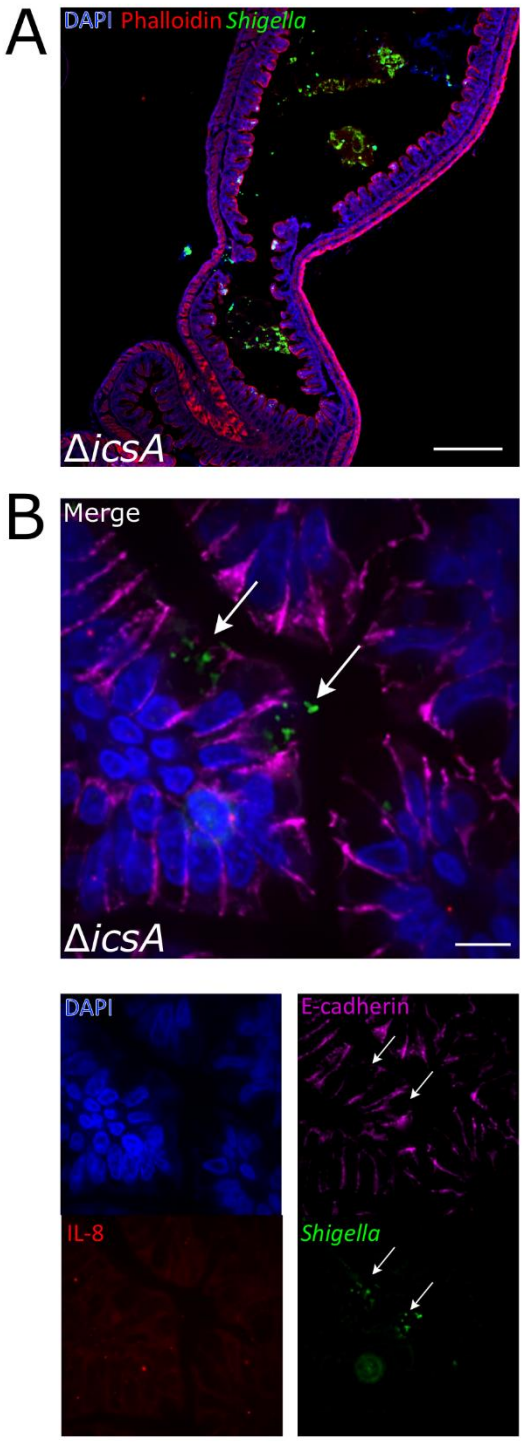
Small animal models of shigellosis that rely on the oral route of infection have been lacking. Here, we found that oro-gastric inoculation of two to three-day-old infant rabbits with *S. flexneri* led to a diarrheal disease and colonic pathology reminiscent of some aspects of human disease. Fasting animals prior to inoculation reduced the variability in infection outcomes, but not all inoculated animals developed disease. The pathogen robustly colonized the colon, where the organism was found primarily in the lumen; however, prominent infection foci were also observed within the colonic epithelium. Robust *S. flexneri* intestinal colonization, invasion of the colonic epithelium and colonic epithelial sloughing required IcsA and the T3SS, which are both canonical *S. flexneri* virulence factors. Despite the reduced intestinal colonization of the *icsA* and *mxiM* mutants, these strains elicited more pronounced colonic inflammation (characterized by infiltration of heterophils) than the WT strain. IL-8 expression, detected with in situ mRNA labeling, was higher in animals infected with the WT versus the mutant strains, suggesting that epithelial invasion promotes expression of this chemokine. Interestingly, IL-8 expression was greater in uninfected cells near infected epithelial cells than in infected epithelial cells themselves. Collectively, our findings suggest that oral infection of infant rabbits offers a useful experimental model for investigations of the pathogenesis of shigellosis.

Figure 3.10. Intestinal localization and IL-8 transcripts in colons from animals infected with an *icsA* mutant.

A. Immunofluorescence micrograph of Δ*icsA* in colonic tissue of infected rabbits 36 hpi. Blue, DAPI; green, FITC-conjugated anti-*Shigella* antibody; red, phalloidin-Alexa 568. Scale bar is 500 μm.

B. Immunofluorescence micrograph of sections stained with a RNAscope probe to rabbit IL-8 (red), antibodies to *Shigella* (green) and E-cadherin (magenta), and DAPI (blue). Panels on right depicts channels of merged left image. Arrows point to multiple *icsA* bacteria in the cytoplasm of two infected cells. Scale bar is 10 μm.

Figure 3.10 (Continued)



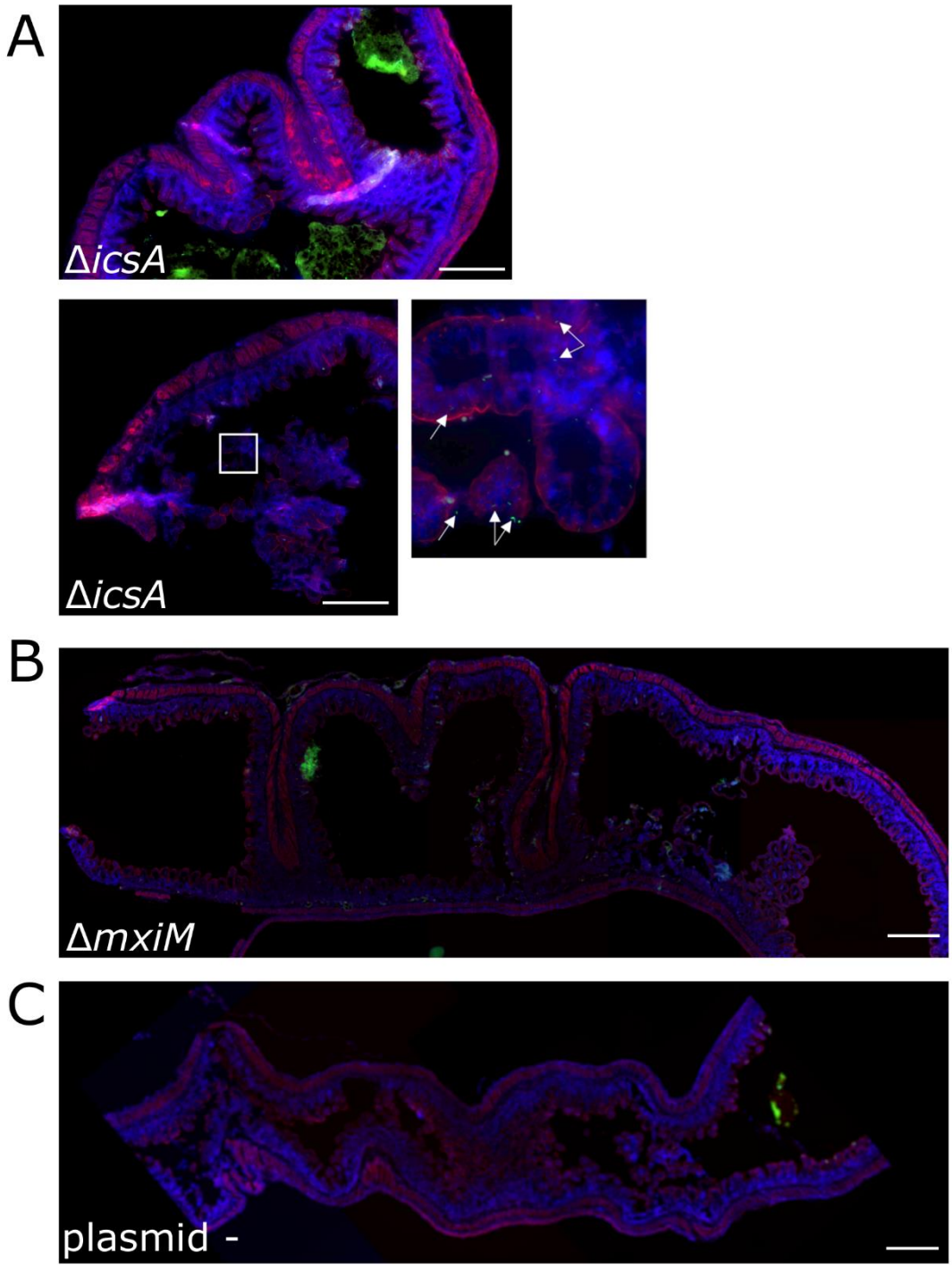
Fasted animals developed disease more frequently and had elevated intestinal colonization compared to animals who fed ad libidum prior to inoculation. The presence of inhibitory substances in milk, such as lactoferrin, which degrades components of the *Shigella* T3SS apparatus (39), may limit bacterial establishment in the intestine, but have less potent effects once colonization is established. Mean colonic colonization was higher in animals that developed disease than those that did not (Figure 3.2A). However, high bacterial burdens are not the only factor predictive of disease; several animals with high pathogen burdens did not exhibit signs of disease (Figure 3.2A). Also, initial rabbit body weights did not strongly influence clinical outcomes (Figure 3.2C). Several additional factors likely modulate *Shigella* colonization and disease manifestation in infant rabbits. For example, variations in the intestinal microbiota of the infant rabbits may limit or potentiate *S. flexneri* virulence and/or colonization, as described for infections caused by other enteric pathogens, including *Clostridium difficile* (60), *L. monocytogenes* (61), and *V. cholerae* (62). Differences in dam feeding patterns also likely influence colonization and disease outcomes. Further elucidation of factors that modulate outcomes will be valuable to improve this model because they may point to ways to elevate the fraction of animals that develop clinical signs of infection.

A high inoculum dose (10^9 cfu) was required to achieve reliable disease development following oral inoculation of two to three-day-old infant rabbits. Animals inoculated with lower doses (e.g., 10^7 cfu) of *S. flexneri* developed disease and robust intestinal colonization at lower frequencies. Interestingly, even in oral non-human primate models, the standard inoculum dose (10^{10} cfu) to ensure consistent development of disease (63, 64) is orders of magnitude greater than the dose used in human challenge studies (typically 10^3-10^4 cfu) (37, 65, 66). The reasons accounting for these marked differences in infectious doses warrant further exploration. It is

Figure 3.11. Localization of *S. flexneri* mutants in infected infant rabbits.

A-C. Immunofluorescence micrographs of *S. flexneri* mutants in colonic tissue of infected rabbits 36 hpi. (A) Inset and white arrows show individual *S. flexneri* Δ*icsA* closely associated with the colonic epithelium. Blue, DAPI; green, FITC-conjugated anti-*Shigella* antibody; red, phalloidin-Alexa 568. Scale bar is 500 μm (A-C).

Figure 3.11 (Continued)



unlikely that older rabbits infected via the oral route will be susceptible to colonization and disease, since our findings with other pathogens (35) suggest that infant rabbits become resistant to oral inoculation with enteric pathogens when they are ~5 days old.

In human infections, *Shigella* causes colonic pathology characterized by an acute inflammatory response with mucosal ulceration and erosions, neutrophil infiltration, congestion, and hemorrhage (8, 9). In the oral infant rabbit model, the WT strain caused edema and sloughing of epithelial cells in the colon, but we did not observe recruitment of heterophils, suggesting that colonic pathology is not primarily attributable to an acute inflammatory response characterized by heterophil infiltration. Instead, the pathology may be driven by invasion and replication of the pathogen within colonic epithelial cells. Oro-gastric inoculation of infant rabbits with EHEC induces heterophil infiltration in the colon (34), indicating that these animals are capable of mounting an acute inflammatory response in this organ.

The marked colonization defect of the Δ*icsA* strain, matching that observed for the Δ*mxiM* (T3SS -) and plasmidless strains, was unexpected. It seems unlikely the Δ*icsA* mutant's colonization defect is entirely attributable to the mutant's deficiency in cell-to-cell spreading. Zumsteg et al. found that IcsA can also serve as an adhesin (67). Since distinct regions of IcsA are required for its adhesive versus cell spreading activities (67), it may be possible to genetically dissect which of these IcsA functions plays a dominant role in colonization, using *S. flexneri* strains producing mutant versions of IcsA. Passage of the pathogen through the upper gastrointestinal tract may be required to reveal IcsA's adhesive activity, because a Δ*icsA* strain had only a modest colonization defect after intra-rectal instillation (10). It was also surprising that the animals infected with the Δ*icsA* strain recruited heterophils to the colon despite little induction of IL-8 expression. These observations suggest that there are additional factors

contributing to heterophil recruitment to the rabbit colon. Moreover, since there is minimal heterophil recruitment in animals infected with the WT strain, IcsA-mediated pathogen adherence to colonic epithelial cells (and potentially concomitant increased invasion) may increase delivery of T3SS effectors into host cells, thereby repressing a host-derived heterophil recruitment factor.

