SSRB-DEPENDENT REGULATION DURING *SALMONELLA* PATHOGENESIS

SSRB-DEPENDENT REGULATION DURING *SALMONELLA* PATHOGENESIS

By

ANA MARIA TOMLJENOVIC-BERUBE, B.Sc.

A Thesis Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements for
the Degree

Doctor of Philosophy

McMaster University DOCTOR OF PHILOSOPHY (2012) Hamilton, Ontario
(Biochemistry and Biomedical Sciences)

TITLE: SsrB-Dependent Regulation During *Salmonella* Pathogenesis

AUTHOR: Ana Maria Tomljenovic-Berube, B.Sc. (McMaster University)

SUPERVISOR: Dr. Brian K. Coombes

NUMBER OF PAGES: xvi, 194

ABSTRACT

Bacteria demonstrate an extraordinary capacity to survive and adapt to changing environments. In part, this ability to adapt can be attributed to horizontal gene transfer, a phenomenon which introduces novel genetic information that can be appropriated for use in particular niches. Nowhere is this more relevant than in pathogenic bacteria, whose acquisition of virulence genes have provided an arsenal that permits them to thrive within their selected host. Regulatory evolution is necessary for timely regulation of these acquired virulence genes in the host environment. *Salmonella enterica* serovar Typhimurium is an intracellular pathogen which possesses numerous horizontally-acquired genomic islands encoding pathogenic determinants that facilitate its host lifestyle. One island, *Salmonella* Pathogenicity Island (SPI)-2, encodes a type-III secretion system (T3SS) which is regulated by the two-component regulatory system SsrA-SsrB. This system coordinates expression of the SPI-2 T3SS as well as an array of virulence effectors encoded in horizontally-acquired regions throughout the *Salmonella* genome. The studies presented here investigated the mechanisms in which the transcription factor SsrB functions to integrate virulence processes through regulatory adaptation. This work identified the regulatory logic controlling SsrB and defined the associated SsrB regulon. Furthermore, SsrB was found to induce a regulatory cascade responsible for the expression of bacteriophage genes encoded within SPI-12, an island that also contributes to *Salmonella* virulence. These findings demonstrate the important contribution of regulatory evolution in pathogen adaptation to the host, and show that

horizontally-acquired genes, once integrated into appropriate regulatory networks, can contribute to pathogen fitness in specific niche environments.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my supervisor, Dr. Brian Coombes, for taking me on as one of his first graduate students and for teaching me what true passion for research is all about. I just knew when we first met that to work with him was an opportunity I couldn't afford to pass up. We started off in a small corner of 4H13, just four students in two bays, and I am so proud and happy for all of the success we've had these last six years. Brian, thank you for inviting me to come work with you, for giving me the space, support, and resources to pursue this work, and for being a wonderful mentor. I am truly thankful for the experience. Thank you also to my committee members Dr. Justin Nodwell and Dr. Herb Schellhorn for challenging me to think harder, read more, and do better. Your guidance and help has been greatly appreciated.

To all of the fantastic Coombes Lab members, past and present: Thank you for all of the great memories! In spite of the ups and downs of science research, I am extremely fortunate that I had a wonderful support group in all of you. I am so glad I got to work with such a fantastic group of people who made the lab environment a fun, loving, and inspiring place to work. Regardless of my occasional grumbling, I loved being the "lab mom" and enjoyed seeing you develop and grow as researchers. I know you all have brilliant careers ahead of you! Suzanne, Christine, and Colin – you were with me from the ground-up and your help was immeasurable over the years. Our successes and frustrations were shared, making the journey that much easier. Thank you from the bottom of my heart.

Of all people, my greatest cheerleaders have been those closest to me, my family. To my parents: *Mama i Tata, obične riječi nemogu prenjeti koliko vaša ljubav i pomoć je meni značilo preko mog života. Vi ste moja snaga, moja inspiracija, is sav ovaj naporni rad je bio vami u čast. Volim vas jako puno. Hvala od dno srca!* My mom once said to me, "Nobody is smarter than you. You can do whatever you want to do, and be whatever you want to be. Be strong, Ana. You can do it." I have carried those words with me every day, and they light the way when life is at its darkest. Thank you for being the source of my strength and confidence. To Colleen and Richard: Thank you your support, for listening to my worries, for helping me stay grounded, and for sending me care packages when I needed them. I am truly blessed to have you in my life. To my siblings, Tomislav, Kerri, Jen, and Chad: Thank you for being a source of strength, for being interested in my work even when it seems incomprehensible, and for all the help outside of my science life. To my best friend Jenn: Thank you for always being a shoulder to cry on, a friend to cheer with, and a distraction from the chaos. I love you all so very much.