Our attempts to utilize TIS to identify novel genetic loci contributing to *S. flexneri* colonization in the infant rabbit intestine were stymied by a narrow infection bottleneck. The tight bottleneck leads to large, random losses of genetic diversity of the input library. The underlying causes of in vivo bottlenecks vary and may include stomach acidity, host innate immune defenses, such as antimicrobial peptides, the number of available niches in the intestine, and competition with the endogenous commensal microbiota (68). Modifications to either the inoculation protocol or library generation could facilitate future in vivo TIS screens. For example, the diversity of the inoculum could be reduced by generating a defined library of transposon mutants with only one or two mutants per gene (e.g. as has been done in *Edwardsiella piscicida* (69)). Regarding the infection protocol, it is possible that the fraction of the inoculum that initially seeds and colonizes the intestine could be elevated by reducing the number of commensal organisms in the intestine that may compete for a niche similar to that occupied by *S. flexneri*. Similar strategies have been utilized to facilitate studies of other enteric pathogens (61, 70).

The intra-rectal infant rabbit model of shigellosis reported by Yum et al. has some beneficial features compared to the oral infection model. Using this route, Yum et al. reported that all animals developed bloody diarrhea and colonic pathology that included substantial recruitment of heterophils (10). As noted above, for unknown reasons, oral inoculation of WT *S.*

flexneri did not lead to heterophil recruitment to sites of damage in the colon. An additional difference is that intra-rectal instillation of a Δ*icsA* mutant led to induction of cytokine expression, heterophil recruitment, and only slightly reduced colonization of the strain, whereas following oral inoculation, a Δ*icsA S. flexneri* exhibited a marked colonization defect and did not induce IL-8 mRNA expression. Additional studies are required to elucidate the reasons that account for the differential importance of IcsA in these models. While some features of the intra-rectal model are attractive, Yum et al. used 2 week old rabbits that were carefully hand reared in an animal facility from birth using a complex protocol that may prove difficult for others to adopt (10). In addition to the physiologic route of infection, the oral infant rabbit model requires far less specialized animal husbandry than the intra-rectal model and may therefore prove more accessible.

In summary, oral inoculation of infant rabbits with *Shigella* provides a feasible small animal model to study the pathogenesis of this globally important enteric pathogen. The model should also be useful to test new therapeutics for shigellosis, an issue of increasing importance given the development of *Shigella* strains with increasing resistance to multiple antibiotics (71–74).

Acknowledgements

This study was supported by the NIGMS grant T32GM007753 (J.D.D.), NIAID grant T32AI-132120 (J.D.D. & A.R.W.), and NIAID grant R01-AI-043247 and the Howard Hughes Medical Institute (M.K.W.).

We gratefully acknowledge Marcia Goldberg for providing *S. flexneri* 2a strains 2457T and BS103 (the virulence plasmidless derivative), and for transducing the streptomycin resistance allele into BS103. We thank Angelina Winbush for help with construction of the Δ*icsA* mutant strain. We thank the Dana-Farber/Harvard Cancer Center in Boston, MA, for the use of the Rodent Histopathology Core, which provided tissue embedding, sectioning, and staining service (NIH 5 P30 CA06516). We thank Rod Bronson at the Rodent Histopathology Core for providing blinded pathology scoring of tissue sections. We thank Brigid Davis and members of the Waldor lab for comments on the manuscript.

Materials and Methods

Bacterial strains and growth

S. flexneri were routinely grown aerobically in Miller lysogeny broth (LB) or LB agar at 30°C or 37°C. Antibiotics, when used, were included at the following concentrations: Streptomycin (Sm) 200 μg/mL, Kanamycin (Km) 50 μg/mL, Carbenicillin (Carb) 100 μg/mL, Chloramphenicol (Cm) 10 μg/mL. To check for the presence of the virulence plasmid, bacteria were grown on media with Congo red added at 0.1% w/v.

E. coli were routinely grown in LB media or agar. Antibiotics were used at the same concentrations as *S. flexneri* except for Cm, which was 30 μg/mL. When required, di-aminopimelic acid (DAP) was added at a concentration of 0.3 mM.

Strain construction

S. flexneri 2a strain 2457T and BS103 (a derivative lacking the virulence plasmid) were gifts of Marcia Goldberg. A spontaneous streptomycin resistant strain of *S. flexneri* 2a strain 2457T was generated by plating overnight LB cultures of *S. flexneri* 2a 2457T on 1000 μg/mL Sm LB plates and identifying Sm resistant (Sm^R) strains that grew as well as the parent strain. The Sm^R strain was used as the wild type strain for all subsequent experiments, including animal experiments and construction of mutant strains. Primers were used to amplify the *rpsL* gene in the strain and Sanger sequencing was performed to determine the nature of the mutation resulting in streptomycin resistance. The streptomycin resistance allele was transferred from the Sm^R wild type strain into strain BS103 by P1 transduction, yielding a Sm^R plasmidless (plasmid -) strain. Single gene deletion mutants were generated in the WT Sm^R strain using the lambda red

recombination method, as previously described (75). Resistance cassettes used in the process were amplified from pKD3 (Cm). Mutations generated by lambda red were moved into a clean genetic background by transferring the mutation to the Sm^R wild type strain via P1 transduction. Subsequently, antibiotic resistance cassettes were removed via FLP-mediated recombination using pCP20. Retention of the virulence plasmid throughout P1 transduction of the mutation into the parental WT Sm^R strain was monitored by plating bacterial mutants on LB + Congo Red to identify red colonies, and by performing multiplex PCR for various genes spread across the virulence plasmid.

Animal Experiments

Rabbit experiments were conducted according to the recommendations of the National Institutes of Health Guide for the Care and Use of Laboratory Animals, the Animal Welfare Act of the United States Department of Agriculture, and the Brigham and Women's Hospital Committee on Animals, as outlined in Institutional Animal Care and Use Compliance protocol #2016N000334 and Animal Welfare Assurance of Compliance number A4752-01.

Litters of two to three-day-old New Zealand White infant rabbits with lactating adult female (dam) obtained from a commercial breeder (Charles River, Canada or Pine Acres Rabbitry Farm & Research Facility, Norton, MA) were used for animal experiments.

Infant rabbits were administered a subcutaneous injection of Zantac (ranitidine hydrochloride, 50 mg/kg; GlaxoSmithKline) 3 hours prior to inoculation with the wild type (Sm^R) or isogenic mutants. We attempted to utilize a bicarbonate solution to administer bacteria, but found that *S. flexneri* do not survive when re-suspended in a sodium bicarbonate solution. For initial experiments, a day after arrival, infant rabbits were oro-gastrically inoculated with 1e9

cfu of log phase *S. flexneri* suspended in LB. To prepare the inoculum, an overnight bacterial culture grown at 30°C was diluted 1:100 and grown at 37°C for 3 hours. The bacteria were subsequently pelleted and re-suspended in fresh LB to a final concentration of 2e9 cfu/mL. Rabbits were oro-gastrically inoculated using a PE50 catheter (Becton Dickson) with 0.5 mL of inoculum (1e9 cfu total). In later experiments, infant rabbits were first separated from the dam for 24 hours prior to inoculation, after which they were immediately returned to the dam for the remainder of the experiment.

The infant rabbits were then observed for 36-40 hours post-inoculation and then euthanized via isoflurane inhalation and subsequent intracardiac injection of 6 mEq KCl at the end of the experiment or when they became moribund. Animals were checked for signs of disease every 10-12 hours. Body weight and body temperature measurements were made 1-2 times daily until the end of the experiment. Body temperature was measured with a digital temporal thermometer (Exergen) and assessed on the infant rabbit chest, in between the front legs. Temperatures reported in Figure 3.1E are the final temperatures prior to euthanasia and change in body weight in Figure 3.1F is a comparison of the final to initial body weight. Diarrhea was scored as follows: no diarrhea (solid feces, no adherent stool on hindpaw region) or diarrhea (liquid fecal material adhering to hindpaw region). Animal experiments with isogenic mutants were always conducted with litter-mate controls infected with the WT Sm^R strain to control litter variation.

At necropsy, the intestine from the duodenum to rectum was dissected, and divided into separate anatomical sections (small intestine, colon) as previously described (54, 76). 1-2 cm pieces of each anatomical section were used for measurements of tissue bacterial burden. Tissue samples were placed in 1x phosphate buffered saline (PBS) with 2 stainless steel beads and

homogenized with a bead beater (BioSpec Products Inc.). Serial dilutions were made using 1xPBS and plated on LB+Sm media for enumeration of bacterial cfu. For processing of tissue for microscopy, 1-2 cm pieces of the tissue adjacent to the piece taken for enumeration of bacterial cfu were submerged in either 4% paraformaldehyde (PFA) for frozen sections or 10% neutral-buffered formalin (NBF) for paraffin sections.

For gentamicin tissue assays, a 1-2 cm portion of the colon was cut open longitudinally and washed in 1X PBS to remove luminal contents and then incubated in 1mL of 1xDMEM with 100 μg/mL gentamicin for 1 hour at room temperature. The tissue was subsequently washed 3x with 20x volumes of 1x PBS for 30 min with shaking. The tissue was then homogenized and serial dilutions were plated on LB+Sm media for enumeration of bacterial burden.

For Tn-seq experiments, aliquots of the transposon library were thawed and aerobically cultured in LB for 3 hours. The bacteria were pelleted and resuspended in fresh LB to a final concentration of 1e9 cfu per 0.5 mL inoculum. A sample of the input library (1e10 cfu) was plated on a large LB+Sm+Km plate (245 cm^2; Corning). Bacterial burdens in infected rabbit tissues were determined by plating serial dilutions on LB+Sm+Km plates. The entire colon was homogenized and plated onto a large LB+Sm+Km plate to recover transposon mutants that survived in the colon. Bacteria on large plates were grown for ~20-22 hours at 30°C, scraped off with ~10 mL fresh LB, and ~ 1 mL aliquots were pelleted. The pellets were frozen at -80°C prior to genomic DNA extraction for Tn-seq library construction.

Data from animal experiments were analyzed in Prism (ver. 8; GraphPad). The Mann-Whitney U test or the Kruskal-Wallis test with Dunn's post-test for multiple comparisons were used to compare the tissue bacterial burdens. A Fisher's exact test was used to compare the proportion of rabbits that developed diarrhea after infection with various bacterial strains.

Immunofluorescence Microscopy

Immunofluorescence images were analyzed from 20 wild-type and at least 4 rabbits infected with each of the various mutant bacterial strains, or uninfected rabbits; 2-3 colon sections per rabbit were examined. Tissue samples used for immunofluorescence were fixed in 4% PFA, and subsequently stored in 30% sucrose prior to embedding in a 1:2.5 mixture of OCT (Tissue-Tek) to 30% sucrose and stored at -80°C, as previously described (35). Frozen sections were made at 10 μm using a cryotome (Leica CM1860UV). Sections were first blocked with 5% bovine serum albumin (BSA) in PBS for 1 hour. Sections were stained overnight at 4°C with a primary antibody, diluted in PBS with 0.5% BSA and 0.5% Triton X-100: anti-*Shigella*-FITC (1/1000; #0903, Virostat); anti-E-cadherin 1:100 (610181, BD Biosciences). After washing with 1xPBS - 0.5% Tween20 (PBST), sections were incubated with 647 phalloidin (1/1000; Invitrogen) for 1 hour at room temperature, washed and stained for 5 min with 4′,6-diamidino-2-phenylindole (DAPI) at 2 μg/mL for 5 min, and covered with ProLong Diamond or Glass Antifade (Invitrogen) mounting media. Slides were imaged using a Nikon Ti Eclipse equipped with a spinning disk confocal scanner unit (Yokogawa CSU-Xu1) and EMCCD (Andor iXon3) camera, or with a sCMOS camera (Andor Zyla) for widefield microscopy.

Histopathology

Tissue samples used for histopathology analysis were fixed in 10% NBF and subsequently stored in 70% ethanol prior to being embedded in paraffin, as previously described (36). Formalin fixed, paraffin embedded (FFPE) sections were made at a thickness of 5 μm. Sections were stained with hematoxylin and eosin (H&E). Slides were assessed for various measures of pathology, e.g. heterophil infiltration, edema, epithelial sloughing, hemorrhage, by a

pathologist blinded to the tissue origin. Semi-quantitative scoring for heterophil infiltration were as follows: 0, no heterophils observed; 1, rare heterophils; 2, few heterophils; 3, many heterophils; 4, abundant heterophils. Brightfield micrographs were collected using an Olympus VS120.

In situ RNA hybridization

Freshly cut FFPE sections (5 μm) were made of the indicated anatomical sections and stored with desiccants at 4°C. Subsequently, sections were processed and analyzed using the RNAscope Multiplex Fluorescent v2 Assay (Advanced Cell Diagnostics USA-ACDbio) combined with immunofluorescence. Briefly, sections were processed following ACDbio recommendations for FFPE sample preparation and pretreatment using 15-minute target retrieval and 25-minute Protease Plus digestion using the RNAscope HybEZ oven for all incubations. An RNAscope C1 probe (OcIL8) to rabbit CXCL8 was developed and used to stain intestinal sections for CXCL8 mRNA expression. C1 probe was detected with Opal 570 dye (Akoya Biosciences) diluted 1:1000 in Multiplex TSA buffer (ACDbio). Sections were also stained with DAPI (2 μg/mL), anti-*Shigella* FITC (1/1000, Virostat), and anti-mouse E-cadherin (1/100; #610181, BDbiosciences). Slides were imaged using a Nikon Ti Eclipse equipped with a spinning disk confocal scanner unit (Yokogawa CSU-Xu1) and EMCCD (Andor iXon3) camera for high magnification images. Slides were imaged using a widefield Zeiss Axioplan 2 microscope through the MetaMorph imaging system for RNAscope signal quantification.

Quantitative Image Analysis

Images of mid colon tissue sections stained with RNAscope OcIL8, DAPI, and FITC-conjugated anti-*Shigella* antibody were acquired and analyzed using the MetaMorph (7.1.4.0) application. Briefly, tiled 10x images covering the entire length of the tissue section were collected using Multi-Dimensional Acquisition. For analysis of the percentage of IL-8 mRNA expressing cells that were adjacent to infected cells, we analyzed 86 foci of infection at 100x magnification. Exclusive threshold values were set for the DAPI channel or the rhodamine channel independently and applied to all images in the data set. The threshold values for DAPI or rhodamine were used to create a binary mask of each image. The total area under the binary mask was recorded and used to calculate the percent of total tissue (DAPI area under mask) expressing CXCL8 mRNA (rhodamine area under mask) by dividing the values for rhodamine area by the DAPI area for each image. Percentages were graphed using Prism version 8 (GraphPad).

Transposon library construction and analysis

A transposon library was constructed in *S. flexneri* 2a 2457T Sm^R (WT Sm^R) using pSC189 (77), using previously described protocols (54, 78) with additional modifications. Briefly, *E. coli* strain MFD*pir* (79) was transformed with pSC189. Conjugation was performed between WT Sm^R and MFD*pir* pSC189. Overnight LB cultures of WT Sm^R (grown at 30°C) and MFD*pir* (grown at 37°C) were mixed and spotted onto 0.45 μm filters on LB+DAP agar plates. The conjugation reaction was allowed to proceed for 2 hours at 30°C. Subsequently, the bacterial mixtures were resuspended in LB and spread across four 245 cm^2 LB+Sm+Km square plates to generate single separate colonies for a transposon library. The square plates were grown at 30°C

for 20 hours. The colonies that formed (~800,000 total) were washed off with LB (8 mL per plate) and the bacteria from two plates were combined. Two separate 1 mL aliquots of the two combined mixtures were used to start two 100 mL LB+Km liquid cultures. The cultures were grown aerobically at 30°C with shaking for 3 hours. For each flask, the bacteria were pelleted and resuspended in a small amount of LB before being spread across two 245 cm^2 LB+Sm+Km square plates and grown at 30°C for 20 hours. The resulting bacteria on the plate were washed off with LB and resuspended. The OD was adjusted to 10 with LB and glycerol so that the final concentration of glycerol was 25%. 1 mL LB + glycerol aliquots were stored at -80°C for later experiments. In addition, 1 mL aliquots were also pelleted, to generate bacterial pellets to serve as sources of genomic DNA for the initial characterization of the transposon library. The pellets were stored at -80°C prior to genomic DNA extraction for Tn-seq library construction.

Tn-seq library construction and data analysis was performed as previously described (54, 55, 80); briefly, genomic DNA was extracted, transposon junctions were amplified, sequencing was performed on an Illumina MiSeq, and data were analyzed using a modified ARTIST pipeline (54, 55). Sequence reads were mapped onto the *S. flexneri* 2a strain 2457T chromosome (Refseq NC_004741.1) and *S. flexneri* 2a strain 301 virulence plasmid (Refseq NC_004851.1). Reads at each TA site were tallied.

References

1. **Legros D, Pierce N**. 2005. Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1World Health Organization.

2. **Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann CM, Houpt ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK, Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD, Walson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC**. 2018. Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect Dis* **18**:1229–1240.

3. **Troeger C, Blacker BF, Khalil IA, Rao PC, Cao S, Zimsen SR, Albertson SB, Stanaway JD, Deshpande A, Abebe Z, Alvis-Guzman N, Amare AT, Asgedom SW, Anteneh ZA, Antonio CAT, Aremu O, Asfaw ET, Atey TM, Atique S, Avokpaho EFGA, Awasthi A, Ayele HT, Barac A, Barreto ML, Bassat Q, Belay SA, Bensenor IM, Bhutta ZA, Bijani A, Bizuneh H, Castañeda-Orjuela CA, Dadi AF, Dandona L, Dandona R, Do HP, Dubey M, Dubljanin E, Edessa D, Endries AY, Eshrati B, Farag T, Feyissa GT, Foreman KJ, Forouzanfar MH, Fullman N, Gething PW, Gishu MD, Godwin WW, Gugnani HC, Gupta R, Hailu GB, Hassen HY, Hibstu DT, Ilesanmi OS, Jonas JB, Kahsay A, Kang G, Kasaeian A, Khader YS, Khan EA, Khan MA, Khang YH, Kissoon N, Kochhar S, Kotloff KL, Koyanagi A, Kumar GA, Magdy Abd El Razek H, Malekzadeh R, Malta DC, Mehata S, Mendoza W, Mengistu DT, Menota BG, Mezgebe HB, Mlashu FW, Murthy S, Naik GA, Nguyen CT, Nguyen TH, Ningrum DNA, Ogbo FA, Olagunju AT, Paudel D, Platts-Mills JA, Qorbani M, Rafay A, Rai RK, Rana SM, Ranabhat CL, Rasella D, Ray SE, Reis C, Renzaho AM, Rezai MS, Ruhago GM, Safiri S, Salomon JA, Sanabria JR, Sartorius B, Sawhney M, Sepanlou SG, Shigematsu M, Sisay M, Somayaji R, Sreeramareddy CT, Sykes BL, Taffere GR, Topor-Madry R, Tran BX, Tuem KB, Ukwaja KN, Vollset SE, Walson JL, Weaver MR, Weldegwergs KG, Werdecker A, Workicho A, Yenesew M, Yirsaw BD, Yonemoto N, El Sayed Zaki M, Vos T, Lim SS, Naghavi M, Murray CJ, Mokdad AH, Hay SI, Reiner RC**. 2018. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect Dis* **18**:1211–1228.

4. **Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM**. 1999. Global burden of Shigella infections: Implications for vaccine development and implementation of control strategies. *Bull World Health Organ* **77**:651–666.

5. **Ram PK, Crump JA, Gupta SK, Miller MA, Mintz ED**. 2008. Part II. Analysis of data gaps pertaining to Shigella infections in low and medium human development index countries, 1984-2005. *Epidemiol Infect* **136**:577–603.

6. **Livio S, Strockbine NA, Panchalingam S, Tennant SM, Barry EM, Marohn ME, Antonio M, Hossain A, Mandomando I, Ochieng JB, Oundo JO, Qureshi S, Ramamurthy T, Tamboura B, Adegbola RA, Hossain MJ, Saha D, Sen S, Faruque ASG, Alonso PL, Breiman RF, Zaidi AKM, Sur D, Sow SO, Berkeley LY, O'Reilly**

CE, Mintz ED, Biswas K, Cohen D, Farag TH, Nasrin D, Wu Y, Blackwelder WC, Kotloff KL, Nataro JP, Levine MM. 2014. Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin Infect Dis* **59**:933–941.

7. **Thompson CN, Duy PT, Baker S**. 2015. The rising dominance of Shigella sonnei: An intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis* **9**:1–13.

8. **Anand BS, Malhotra V, Bhattacharya SK, Datta P, Datta D, Sen D, Bhattacharya MK, Mukherjee PP, Pal SC**. 1986. Rectal histology in acute bacillary dysentery. *Gastroenterology* **90**:654–60.

9. **Mathan MM, Mathan VI**. 1991. Morphology of rectal mucosa of patients with shigellosis. *Rev Infect Dis* **13 Suppl 4**:S314-8.

10. **Yum LK, Byndloss MX, Feldman SH, Agaisse H**. 2019. Critical role of bacterial dissemination in an infant rabbit model of bacillary dysentery. *Nat Commun* **10**:1–10.

11. **Schroeder GN, Hilbi H**. 2008. Molecular pathogenesis of Shigella spp.: controlling host cell signaling, invasion, and death by type III secretion. *Clin Microbiol Rev* **21**:134–56.

12. **Sansonetti PJ, Kopecko DJ, Formal SB**. 1981. Shigella sonnei plasmids: evidence that a large plasmid is necessary for virulence. *Infect Immun* **34**:75–83.

13. **Sansonetti PJ, Kopecko DJ, Formal SB**. 1982. Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infect Immun* **35**:852–60.

14. **Venkatesan MM, Goldberg MB, Rose DJ, Grotbeck EJ, Burland V, Blattner FR**. 2001. Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri. *Infect Immun* **69**:3271–85.

15. **Sasakawa C, Kamata K, Sakai T, Makino S, Yamada M, Okada N, Yoshikawa M**. 1988. Virulence-associated genetic regions comprising 31 kilobases of the 230-kilobase plasmid in Shigella flexneri 2a. *J Bacteriol* **170**:2480–4.

16. **Makino S, Sasakawa C, Kamata K, Kurata T, Yoshikawa M**. 1986. A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in S. flexneri 2a. *Cell* **46**:551–5.

17. **Goldberg MB, Theriot JA**. 1995. Shigella flexneri surface protein IcsA is sufficient to direct actin-based motility. *Proc Natl Acad Sci U S A* **92**:6572–6.

18. **Lederer I, Much P, Allerberger F, Voracek T, Vielgrader H**. 2005. Outbreak of shigellosis in the Vienna Zoo affecting human and non-human primates [3]. *Int J Infect Dis* **9**:290–291.

19. **Formal SB, Gemski P, Baron LS, Labrec EH**. 1971. A Chromosomal Locus Which Controls the Ability of Shigella flexneri to Evoke Keratoconjunctivitis. *Infect Immun* **3**:73–9.

20. **West NP**. 2005. Optimization of Virulence Functions Through Glucosylation of Shigella LPS. *Science (80)* **307**:1313–1317.

21. **Shim D-H, Suzuki T, Chang S-Y, Park S-M, Sansonetti PJ, Sasakawa C, Kweon M-**

N. 2007. New Animal Model of Shigellosis in the Guinea Pig: Its Usefulness for Protective Efficacy Studies. *J Immunol* **178**:2476–2482.

22. **Arena ET, Campbell-Valois FX, Tinevez JY, Nigro G, Sachse M, Moya-Nilges M, Nothelfer K, Marteyn B, Shorte SL, Sansonetti PJ**. 2015. Bioimage analysis of Shigella infection reveals targeting of colonic crypts. *Proc Natl Acad Sci U S A* **112**:E3282–E3290.

23. **Xu D, Liao C, Zhang B, Tolbert WD, He W, Dai Z, Zhang W, Yuan W, Pazgier M, Liu J, Yu J, Sansonetti PJ, Bevins CL, Shao Y, Lu W**. 2018. Human Enteric α-Defensin 5 Promotes Shigella Infection by Enhancing Bacterial Adhesion and Invasion. *Immunity* **48**:1233-1244.e6.

24. **Tinevez J-Y, Arena ET, Anderson M, Nigro G, Injarabian L, André A, Ferrari M, Campbell-Valois F-X, Devin A, Shorte SL, Sansonetti PJ, Marteyn BS**. 2019. Shigella-mediated oxygen depletion is essential for intestinal mucosa colonization. *Nat Microbiol* **4**:2001–2009.

25. **Singer M, Sansonetti PJ**. 2004. IL-8 Is a Key Chemokine Regulating Neutrophil Recruitment in a New Mouse Model of Shigella- Induced Colitis . *J Immunol* **173**:4197–4206.

26. **Voino-Yasenetsky M V, Voino-Yasenetsky MK**. 1962. Experimental pneumonia caused by bacteria of the Shigella group. *Acta Morphol Acad Sci Hung* **11**:439–54.

27. **Mallett CP, VanDeVerg L, Collins HH, Hale TL**. 1993. Evaluation of Shigella vaccine safety and efficacy in an intranasally challenged mouse model. *Vaccine* **11**:190–196.