Finally, to my one and only, my husband Michael: You came into my life almost exactly to the day that I started on this scholastic journey at McMaster. Through thick and thin, at my highest and at my lowest, I have shared all of these experiences with you. Any successes I have had over these last eleven years are because of you. You have been my rock, my confidant, my voice of reason, my joy, my everything. I would *never* have made it this far without you. This body of work is as much yours as it is mine. Thank you for getting me here. You mean the world to me and I am blessed to have you in my life. I love you, Boo.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	adenine
AmpR	ampicillin-resistant
B1H	bacterial-one hybrid
BLAST	Basic Local Alignment Search Tool
bp	base pair
C	cytosine
CA	catalytic/ATPase domain
Ca^{2+}	calcium
cDNA	complementary DNA
CDS	coding DNA sequence
CFU	colony forming units
ChIP	chromatin immunoprecipitation
ChIP-on-chip	chromatin immunoprecipitation on microarray (chip)
CI	competitive index
Cm	chloramphenicol
COG	Clusters of Orthologous Groups
Cy3	cyanine-3 dye
Cy5	cyanine-5 dye
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DHp	dimerization/histidine phosphotransfer domain
DMEM	Dulbecco's Modified Eagle Medium
DNA	deoxyribonucleic acid
DTT	dithiothreitol
dTTP	deoxythymidine triphosphate
E	eluted fraction
E. coli	*Escherichia coli*
ECL	enhanced chemiluminescence
EHEC	enterohemorrhagic *Escherichia coli*
EPEC	enteropathogenic *Escherichia coli*
FBS	fetal bovine serum
G	guanine
GSP	gene-specific primer
H2H	head-to-head
HA	hemagglutinin

HCl	hydrochloric acid
HGT	horizontal gene transfer
his	histidine
HK	histidine protein kinase
HM	host membrane
HRP	horseradish peroxidase
HTH	helix-turn-helix
IGR	intergenic region
IL	interleukin
IM	inner membrane
IPTG	isopropyl-β-D-1-thiogalactopyranoside
KanR	kanamycin-resistant
LB	Luria-Bertani medium
LEE	Locus of Enterocyte Effacement
LPM	low phosphate, low magnesium minimal medium
M	microfold cells
Mg^{2+}	magnesium
MTOC	microtubule organizing centre
NaOH	sodium hydroxide
NM	minimum glucose medium lacking histidine
NO	nitric oxide
nt	nucleotide
NTD	amino (N)-terminal domain
NTS	non-typhoidal *Salmonella*
NUT	N utilization site
OD	optical density
OM	outer membrane
P	Promoter
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PG	peptidoglycan
phage	bacteriophage
P_L	phage lambda left promoter
PMN	polymorphonuclear lymphocytes
PO_4^{2-}	phosphate
P_R	phage lambda right promoter
P_R'	phage lambda right prime promoter
PVDF	polyvinylidene fluoride

PWM	position weight matrix
qRT-PCR	quantitative real-time polymerase chain reaction
qut	Q utilization site
RACE	random amplification of cDNA ends
RES	reticuloendothelial system
RLU	relative light units
RNAP	RNA polymerase
RNS	reactive nitrogen species
ROD	region of difference
RR	response regulator
RT-PCR	reverse-transcription polymerase chain reaction
S. bongori	*Salmonella bongori*
S. enterica	*Salmonella enterica*
S. Typhimurium	*Salmonella enterica* serovar Typhimurium
S. mitis	*Streptococcus mitis*
SCV	*Salmonella*-containing vacuole
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SIF	*Salmonella*-induced filament
SmR	streptomycin-resistant
SPI	*Salmonella* Pathogenicity Island
SPI-2	*Salmonella* Pathogenicity Island-2
SPI-12	*Salmonella* Pathogenicity Island-12
ssa	secretion system apparatus
ssc	secretion system chaperone
sse	secretion system effector
ssr	secretion system regulator
S. glossinidius	*Sodalis glossinidius*
SSR-3	*Sodalis* Symbiosis Region-3
STEC	Shiga-toxin producing *E. coli*
subsp	subspecies
sv	serovar
T	thymine
T1SS	type-I secretion system
T3SS	type-III secretion system
T6SS	type-VI secretion system
TCS	two-component system
TF	transcription factor

t_L	left terminator
TLR	Toll-like receptor
Tn	transposon
t_R	right terminator
tRNA	transfer ribonucleic acid
U	unbound fraction
v/v	ratio of volume/volume
W	whole-cell lysate
w/v	ratio of weight/volume
λ	phage lambda
Δ	delta, deletion
σ	sigma factor
+	positive control