28. **Van de Verg LL, Mallett CP, Collins HH, Larsen T, Hammack C, Hale TL**. 1995. Antibody and cytokine responses in a mouse pulmonary model of Shigella flexneri serotype 2a infection. *Infect Immun* **63**:1947–1954.

29. **Way SS, Borczuk AC, Dominitz R, Goldberg MB**. 1998. An Essential Role for Gamma Interferon in Innate Resistance to Shigella flexneri Infection. *Infect Immun* **66**:1342–1348.

30. **Fernandez MI, Thuizat A, Pedron T, Neutra M, Phalipon A, Sansonetti PJ**. 2003. A newborn mouse model for the study of intestinal pathogenesis of shigellosis. *Cell Microbiol* **5**:481–491.

31. **Fernandez M-I, Regnault B, Mulet C, Tanguy M, Jay P, Sansonetti PJ, Pédron T**. 2008. Maturation of Paneth Cells Induces the Refractory State of Newborn Mice to Shigella Infection . *J Immunol* **180**:4924–4930.

32. **Mostowy S, Boucontet L, Mazon Moya MJ, Sirianni A, Boudinot P, Hollinshead M, Cossart P, Herbomel P, Levraud J-P, Colucci-Guyon E**. 2013. The zebrafish as a new model for the in vivo study of Shigella flexneri interaction with phagocytes and bacterial autophagy. *PLoS Pathog* **9**:e1003588.

33. **Duggan GM, Mostowy S**. 2018. Use of zebrafish to study Shigella infection. *Dis Model Mech* **11**:1–11.

34. **Ritchie JM, Thorpe CM, Rogers AB, Waldor MK**. 2003. Critical roles for stx2, eae, and tir in enterohemorrhagic Escherichia coli-induced diarrhea and intestinal

inflammation in infant rabbits. *Infect Immun* **71**:7129–39.

35. **Ritchie JM, Rui H, Bronson RT, Waldor MK**. 2010. Back to the future: studying cholera pathogenesis using infant rabbits. *MBio* **1**:e00047-10.

36. **Ritchie JM, Rui H, Zhou X, Iida T, Kodoma T, Ito S, Davis BM, Bronson RT, Waldor MK**. 2012. Inflammation and Disintegration of Intestinal Villi in an Experimental Model for Vibrio parahaemolyticus-Induced Diarrhea. *PLoS Pathog* **8**:e1002593.

37. **Porter CK, Lynen A, Riddle MS, Talaat K, Sack D, Gutiérrez RL, McKenzie R, DeNearing B, Feijoo B, Kaminski RW, Taylor DN, Kirkpatrick BD, Bourgeois AL**. 2018. Clinical endpoints in the controlled human challenge model for Shigella: A call for standardization and the development of a disease severity score. *PLoS One* **13**:e0194325.

38. **Gomez HF, Ochoa TJ, Herrera-Insua I, Carlin LG, Cleary TG**. 2002. Lactoferrin protects rabbits from Shigella flexneri-induced inflammatory enteritis. *Infect Immun* **70**:7050–3.

39. **Gomez HF, Ochoa TJ, Carlin LG, Cleary TG**. 2003. Human lactoferrin impairs virulence of Shigella flexneri. *J Infect Dis* **187**:87–95.

40. **Kadurugamuwa JL, Rohde M, Wehland J, Timmis KN**. 1991. Intercellular spread of Shigella flexneri through a monolayer mediated by membranous protrusions and associated with reorganization of the cytoskeletal protein vinculin. *Infect Immun* **59**:3463–71.

41. **Monack DM, Theriot JA**. 2001. Actin-based motility is sufficient for bacterial membrane protrusion formation and host cell uptake. *Cell Microbiol* **3**:633–47.

42. **Lecuit M, Vandormael-Pournin S, Lefort J, Huerre M, Gounon P, Dupuy C, Babinet C, Cossart P**. 2001. A transgenic model for listeriosis: Role of internalin in crossing the intestinal barrier. *Science (80)* **292**:1722–1725.

43. **Zhang T, Abel S, Abel Zur Wiesch P, Sasabe J, Davis BM, Higgins DE, Waldor MK**. 2017. Deciphering the landscape of host barriers to Listeria monocytogenes infection. *Proc Natl Acad Sci U S A* **114**:6334–6339.

44. **Maury MM, Bracq-Dieye H, Huang L, Vales G, Lavina M, Thouvenot P, Disson O, Leclercq A, Brisse S, Lecuit M**. 2019. Hypervirulent Listeria monocytogenes clones' adaption to mammalian gut accounts for their association with dairy products. *Nat Commun* **10**:2488.

45. **Lawley TD, Bouley DM, Hoy YE, Gerke C, Relman DA, Monack DM**. 2008. Host transmission of Salmonella enterica serovar Typhimurium is controlled by virulence factors and indigenous intestinal microbiota. *Infect Immun* **76**:403–16.

46. **Suwandi A, Galeev A, Riedel R, Sharma S, Seeger K, Sterzenbach T, García Pastor L, Boyle EC, Gal-Mor O, Hensel M, Casadesús J, Baines JF, Grassl GA**. 2019. Std fimbriae-fucose interaction increases Salmonella-induced intestinal inflammation and prolongs colonization. *PLoS Pathog* **15**:e1007915.

47. **Waugh DJJ, Wilson C**. 2008. The interleukin-8 pathway in cancer. *Clin Cancer Res* **14**:6735–41.

48. **Schnupf P, Sansonetti PJ**. 2012. Quantitative RT-PCR profiling of the rabbit immune response: assessment of acute Shigella flexneri infection. *PLoS One* **7**:e36446.

49. **Raqib R, Lindberg AA, Wretlind B, Bardhan PK, Andersson U, Andersson J**. 1995. Persistence of local cytokine production in shigellosis in acute and convalescent stages. *Infect Immun* **63**:289–296.

50. **Raqib R, Wretlind B, Andersson J, Lindberg AA**. 1995. Cytokine secretion in acute shigellosis is correlated to disease activity and directed more to stool than to plasma. *J Infect Dis* **171**:376–84.

51. **Puhar A, Tronchère H, Payrastre B, Tran Van Nhieu G, Sansonetti PJ**. 2013. A Shigella Effector Dampens Inflammation by Regulating Epithelial Release of Danger Signal ATP through Production of the Lipid Mediator PtdIns5P. *Immunity* **39**:1121–1131.

52. **Arbibe L, Kim DW, Batsche E, Pedron T, Mateescu B, Muchardt C, Parsot C, Sansonetti PJ**. 2007. An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses. *Nat Immunol* **8**:47–56.

53. **Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJ, Rubin EJ, Waldor MK**. 2014. ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet* **10**:e1004782.

54. **Hubbard TP, Billings G, Dörr T, Sit B, Warr AR, Kuehl CJ, Kim M, Delgado F, Mekalanos JJ, Lewnard JA, Waldor MK**. 2018. A live vaccine rapidly protects against cholera in an infant rabbit model. *Sci Transl Med* **10**:1–11.

55. **Hubbard TP, Chao MC, Abel S, Blondel CJ, Abel zur Wiesch P, Zhou X, Davis BM, Waldor MK**. 2016. Genetic analysis of *Vibrio parahaemolyticus* intestinal colonization. *Proc Natl Acad Sci* **113**:6283–6288.

56. **Warr AR, Hubbard TP, Munera D, Blondel CJ, Abel Zur Wiesch P, Abel S, Wang X, Davis BM, Waldor MK**. 2019. Transposon-insertion sequencing screens unveil requirements for EHEC growth and intestinal colonization. *PLoS Pathog* **15**:e1007652.

57. **Schuch R, Maurelli AT**. 1999. The Mxi-Spa type III secretory pathway of Shigella flexneri requires an outer membrane lipoprotein, MxiM, for invasin translocation. *Infect Immun* **67**:1982–1991.

58. **Schuch R, Maurelli AT**. 2001. MxiM and MxiJ, base elements of the Mxi-Spa type III secretion system of Shigella, interact with and stabilize the MxiD secretin in the cell envelope. *J Bacteriol* **183**:6991–6998.

59. **Burkinshaw BJ, Strynadka NCJ**. 2014. Assembly and structure of the T3SS. *Biochim Biophys Acta - Mol Cell Res* **1843**:1649–1663.

60. **Buffie CG, Bucci V, Stein RR, McKenney PT, Ling L, Gobourne A, No D, Liu H, Kinnebrew M, Viale A, Littmann E, van den Brink MRM, Jenq RR, Taur Y, Sander**

C, Cross JR, Toussaint NC, Xavier JB, Pamer EG. 2015. Precision microbiome reconstitution restores bile acid mediated resistance to Clostridium difficile. *Nature* **517**:205–8.

61. **Becattini S, Littmann ER, Carter RA, Kim SG, Morjaria SM, Ling L, Gyaltshen Y, Fontana E, Taur Y, Leiner IM, Pamer EG**. 2017. Commensal microbes provide first line defense against Listeria monocytogenes infection. *J Exp Med* **214**:1973–1989.

62. **Zhao W, Caro F, Robins W, Mekalanos JJ**. 2018. Antagonism toward the intestinal microbiota and its effect on Vibrio cholerae virulence. *Science (80)* **359**:210–213.

63. **Formal SB, Kent TH, May HC, Palmer A, Falkow S, LaBrec EH**. 1966. Protection of monkeys against experimental shigellosis with a living attenuated oral polyvalent dysentery vaccine. *J Bacteriol* **92**:17–22.

64. **Formal SB, Oaks E V., Olsen RE, Wingfield-Eggleston M, Snoy PJ, Cogan JP**. 1991. Effect of prior infection with virulent shigella flexneri 2a on the resistance of monkeys to subsequent infection with shigella sonnei. *J Infect Dis* **164**:533–537.

65. **Coster TS, Hoge CW, VanDeVerg LL, Hartman AB, Oaks E V., Venkatesan MM, Cohen D, Robin G, Fontaine-Thompson A, Sansonetti PJ, Hale TL**. 1999. Vaccination against shigellosis with attenuated Shigella flexneri 2a strain SC602. *Infect Immun* **67**:3437–43.

66. **Porter CK, Thura N, Ranallo RT, Riddle MS**. 2013. The Shigella human challenge model. *Epidemiol Infect* **141**:223–232.

67. **Brotcke Zumsteg A, Goosmann C, Brinkmann V, Morona R, Zychlinsky A**. 2014. IcsA is a Shigella flexneri adhesion regulated by the type III secretion system and required for pathogenesis. *Cell Host Microbe* **15**:435–445.

68. **Abel S, Abel zur Wiesch P, Davis BM, Waldor MK**. 2015. Analysis of Bottlenecks in Experimental Models of Infection. *PLoS Pathog* **11**:e1004823.

69. **Wei L, Qiao H, Sit B, Yin K, Yang G, Ma R, Ma J, Yang C, Yao J, Ma Y, Xiao J, Liu X, Zhang Y, Waldor MK, Wang Q**. 2019. A Bacterial Pathogen Senses Host Mannose to Coordinate Virulence. *iScience* **20**:310–323.

70. **Barthel M, Hapfelmeier S, Quintanilla-Martínez L, Kremer M, Rohde M, Hogardt M, Pfeffer K, Rüssmann H, Hardt W-D**. 2003. Pretreatment of mice with streptomycin provides a Salmonella enterica serovar Typhimurium colitis model that allows analysis of both pathogen and host. *Infect Immun* **71**:2839–58.

71. **Holt KE, Baker S, Weill FX, Holmes EC, Kitchen A, Yu J, Sangal V, Brown DJ, Coia JE, Kim DW, Choi SY, Kim SH, Da Silveira WD, Pickard DJ, Farrar JJ, Parkhill J, Dougan G, Thomson NR**. 2012. Shigella sonnei genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet* **44**:1056–1059.

72. **Connor TR, Barker CR, Baker KS, Weill FX, Talukder KA, Smith AM, Baker S, Gouali M, Thanh DP, Azmi IJ, da Silveira WD, Semmler T, Wieler LH, Jenkins C, Cravioto A, Faruque SM, Parkhill J, Kim DW, Keddy KH, Thomson NR**. 2015.

Species-wide whole genome sequencing reveals historical global spread and recent local persistence in Shigella flexneri. *Elife* **4**:1–16.

73. **Puzari M, Sharma M, Chetia P**. 2018. Emergence of antibiotic resistant Shigella species: A matter of concern. *J Infect Public Health* **11**:451–454.

74. **Abbasi E, Abtahi H, van Belkum A, Ghaznavi-Rad E**. 2019. Multidrug-resistant Shigella infection in pediatric patients with diarrhea from central Iran. *Infect Drug Resist* **12**:1535–1544.

75. **Datsenko KA, Wanner BL**. 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**:6640–5.

76. **Abel S, Waldor MK**. 2015. Infant Rabbit Model for Diarrheal Diseases. *Curr Protoc Microbiol* **38**:6A.6.1-15.