CHAPTER ONE

INTRODUCTION

INTRODUCTION

Integrated Virulence, an Introduction

The traditional perspective on evolution dictates that changes in the genetic makeup of a particular organism are acquired over long periods of time through random, isolated genetic changes that are inherited from parent to offspring. These changes may ultimately be retained or lost, depending on the fitness impact to the organism in a particular environment, and can be readily observed in the polymorphisms of various species today. In addition to this vertical, hereditary form of evolution, bacteria have adapted another mechanism to fast-forward evolution by way of lateral or horizontal gene transfer (HGT), in which genetic information is exchanged across bacterial species in modules that can ultimately provide benefits to the new carrier. This is particularly true amongst pathogens, whose acquisition of horizontally-acquired DNA has greatly impacted their success in establishing new niches within host organisms [1]. Researchers have come to understand the mechanisms used by bacteria to facilitate this process and generally appreciate the impact that HGT has made on bacterial evolution, however many questions remain. How do bacteria take this novel genetic information and use it to their advantage? More specifically, how do they regulate newly acquired DNA to ensure appropriate timing and usage of newly acquired genes? We know that nucleoid-associated proteins such as H-NS and Sfh function to silence horizontally-acquired DNA to prevent aberrant expression of potentially lethal genes in a new genome environment [2-4], but how do these genes ultimately get incorporated into existing regulatory gene networks in order to make use of their potential functions?

Recent work has revealed that regulatory evolution likely plays a role. Regulatory evolution is the process by which changes in non-coding *cis*-regulatory regions of genes results in an altered phenotype in the affected organism due to changes in downstream gene expression [5, 6]. Examples of this phenomenon have been identified in *Salmonella*, where genes shared between non-pathogenic and pathogenic species show evidence of *cis*-regulatory changes that introduce the use of a new transcriptional regulator to influence the expression of these factors specifically in the pathogen during infection [7, 8]. Ultimately, this phenomenon has integrated these genes into the intracellular virulence program in *Salmonella enterica*. Thus, *integrated virulence* is a process in which multiple virulence factors or genes are incorporated into one or more regulatory network(s) to facilitate survival in the host environmental niche. This integration is mediated by changes at the DNA sequence level to ensure recognition by transcription factors involved in virulence processes. This is particularly relevant for horizontal DNA acquisitions, whose regulatory sequences must be reprogrammed in order to be exploited by the carrier when conquering new environments. Therefore, in bacterial pathogens, integrated virulence occurs at the intersection between regulatory evolution and horizontal gene transfer, ultimately contributing to bacterial evolution overall.

Horizontal Gene Transfer and Pathogen Evolution

The lateral exchange of genetic information can proceed through several mechanisms: transformation, conjugation, and transduction. In general, HGT by way of

transformation involves the uptake of extracellular DNA. This takes place in certain bacterial species that encode and express competence factors, which are proteins that mediate DNA processing and acquisition [9]. Natural transformation thus occurs as competent bacteria encounter DNA in their environment. Conjugation is the process in which genetic information is transferred directly between a donor to a recipient bacterium through a pilus appendage, which acts as a conduit for DNA to pass through [10]. In this case, the donor must carry the genes necessary for pilus synthesis. Upon successful transfer, the recipient, called the transconjugant, integrates the newly acquired genetic content into its own genome or carries it as a plasmid. Transfer of DNA is mediated either by a plasmid or a chromosomally integrated conjugative element, such as a conjugative transposon [11]. Finally, transduction is the process in which bacterial viruses, known as bacteriophage or simply phage, mediate HGT. This generally occurs due to packaging of bacterial DNA by phage during virus synthesis in an infected cell. DNA transfer is mediated when newly synthesized phage particles carrying bacterial DNA from one host move on to infect another bacterium, injecting DNA into the new target which can then be integrated within the genome of this new host. All three of these processes ultimately lead to the acquisition of novel genetic information that can be integrated into the bacterial chromosome and subsequently passed on vertically from generation to generation [12].