77. **Chiang SL, Rubin EJ**. 2002. Construction of a mariner-based transposon for epitope-tagging and genomic targeting. *Gene* **296**:179–85.

78. **Chao MC, Pritchard JR, Zhang YJ, Rubin EJ, Livny J, Davis BM, Waldor MK**. 2013. High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Res* **41**:9033–9048.

79. **Ferrières L, Hémery G, Nham T, Guérout A-M, Mazel D, Beloin C, Ghigo J-M**. 2010. Silent mischief: bacteriophage Mu insertions contaminate products of Escherichia coli random mutagenesis performed using suicidal transposon delivery plasmids mobilized by broad-host-range RP4 conjugative machinery. *J Bacteriol* **192**:6418–27.

80. **Kimura S, Waldor MK**. 2019. The RNA degradosome promotes tRNA quality control through clearance of hypomodified tRNA. *Proc Natl Acad Sci* **116**:1394–1403.

Chapter 4

Discussion

In this thesis, I have described our work on the virulence mechanisms of several bacterial pathogens enabled by high-throughput next generation sequencing (NGS) approaches and the development of a small animal model of infection. Research on host-pathogen interactions and microbial pathogenesis increasingly relies on interdisciplinary approaches to identify and characterize various aspects of infection. I have specifically demonstrated how through using a multi-modal approach we have identified novel virulence factors in an important pathogen of domesticated animals and described a relevant infection model for the leading bacterial cause of diarrheal deaths worldwide. Our findings have given rise to numerous additional interesting questions and we are hoping to pursue these areas of research in the future. The remainder of this chapter summarizes the results and places them in the larger context of their respective fields by addressing some future directions for each of the projects described in the previous chapters.

Virulence regulation in the *S. equi* group

In Chapter 2, I described how we discovered a novel, conserved virulence regulator in a key pathogen of farmed animals, *Streptococcus equi* subspecies *zooepidemcus* (SEZ). We also described an unusual reactivity of an antibody that recognized a surface polysaccharide, poly-N-acetylglucosamine (PNAG), with a highly prevalent cell-wall anchored surface protein in SEZ, SzM, which is a member of the M/M-like protein family. Through a combination of multiple NGS technologies, including transposon insertion sequencing (Tn-seq, TIS), whole genome sequencing (WGS), and RNA-sequencing (RNA-seq), we identified and characterized a novel virulence regulator, which we termed Sez Virulence (SezV), that activated expression of a neighboring gene that encoded SzM (Figure 4.1). We found that SezV primarily activated

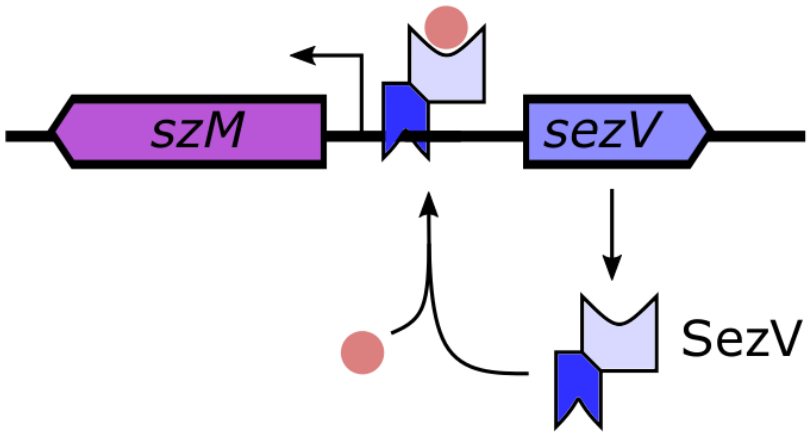


Figure 4.1. Putative model of SezV activitation of *szM*.

Putative model for SezV activation of transcription of *szM*. SezV contains two domains, a ligand binding domain (light blue) and a helix-turn-helix domain (dark blue), which likely binds DNA. In this proposed model, binding to a ligand (red circle), whose identity is unknown, is required for binding of SezV to the promoter of *szM*.

expression of only one gene, *szM*, and repressed expression of genes in a specific pilus operon. Using a mouse model of infection, we found that both *sezV* and *szM* mutant strains were attenuated in virulence and had reduced colonization of the brain, a primary site of infection in the model. The survival curves between the two mutant strains were not statistically different, which may suggest that the loss of virulence in the *sezV* mutant was due to absent expression of *szM*. Through large-scale genomic analyses of publicly available genomes and new genomes of strains that we sequenced, we found that the *sezV-szM* locus was restricted to strains of the *S. equi* group (which includes SEZ and *Streptococcus equi* subspecies *equi* [SEE]) and group A streptococci (GAS). Furthermore, the locus was found to be conserved in every strain of the *S. equi* group and in a small fraction of GAS strains, notably all strains of the M18 type contained the locus.

We also discovered that *szM* expression was required for whole bacterial reactivity to an anti-PNAG antibody, and subsequently found through biochemical experiments that the anti-PNAG antibody directly recognized the SzM protein. This led us to explore whether PNAG decorated SzM, and selective ablation of PNAG through chemical or enzymatic means led to a loss of reactivity of SzM with the anti-PNAG antibody, providing evidence for a glycosylation of SzM with PNAG (Figure 4.2). However, glycan mass spectrometry did not detect any glycans liberated from SzM, and thus we could not definitively conclude that SzM was glycosylated.

Hence, our findings extended our knowledge of virulence regulation in streptococci and the immunochemical properties of M-like proteins, which are important virulence factors in streptococci (1). These discoveries have led to several additional questions regarding regulation

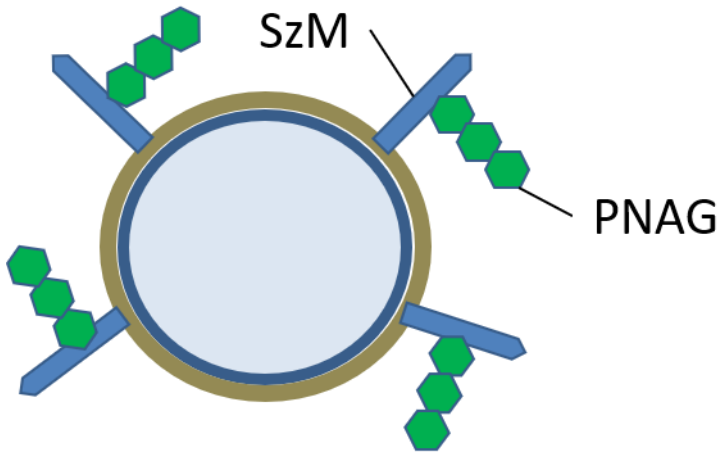


Figure 4.2. Putative model of PNAG association with SzM in SEZ bacteria.
Putative model of how a PNAG-like polysaccharide (possibly a shorter oligosaccharide form) is linked to the SzM protein, which is anchored to the cell wall in SEZ. We do not yet have direct evidence to support glycosylation of SzM as glycan mass spectrometry did not detect any glycans on SzM, but immunochemical experiments provide support for a glycan modification. Brown ring: cell wall; blue ring: cell membrane; light blue circle: cytoplasm.

of virulence in SEZ and the role of specific surface proteins.

Upon describing SezV, a previously uncharacterized protein, we found that it contained two predicted domains – a ligand binding domain (LBD) similar to those found in sensor kinases and an AraC type helix-turn-helix (HTH) domain. We hypothesized that the HTH domain bound DNA, and in our work found evidence that SezV is a virulence regulator. The function of the LBD is unclear and it is unknown whether it even binds a ligand. The identity of any putative ligand remains to be discovered as well as the activities of the bound and un-bound forms of SezV. A potential ligand that modulates the activity of a transcriptional regulator would suggest that SezV activity can be post-translationally regulated. However, it is not known whether SezV activity is regulated at the transcriptional level as well, e.g. via auto-regulation or by other, unknown regulators. Future work can serve to illuminate the dynamics of the expression and activity of *sezV* in vivo using small animal models of infection to understand the spatial (i.e. specific tissue or organ) and temporal (i.e. kinetics) regulation of SezV activity, as has recently been done for in vivo virulence regulation of other pathogens (2).

One of the surprising findings we made when studying the function of SezV was that in its presence, the expression of a specific pilus operon in SEZ was severely repressed. The operon has been previously shown to produce a pilus (3, 4), and the major pilus subunit has been utilized as a component in a candidate vaccine (5). While the role of this pilus in vivo is unknown, we hypothesized that its expression may hinder widespread dissemination by constraining the bacteria to adhere to specific tissues and promote commensal colonization (Figure 4.3). Once *sezV* is activated, reduced expression of the pilus may facilitate development of invasive disease. It is likely that the expression dynamics of the pilus are the inverse of *sezV* dynamics in vivo. The expression of the pilus may increase bacterial adherence to specific tissues and prevent

widespread dissemination. Thus, it would be interesting to explore whether constitutive expression of the pilus operon, for example through expression from a stable plasmid or promoter replacement, prevents dissemination in an animal model.

M and M-like proteins are important virulence factors in pyogenic streptococci such as *S. pyogenes* and *S. equi* group strains. *S. equi* strains are thought to contain only two M-like proteins, SzM/SeM and SzP/SzPSe. Both play roles in the virulence of *S. equi* strains (6–8). In *S. pyogenes*, all the M/M-like proteins are regulated by a single protein, Mga (multiple gene activator), which is found in the same locus as the M/M-like proteins (9). In *S. equi* strains, prior to our work, no regulator for either of the two M-like proteins (SzM or SzP) had been discovered. We discovered that SezV activates expression of SzM and that they are genetically linked. Surprisingly, SezV does not affect expression of SzP and the gene encoding SzP is located at a site in the genome that is distant from the *sezV* locus. Hence, we discovered that regulation of M-like proteins in SEZ, which is likely conserved in SEE, is not coordinated by a single activator. Furthermore, it is also likely that in strains of GAS that contain *sezV* and an *szM* ortholog, Mga does not activate the *szM* ortholog. An unresolved question is the identity of the regulator of SzP expression. Similar to the activity of SezV, the function and activity of the SzP regulator in vivo will be exciting to explore. During our genomic analyses, we have also surprisingly discovered that strains of the *S. equi* group contain a single *mga* homolog, and the gene is not located near *szM* nor *szP* in the genome. It is interesting to consider the possible role of the *mga* homolog in *S. equi* group strains and implications for understanding the evolution of the *S. equi* group and *S. pyogenes*, which is a streptococcal species closely related to *S. equi* (10).

Figure 4.3. Putative model of SEZ pilus regulation and pathogen spread.

In the proposed model, SEZ bacteria can exist in one of two states.

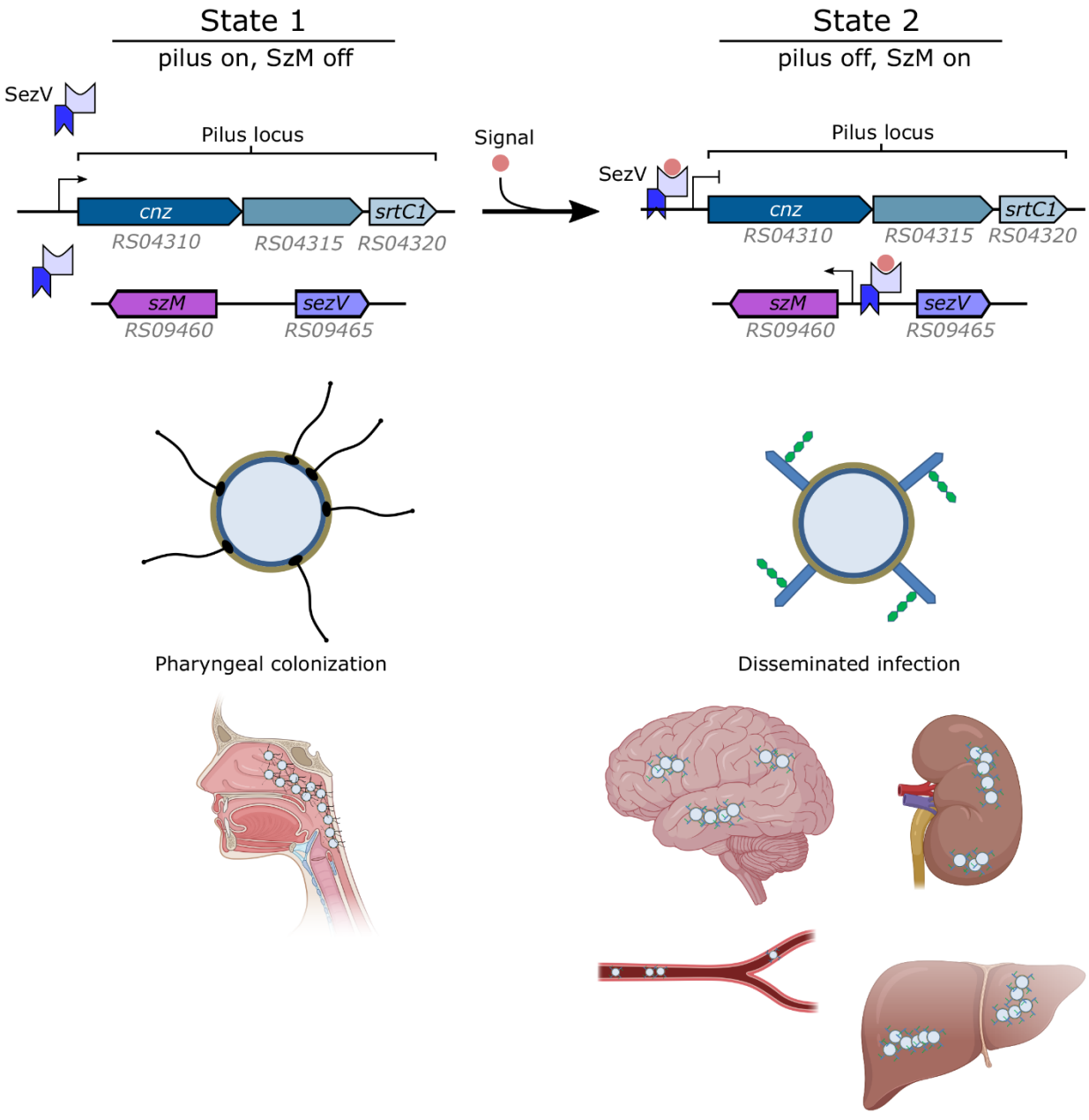
In state 1, SezV activity is off (cannot bind DNA) so SzM is not expressed while the pilus is expressed, since SezV cannot repress expression of the pilus locus. This restricts the pathogen's colonization to that of the animal's nasopharynx and oropharynx.

In state 2, SezV activity is on (binds DNA) so SzM is expressed while the pilus is not expressed, since SezV is bound to its ligand and represses expression of the pilus locus. The absence of the pilus permits bacteria to spread, via the bloodstream, and seed other organs such as the brain, kidneys, and liver.

In this model, SezV activity requires binding to a ligand. The signal that switches the SEZ bacteria from state 1 to state 2 may be a potential ligand for SezV, e.g. a host-derived small molecule or metabolite.

Pilus locus spans from *cnz* to *srtC1*; Cnz: collagen binding protein in SEZ and homolog of Cne, SrtC1: sortase C. Circles in middle represent SEZ bacteria: larger brown ring: cell wall; smaller blue ring: cell membrane; smaller light blue circle: cytoplasm; black wavy lines: pili; blue rods decorated with green hexagons: SzM protein.

Figure 4.3 (Continued)



Advancing small animal models of *Shigella* infection

In Chapter 3, I described the development of an oro-gastric small animal model of *Shigella* infection that utilized infant rabbits. In this project, we demonstrated that two to three-day-old infant rabbits could serve as useful models for *Shigella* infection, since infected animals developed disease, e.g. diarrhea, as well as intestinal pathology, and were robustly colonized by the pathogen (Figure 4.4). Surprisingly, we found that infant rabbits fed ad libitum prior to inoculation were relatively resistant to being colonized and developing disease, however, fasting rabbits prior to inoculation led to more robust development of disease and colonization. In the colon, infected animals developed pathology reminiscent of aspects of pathology seen in human infections, such as epithelial cell sloughing and edema. Notably, animals infected with the WT strain did not recruit heterophils (rabbit neutrophil equivalent) to the intestine; human infections are marked by substantial recruitment of neutrophils to colon. Immunofluorescence microscopy of tissue from infected animals demonstrated large areas of infected epithelial cells in the colon and substantial bacteria in the intestinal lumen. With three-dimensional confocal microscopy of tissue sections, we confirmed intracellular localization of *S. flexneri* bacteria within epithelial cells in the colon. Furthermore, by staining polymerized actin, we visualized actin tails at the poles of intracellular bacteria and the formation of membrane protrusions from primary infected cells into neighboring uninfected cells. Both these observations suggest that the pathogen undergoes intracellular movement in the infected cell and is spreading from cell-to-cell among colonic epithelial cells. We also explored the host innate immune response to infection by performing in situ measurements of the expression level of a key cytokine involved in acute inflammatory responses, the CXC chemokine IL-8 (11). We found elevated expression of IL-8

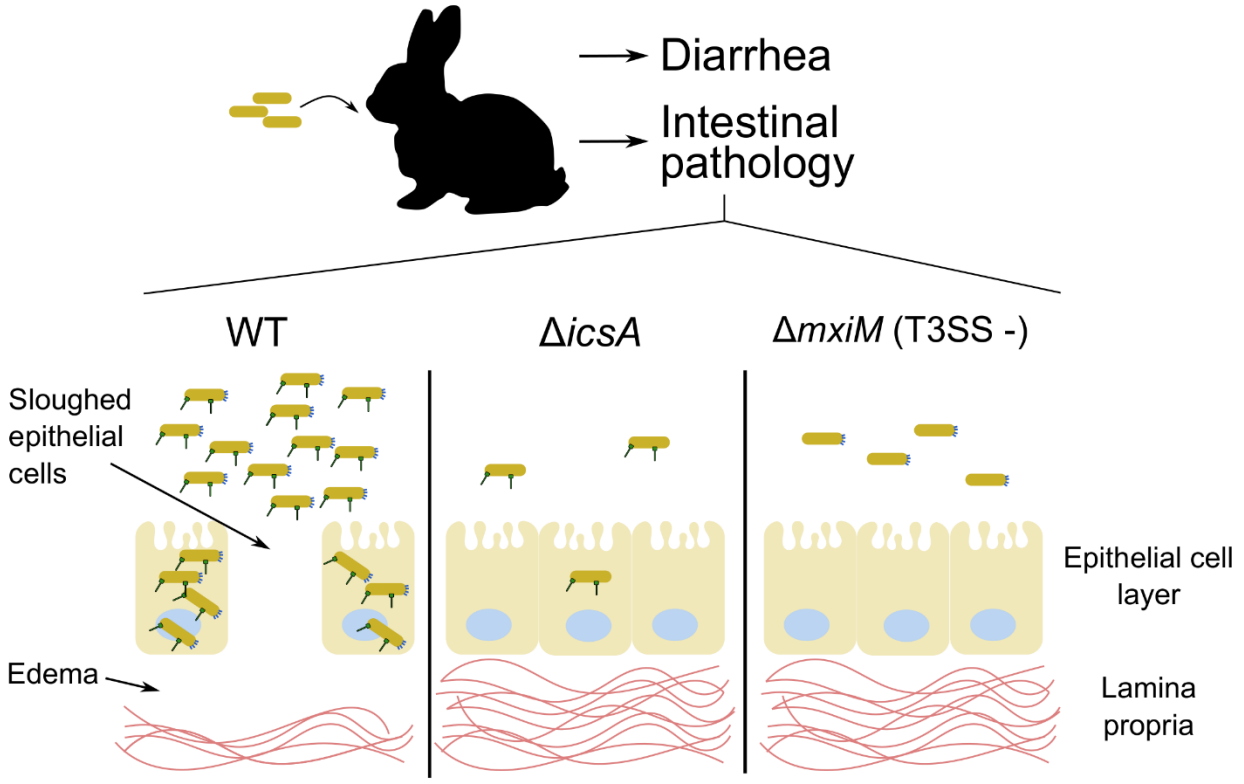


Figure 4.4. Model for oral inoculation of infant rabbits with *S. flexneri*.

Oro-gastric inoculation with *S. flexneri* bacteria (yellow) results in diarrhea and intestinal pathology in the colon. Infection with the WT strain, which expresses the T3SS and IcsA at body temperature (37°C), robustly colonizes the colon, invades epithelial cells, and causes pathology (edema and sloughing of epithelial cells). The Δ*icsA* mutant has reduced colonizing capacity, does not cause pathology, and was very rarely found within colonic epithelial cells. The Δ*mxiM* mutant, which represents a T3SS negative (T3SS -) strain, had reduced intestinal colonization, did not cause pathology, and was not found in an intracellular location. Red wavy lines represent connective tissue (e.g. collagen) in lamina propria.

mRNA near infected epithelial cells and that while infected epithelial cells were less likely to express IL-8, neighboring uninfected cells more frequently expressed the cytokine.

In addition, we demonstrated that two key virulence factors in *S. flexneri*, IcsA, a protein required for cell-to-cell spreading and adherence to host cells, and the type III secretion system (T3SS) are required for disease, pathology, and intestinal colonization (Figure 4.4). Surprisingly, although mutants in IcsA and the T3SS did not induce epithelial cell sloughing and edema, they did recruit heterophils to the intestine. Comparison of intestinal colonization of the IcsA mutant strain with the WT strain demonstrated that there were statistically significant differences between colonization in the small and large intestine. For the T3SS mutant strain, comparison with the WT strain only demonstrated a statistically significant difference in the small intestine. However, direct comparison of the IcsA mutant strain with the T3SS mutant strain did not demonstrate a statistically significant difference in colonization between the two strains in either the small or the large intestine. The absence of IcsA also led to loss of infection foci, and in animals infected with the IcsA mutant strain, we only found rare isolated epithelial cells containing intracellular bacteria. Absence of the T3SS prevented the bacteria from entering epithelial cells in vivo.

We attempted to perform in vivo Tn-seq with *S. flexneri* to identify additional genes required for colonization and pathogenesis but were limited by a narrow bottleneck for colonization of the colon (Figure 4.5). Thus, the random loss in genetic diversity in the library caused by the narrow bottleneck precluded any further identification of additional genes required for colonization. Nevertheless, with modifications to the mutant library or inoculation protocol, high-throughput identification of new genes involved in *S. flexneri* colonization of the intestine may become feasible.

Overall, our work addressed a key challenge in the field of *Shigella* pathogenesis – translating findings made in tissue cultured cells, in which the lifecycle of the bacteria and interactions with the host cell have been described in much detail, to in vivo models of infection.

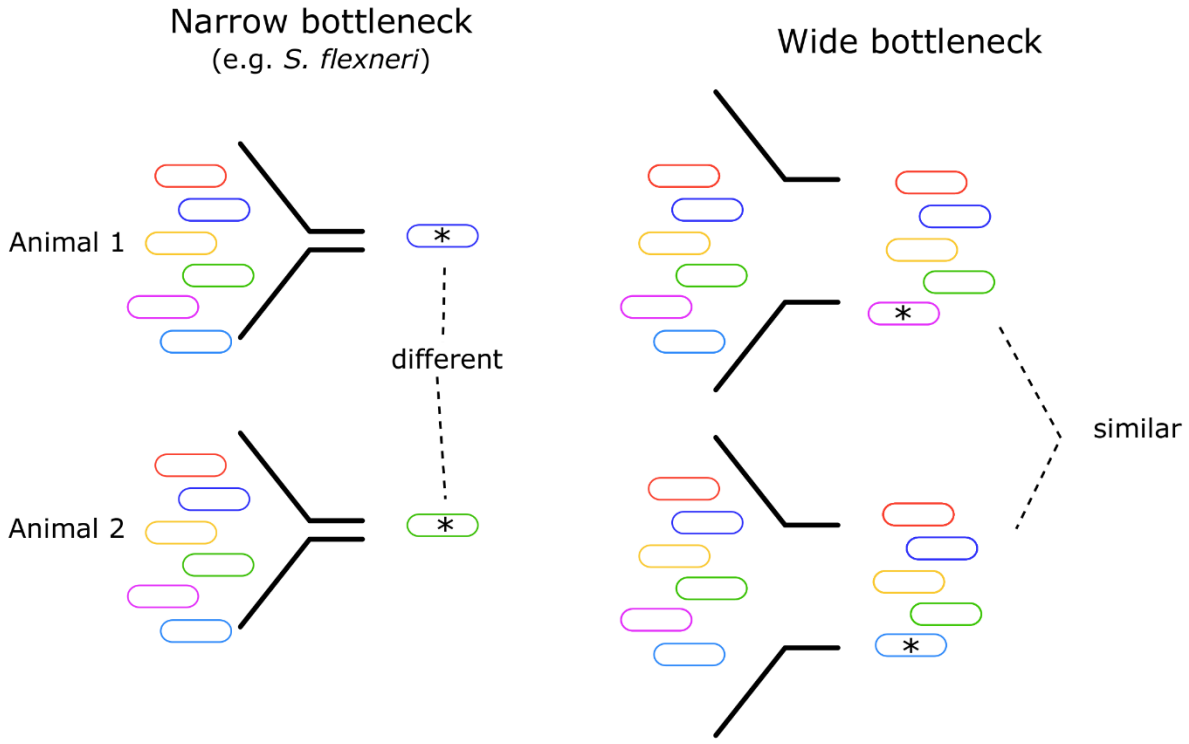


Figure 4.5. Model of magnitude of bottlenecks in intestinal colonization.

Bottlenecks result in a stochastic loss of genetic diversity in a population. Initial seeding and colonization of the intestine requires passage through a bottleneck. (Left) When a transposon (Tn) insertion library (multicolored bacteria) is used to inoculate an animal, narrow bottlenecks result in loss of most Tn mutants. Hence, the mutants recovered from replicate experiments, e.g. two animals, are different. This makes it challenging to determine whether Tn mutants are lost due to the bottleneck or selective pressure in the intestine. (Right) In contrast, in wide bottlenecks, most of the Tn mutants are able to seed the intestine and any subsequent loss of Tn mutants can be attributed to selection in the intestine.

Oral inoculation of infant rabbits may prove to be a versatile tool for studies on *Shigella* pathogenesis and we expect it to be a useful platform for future research and drug development.

During our experiments, we found that rabbits infected with the WT strain developed disease, such as diarrhea, reduced body temperature, and lack of weight gain. There are several possible physiological and behavioral explanations for these clinical signs. Animals likely develop diarrhea due to dysfunction and destruction of colonic epithelial cells that leads to disruption of the normal functioning of the colon, which is largely to reabsorb water. This colonic tissue dysfunction results in liquid feces and diarrhea, as less water will be reabsorbed from the feces. Reduced body temperature in animals may arise from reduced metabolism in infected animals secondary to nutritional deficiency caused by intestinal dysfunction, behavioral adaptations reducing huddling with other kits or the dam leading to a reduction in warmth gained from other animals, or neurogenic causes such as an altered central thermostat. Animals without disease gained weight over the course of the experiment; the lack of weight gain in diseased animals could be due to water loss from diarrhea and/or behavioral changes reducing feeding frequency.

We also found that some animals succumbed to infection, which we think is likely due to bacteremia with the infecting pathogen, *S. flexneri*. In preliminary animal experiments, we had checked whether *S. flexneri* was found in the blood of inoculated animals by measuring bacterial burden in the blood. We did not find *S. flexneri* in the blood of infected animals, regardless of whether they had developed diarrhea or not, suggesting that in the infant rabbit model *S. flexneri* does not typically enter the bloodstream and does not result in a disseminated infection. These results are similar to those observed in human infections, where *Shigella* bacteremia and disseminated infections are not characteristics of the disease and are rare. The pathogen's

susceptibility to complement (12) as well as destruction by neutrophils (13) may explain why the pathogen cannot survive in the blood. However, in some of the few cases in our animal model where infection resulted in mortality, we tested for bacterial burden in the blood and found evidence for bacteremia.

In our project, we found that a majority, but not all rabbits infected with *S. flexneri* developed diarrhea; there are numerous risk factors that may affect the development of diarrhea. Infant rabbits have a microbiome with some diversity (14), and in other animals, the intestinal microbiome has been shown to influence the development of disease during infection with other enteric bacterial pathogens. Hence, it is possible that differences in the intestinal microbiota of infant rabbits, e.g. among litters, may alter the probability for developing disease. Quality of dam care and frequency of feeding of kits also may affect survival, as experiments we performed as well as other published reports suggest that components in milk may inhibit colonization of the bacteria. Furthermore, adequate feeding ensures that kits obtain nutrients that are required both for basal growth and to mount productive responses to infections. Lack of feeding will also directly contribute to reduced weight gain in the kits. We also found that, as a group, animals that developed disease had statistically significantly higher bacterial burdens of *S. flexneri* in the distal small intestine and colon; however, there was a wide spread of intestinal bacterial burden such that, at the individual level, some infected animals without disease had higher absolute burdens than animals with disease. We believe that intestinal bacterial burden is likely another component that can increase the likelihood of disease development. One of the better predictors of disease was the body temperature at 36 hours post inoculation, as both (1) almost all animals with disease had low body temperatures and (2) the majority of animals with low temperatures had disease. Overall, on their own, each of the factors discussed here as well as other unknown

factors likely independently and variably increase the likelihood of developing disease upon infection. Thus, animals that do not develop disease may still have some aspects of infection that are traditionally seen in animals with disease, i.e. animals may be highly colonized with *S. flexneri* but will not develop disease such as diarrhea. It is likely that the presence of several risk factors for disease, some of which were not measured in our project, may be required to increase the probability for disease to a high level. Identifying the factor(s) that contribute to disease in infant rabbits could help improve the reproducibility of the model as well as deepen our understanding of infections in humans, for example by aiding in the identification of people who may have a high probability for developing severe disease and by informing treatment or preventative measures for *Shigella* infections.

While our model examined the virulence properties of *S. flexneri*, which is the species responsible for the greatest percentage of cases worldwide, the prevalence of another species, *S. sonnei*, is rising. The majority of infections in developed nations are attributable to *S. sonnei*, but more recently it is becoming a common cause of infections in transition nations and of travelers' diarrhea (15). *S. sonnei* has unique features when compared with *S. flexneri*, such as a less stable virulence plasmid with slightly different genetic content (16) and the presence of an active type VI secretion system (T6SS) (17). Due to its growing importance, testing whether this species also leads to disease in infant rabbits would be valuable for future research, particularly since the majority of *Shigella* research over the past few decades has focused on *S. flexneri*. In addition, very little research has and is being performed on *S. dysenteriae*, which is a rare cause of sporadic infections, but still occasionally causes deadly outbreaks due to its potential to encode Shiga toxin. It is known that Shiga toxin causes pathology and diarrhea in the infant rabbit model, based on work performed using Enterohemorrhagic *E. coli* (EHEC), which encodes two

Shiga-like toxins (18). However, the interplay between *Shigella* infection and Shiga toxin in pathology and diarrhea in the infant model remains to be explored. In addition to analyzing differences in disease and pathogenesis among each *Shigella* species, the infant rabbit model may provide a straightforward approach to evaluate the relative virulence of different strains of a single species. Such an approach is commonly used to test the virulence of strains of other diarrheal pathogens, such as *V. cholerae* (19) and EHEC (20). We only tested one strain in the infant model, *S. flexneri* serotype 2a strain 2457T, a strain that has been intensively researched and is commonly used as a challenge strain in human studies on vaccine research (21). Since multiple serotypes of *S. flexneri* and other species are currently circulating, multiple challenge strains are required for testing the efficacy of a vaccine. Furthermore, for live-attenuated vaccines, the starting virulent strain chosen should be sufficiently immunogenic to ensure the vaccine stimulates a strong, protective immune response. One concern in *Shigella* vaccine research has been the choice of a parental strain for developing a vaccine that is inadequately virulent, as some strains are substantially less virulent and immunogenic in humans. The use of a weakly virulent strain has led to an unintentional failure of a vaccine candidate, which could have been avoided by pre-screening strains for virulence (22, 23). Overall, the infant rabbit model could provide a potential platform to test both the development of disease and virulence of various candidate strains.

Although we were unsuccessful in conducting in vivo Tn-seq with *S. flexneri* in the oral infant rabbit model, the appeal of studying the role of bacterial genes in colonization and pathogenesis remains attractive. For example, while we found that two key virulence factors in *S. flexneri*, IcsA and the T3SS, were required for development of disease (e.g. diarrhea and intestinal pathology) and intestinal colonization, the contribution of individual T3SS secreted

effectors in each of these aspects of *Shigella* infection is unknown. Presumably with modifications, Tn-seq or similar genome-scale methodologies (e.g. CRISPR interference [CRISPRi]) (24, 25) can be utilized in the oral inoculation model to perform high-throughput genetic analysis of *S. flexneri* colonization. Testing of additional *Shigella* strains in the infant rabbit model could also permit genome-scale analyses with these strains to identify species or strain specific dependencies in vivo that may provide insights into unique physiological differences among strains and species and/or the development of novel, targeted therapeutics, e.g. for multi-drug resistant strains.

A surprising observation from our study was that certain mutant strains of *Shigella* recruited large numbers of heterophils, a key immune cell of the acute inflammatory response, to the intestine of infected animals. However, these strains did not induce high levels of the CXC family chemokine IL-8, which is considered a primarily heterophil/neutrophil chemotactic factor. These results suggested that during *Shigella* infection, there are other chemotactic factors that recruit heterophils to the intestine and that IL-8 does not directly lead to heterophil recruitment. Additional factors may include canonical neutrophil chemokines, such as complement component C5a (26), or previously undescribed host or bacterial chemokines. Another unexpected finding in our work was that the wild type (WT) *Shigella* strain did not robustly recruit heterophils to the intestine. This is in contrast with both histological reports of human biopsies taken during *Shigella* infection, which display large numbers of neutrophils (27, 28), and most other animal models of *Shigella* infection, including the intra-rectal infant rabbit model, in which WT infection induces massive heterophil recruitment to the intestine (29). We do not have a good reason for why the WT strain does not induce heterophil recruitment in the

oral inoculation infant rabbit model. However, our results may point to an interesting, unique biological finding about the nature of host-pathogen interactions in the infant rabbit intestine.

Our oral infant rabbit model also provides a platform to investigate numerous aspects of *Shigella* pathogenesis, prevention, and treatment. Our oral model could be used to study the viability of new therapeutics as well as to better describe the function and outcomes after treatment with existing therapeutics. While most *Shigella* infections are self-resolving and only require supportive care without chemotherapy, in certain patient populations or to reduce the duration of the disease, patients are administered antibiotics (30). Use of an animal model could enable a much higher resolution analysis of changes in clinical outcome upon antibiotic treatment and any differences in outcome attributable to different types of antibiotics. For example, an animal model could enable a kinetic analysis of bacteria clearance and sites of relative resistance in the intestine, and offer a detailed view of the resolution of pathology after infection. In addition, the oral infant model provides a useful in vivo model to test the effect of novel anti-bacterial compounds as well as to compare efficacy to existing compounds. Furthermore, the model could be used to develop new types of non-destructive (i.e. not bacteriostatic or bacteriocidal) anti-bacterial agents and therapeutics, such as those targeting virulence factor expression, masking virulence factor binding sites, or interfering with intestinal colonization. Apart from therapeutic development and testing, the oral model also provides a useful system to investigate *Shigella* colonization of the intestine. For example, the role of virulence plasmid and chromosomal genes in intestinal colonization could be monitored. The oral model will also likely be suitable for in vivo genetic screens of factors required for establishment and maintenance of colonic colonization. While we found that an in vivo genetic screen using a large, pooled library of random transposon mutants of *S. flexneri* was not

technically feasible, we believe that modifications of the inoculation protocol or transposon library will facilitate an in vivo genetic screen. For example, we could construct a defined, arrayed transposon library of smaller size and inoculate animals with the entire library or subpools of the arrayed transposon library. Alternatively, we could use CRISPR interference (CRISPRi) to conduct a genetic screen, as CRISPRi has recently been shown to be useful for conducting genetic screens in bacteria (24, 25), and library size can be precisely controlled by the user with this technology. Unlike other animal models of *Shigella* infection, our oral model provides an opportunity to accurately model the natural infection process involving passage through the upper gastrointestinal tract (e.g. esophagus and stomach), and seeding of the colon. Hence, our oral model can be used to study genes that may affect passage through the gastrointestinal tract. Furthermore, our detection of *Shigella* infection of colonic epithelial cells after oral inoculation can be used to study aspects of the intracellular *Shigella* lifecycle in a more native context, as most previous work has been done in tissue culture cells. In our model we detected numerous infectious foci forming in the colon of infected animals as well as evidence of cell-to-cell spreading, the formation of actin tails, and replication within colonic epithelial cells. Hence, infection results in a substantial number of sites of infected colonic tissue that would allow collection of sufficient amounts of data (for example with a more automated imaging processing platform) to permit a more detailed quantitative analysis of various aspects of infection of epithelial cells. This could include analysis of basic aspects of epithelial cell infection such as the number of bacteria per cell and the number of epithelial cells in an infection focus, as well as analysis of host proteins or mRNA (by using antibodies or in situ mRNA hybridization) whose levels or localization are altered during infection. Finally, the oral model provides a platform to study the effect of *Shigella* infection not only in the colon and rectum, but

also in the small intestine. In the oral model, the small intestine is highly colonized but does not display the pathology seen in the colon. However, whether infection alters the normal function of the small intestine remains unknown.

Recently, an intra-rectal model of infection of infant rabbits has been reported (29). Unlike our oral model, the intra-rectal model focuses on pathology and disease in the rectum because the bacteria are directly inoculated into this location. The intra-rectal model has higher reproducibility than the oral model, which is likely due to direct, rapid inoculation of the rectum with a high dose of *Shigella*, and animals in the intra-rectal model develop grossly bloody diarrhea. Furthermore, the rectal pathology observed in the intra-rectal model may be more representative (than the oral model) of the colonic pathology of human patients, as the infected animals develop ulcerations, influxes of heterophils (neutrophil equivalents), and elevated expression of proinflammatory cytokines (e.g. IL-1β, TNF-α, and IL-8). Hence, the intra-rectal model could be used to study the effect of various bacterial genes on the development of pathology and diarrhea. It is unclear why the intra-rectal model results in pathology different from that seen in the oral model. Unlike the oral model, the intra-rectal model cannot be used to study the transit of the bacteria through the gastrointestinal tract and may not be an ideal model to study intestinal colonization. Furthermore, the intra-rectal model requires involved, specialized animal husbandry that may limit its widespread adoption and use. Finally, due to the requirement to infect infant animals, neither the intra-rectal nor oral infant rabbit model can be used to study adaptive immune responses to infection and long-term protective immunity conferred by a vaccine.

Despite the existence of many useful animal models for *Shigella* infection, the field of *Shigella* pathogenesis still faces a yet insurmountable hurdle. The key research priority is the

development of a broad-spectrum *Shigella* vaccine, particularly to prevent infections with multi-drug resistant strains. A prerequisite for development of such a vaccine will likely be a relevant small animal model that contains a mature immune system and can be used to develop and test vaccine candidates. Unfortunately, no such small animal model currently exists. While we and others have recently demonstrated that infant rabbit models are useful small animal models of disease, the small window in which they can be infected and lack of a mature immune system limits their applicability to the study of *Shigella* pathogenesis and non-vaccine therapeutics. A model to test vaccine candidates typically requires an individual animal to be immunized and subsequently challenged with a virulent strain several days to weeks post immunization. The current model utilized for vaccine experiments is the pulmonary adult mouse model, in which intranasal delivery of virulent bacteria results in a lethal pneumonitis, with some histologic similarities to intestinal infections, since the lungs are also mucosal tissues (31, 32). While this model takes advantage of a robust phenotype and the power of mouse genetics, pulmonary infections do not occur in humans and thus represents an unnatural site of infection. Thus, disregarding the discrepancies arising due to the differences between mice and humans, findings gleaned from the model regarding aspects of pulmonary immunity may not be accurate when translated to the intestine. Testing vaccines in a more natural infection requires using higher non-human primates (NHPs), or more typically, human challenge studies, but these typically end up being early phase clinical trials (33), which are complex and labor-intensive to conduct. NHPs are the only known animal to develop shigellosis upon oral inoculation as adults. However, their use is highly restricted, which is in part due to their high cost. Human challenge studies are valuable, yet in these studies, it is not as easy as in animal models to rapidly test multiple approaches or acquire intestinal tissue. An attractive alternative to NHPs and human testing

could be the intra-rectal adult guinea pig model. While the model has been successfully used over the past several years for numerous studies on *Shigella* pathogenesis (34–36), in the original paper, the model was shown to be useful for studying the efficacy of intestinally delivered vaccines (37). The lack of its broader acceptance and use in *Shigella* vaccine research may be attributable to its reliance on a non-traditional animal species, which makes adoption of the model challenging. However, the intra-rectal adult guinea pig model may provide a much more relevant, translatable model than the mouse pulmonary model for studies on *Shigella* vaccine development.

Summary

Our work has resulted in several unexpected findings and advanced our knowledge of disease pathogenesis in multiple bacterial pathogens, yet much remains to be uncovered about these infectious agents and host responses to them. Future work will aim to deepen our understanding of the myriad ways in which pathogens cause disease in their hosts and how we can overcome these infections, through either innate approaches or novel therapeutics.

References

1. **Frost HR, Sanderson-Smith M, Walker M, Botteaux A, Smeesters PR**. 2018. Group A streptococcal M-like proteins: From pathogenesis to vaccine potential. *FEMS Microbiol Rev* **42**:193–204.

2. **Connolly JPR, Slater SL, O'Boyle N, Goldstone RJ, Crepin VF, Ruano-Gallego D, Herzyk P, Smith DGE, Douce GR, Frankel G, Roe AJ**. 2018. Host-associated niche metabolism controls enteric infection through fine-tuning the regulation of type 3 secretion. *Nat Commun* **9**:4187.

3. **Lannergård J, Frykberg L, Guss B**. 2003. CNE, a collagen-binding protein of Streptococcus equi. *FEMS Microbiol Lett* **222**:69–74.

4. **Steward KF, Robinson C, Maskell DJ, Nenci C, Waller AS**. 2017. Investigation of the Fim1 putative pilus locus of Streptococcus equi subspecies equi. *Microbiology* **163**:1217–1228.

5. **Flock M, Karlström A, Lannergård J, Guss B, Flock J-I**. 2006. Protective effect of vaccination with recombinant proteins from Streptococcus equi subspecies equi in a strangles model in the mouse. *Vaccine* **24**:4144–51.

6. **D'Gama JD, Ma Z, Zhang H, Liu X, Fan H, Morris ERA, Cohen ND, Cywes-Bentley C, Pier GB, Waldor MK**. 2019. A Conserved Streptococcal Virulence Regulator Controls the Expression of a Distinct Class of M-Like Proteins. *MBio* **10**:e02500-19.

7. **Hong-Jie F, Fu-yu T, Ying M, Cheng-ping L**. 2009. Virulence and antigenicity of the szp-gene deleted Streptococcus equi ssp. zooepidemicus mutant in mice. *Vaccine* **27**:56–61.

8. **Timoney JF, Artiushin SC, Boschwitz JS**. 1997. Comparison of the sequences and functions of Streptococcus equi M-like proteins SeM and SzPSe. *Infect Immun* **65**:3600–3605.

9. **Hondorp ER, McIver KS**. 2007. The Mga virulence regulon: Infection where the grass is greener. *Mol Microbiol* **66**:1056–1065.

10. **Gao X-Y, Zhi X-Y, Li H-W, Klenk H-P, Li W-J**. 2014. Comparative Genomics of the Bacterial Genus Streptococcus Illuminates Evolutionary Implications of Species Groups. *PLoS One* **9**:e101229.

11. **Waugh DJJ, Wilson C**. 2008. The interleukin-8 pathway in cancer. *Clin Cancer Res* **14**:6735–41.

12. **Reed WP, Albright EL**. 1974. Serum factors responsible for killing of Shigella. *Immunology* **26**:205–15.

13. **Weinrauch Y, Drujan D, Shapiro SD, Weiss J, Zychlinsky A**. 2002. Neutrophil elastase targets virulence factors of enterobacteria. *Nature* **417**:91–4.

14. **Combes S, Michelland RJ, Monteils V, Cauquil L, Soulié V, Tran NU, Gidenne T, Fortun-Lamothe L**. 2011. Postnatal development of the rabbit caecal microbiota

composition and activity. *FEMS Microbiol Ecol* **77**:680–9.

15. **Thompson CN, Duy PT, Baker S**. 2015. The rising dominance of Shigella sonnei: An intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis* **9**:1–13.

16. **Pilla G, McVicker G, Tang CM**. 2017. Genetic plasticity of the Shigella virulence plasmid is mediated by intra- and inter-molecular events between insertion sequences. *PLoS Genet* **13**:e1007014.

17. **Anderson MC, Vonaesch P, Saffarian A, Marteyn BS, Sansonetti PJ**. 2017. Shigella sonnei Encodes a Functional T6SS Used for Interbacterial Competition and Niche Occupancy. *Cell Host Microbe* **21**:769-776.e3.

18. **Ritchie JM, Thorpe CM, Rogers AB, Waldor MK**. 2003. Critical roles for stx2, eae, and tir in enterohemorrhagic Escherichia coli-induced diarrhea and intestinal inflammation in infant rabbits. *Infect Immun* **71**:7129–39.

19. **Nishibuchi M, Seidler RJ, Rollins DM, Joseph SW**. 1983. Vibrio factors cause rapid fluid accumulation in suckling mice. *Infect Immun* **40**:1083–91.

20. **Duchet-Suchaux M, Le Maitre C, Bertin A**. 1990. Differences in susceptibility of inbred and outbred infant mice to enterotoxigenic Escherichia coli of bovine, porcine and human origin. *J Med Microbiol* **31**:185–90.

21. **Porter CK, Thura N, Ranallo RT, Riddle MS**. 2013. The Shigella human challenge model. *Epidemiol Infect* **141**:223–232.

22. **Levine MM, Kotloff KL, Barry EM, Pasetti MF, Sztein MB**. 2007. Clinical trials of Shigella vaccines: Two steps forward and one step back on a long, hard road. *Nat Rev Microbiol* **5**:540–553.

23. **Li A, Kärnell A, Huan PT, Cam PD, Minh NB, Trâm LN, Quy NP, Trach DD, Karlsson K, Lindberg G**. 1993. Safety and immunogenicity of the live oral auxotrophic Shigella flexneri SFL124 in adult Vietnamese volunteers. *Vaccine* **11**:180–9.

24. **Lee HH, Ostrov N, Wong BG, Gold MA, Khalil AS, Church GM**. 2019. Functional genomics of the rapidly replicating bacterium Vibrio natriegens by CRISPRi. *Nat Microbiol* **4**:1105–1113.

25. **Wang T, Guan C, Guo J, Liu B, Wu Y, Xie Z, Zhang C, Xing X-H**. 2018. Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. *Nat Commun* **9**:2475.

26. **Sokol CL, Luster AD**. 2015. The Chemokine System in Innate Immunity. *Cold Spring Harb Perspect Biol* **7**:a016303.

27. **Anand BS, Malhotra V, Bhattacharya SK, Datta P, Datta D, Sen D, Bhattacharya MK, Mukherjee PP, Pal SC**. 1986. Rectal histology in acute bacillary dysentery. *Gastroenterology* **90**:654–60.

28. **Mathan MM, Mathan VI**. 1991. Morphology of rectal mucosa of patients with shigellosis. *Rev Infect Dis* **13 Suppl 4**:S314-8.

29. **Yum LK, Byndloss MX, Feldman SH, Agaisse H**. 2019. Critical role of bacterial dissemination in an infant rabbit model of bacillary dysentery. *Nat Commun* **10**:1–10.

30. **Christopher PR, David K V, John SM, Sankarapandian V**. 2010. Antibiotic therapy for Shigella dysentery. *Cochrane database Syst Rev* CD006784.

31. **Mallett CP, VanDeVerg L, Collins HH, Hale TL**. 1993. Evaluation of Shigella vaccine safety and efficacy in an intranasally challenged mouse model. *Vaccine* **11**:190–196.

32. **Van de Verg LL, Mallett CP, Collins HH, Larsen T, Hammack C, Hale TL**. 1995. Antibody and cytokine responses in a mouse pulmonary model of Shigella flexneri serotype 2a infection. *Infect Immun* **63**:1947–1954.

33. **Coster TS, Hoge CW, VanDeVerg LL, Hartman AB, Oaks E V., Venkatesan MM, Cohen D, Robin G, Fontaine-Thompson A, Sansonetti PJ, Hale TL**. 1999. Vaccination against shigellosis with attenuated Shigella flexneri 2a strain SC602. *Infect Immun* **67**:3437–43.

34. **Arena ET, Campbell-Valois FX, Tinevez JY, Nigro G, Sachse M, Moya-Nilges M, Nothelfer K, Marteyn B, Shorte SL, Sansonetti PJ**. 2015. Bioimage analysis of Shigella infection reveals targeting of colonic crypts. *Proc Natl Acad Sci U S A* **112**:E3282–E3290.

35. **Xu D, Liao C, Zhang B, Tolbert WD, He W, Dai Z, Zhang W, Yuan W, Pazgier M, Liu J, Yu J, Sansonetti PJ, Bevins CL, Shao Y, Lu W**. 2018. Human Enteric α-Defensin 5 Promotes Shigella Infection by Enhancing Bacterial Adhesion and Invasion. *Immunity* **48**:1233-1244.e6.

36. **Tinevez J-Y, Arena ET, Anderson M, Nigro G, Injarabian L, André A, Ferrari M, Campbell-Valois F-X, Devin A, Shorte SL, Sansonetti PJ, Marteyn BS**. 2019. Shigella-mediated oxygen depletion is essential for intestinal mucosa colonization. *Nat Microbiol* **4**:2001–2009.

37. **Shim D-H, Suzuki T, Chang S-Y, Park S-M, Sansonetti PJ, Sasakawa C, Kweon M-N**. 2007. New Animal Model of Shigellosis in the Guinea Pig: Its Usefulness for Protective Efficacy Studies. *J Immunol* **178**:2476–2482.