Evidence of HGT is most readily observed by comparisons of genome content between related organisms and through evaluation of the average G+C content of DNA across the genome. Foreign DNA that was recently acquired exhibits deviations from the

average genome G+C content, typically possessing a higher A+T skew and can demonstrate codon bias [13, 14]. Additionally, DNA which has been horizontally-acquired is frequently associated with common sites of integration, such as tRNA loci, and/or encodes sequences that imply mobility, such as dinucleotide bias, transfer origins of plasmids, remnants of translocatable DNA elements, redundant sequences, and genes encoding integrases, amongst others [12, 15, 16]. All together, regions of DNA that appear to be different from surrounding chromosomal regions are typically referred to as "genomic islands," while regions carrying genetic content that specifically confer virulence attributes are labelled as "pathogenicity islands."

The fitness contributions that HGT has made to pathogens within the context of the host environment have been well-documented. Spread of antibiotic resistance by way of plasmid-mediated HGT has occurred since the first use of antibiotics in the 1940's, and is the most common mechanism of resistance gene exchange [17, 18]. They have had a prominent impact on the treatment of bacterial infections, particularly in hospital environments, where multi-drug resistant strains of pathogens such as methicillin-resistant *Staphylococcus aureus, Salmonella enterica* strain DT104, and *Acinetobacter baumanii* manage to subvert most antimicrobial therapies, making treatment increasingly difficult, particularly for immuno-compromised individuals [19-23].

In addition to plasmid-encoded virulence attributes, transposon- and phage-integrated DNA elements are ubiquitous among pathogenic bacteria and confer many advantages to their carriers. Pathogenicity islands have made a significant impact on pathogen evolution, particularly in enteric pathogens. For example, the Locus for

Enterocyte Effacement (LEE) in enteropathogenic *Escherichia coli* (EPEC) as well as *Salmonella* Pathogenicity Island-1 and -2 in *Salmonella enterica* encode a specialized secretion system that facilitates interaction with and manipulation of host cells; the LEE mediates attachment and effacement of gut epithelial cells, while the systems in *Salmonella* cause invasion, and mediate survival of these pathogens in host cells. Many pathogenicity islands also encode genes for virulence effector proteins, fimbriae, toxins, adhesins, as well as other secretion systems that facilitate host manipulation [15]. Prophages in particular have been found to encode toxins which ultimately serve as virulence determinants in their host bacterium. Prominent examples include botulinum toxin in *Clostridium botulinum*, cholera toxin in *Vibrio cholerae* and Shiga toxin in various pathogenic *Escherichia coli* (*E. coli*) isolates. These genes, called lysogenic conversion genes, result in the automatic conversion of non-pathogenic bacterial strains into a pathogenic variant simply through latent infection by a toxin-encoding phage [24]. Prophages can carry additional genes that frequently encode virulence factors which contribute to fitness in the host environment [25]. These genes often function as translocated effectors for secretion systems, as best exemplified in the enteric pathogen *Salmonella enterica*, which employs a number of phage-encoded effectors, including SopE, GogB, and SseI [26-29].

Upon incorporation of new DNA into the genome, regulation of horizontally-acquired DNA is believed to be facilitated by nucleoid-associated proteins, such as H-NS, YdgT, and Hha amongst others [2, 3, 30, 31]. These proteins, exemplified by H-NS, have been found to preferentially bind to A-T rich DNA sequences and suppress transcriptional

activity from these loci, which often include horizontally-acquired DNA [2, 3]. Sfh, a plasmid-encoded nucleoid-associated protein was even found to specifically function in a stealth capacity to mask the introduction of the virulence plasmid into new *Salmonella* strains, thereby ensuring its preservation in the new carrier [4]. These proteins thus play an important role in suppressing expression of newly acquired gene content and ensure that aberrant expression does not inadvertently decrease the fitness of the new carrier.

Eventual integration of horizontally-acquired DNA into existing regulatory networks has been observed in several organisms, particularly with respect to virulence attributes which require careful contextual regulation in specific environments. For instance, expression of the horizontally-acquired virulence regulator ToxT has been incorporated into the ancestral TcpP and ToxR regulatory networks in *Vibrio cholerae* [32, 33]. Similarly in *Yersinia* species, several horizontally-acquired genes have come under the control of the RovA ancestral regulator, while the master virulence regulator PhoP in *Salmonella enterica* has also demonstrated the same capacity for integrated virulence, recognizing alternative *cis*-regulatory sequences for ancestral and horizontally-acquired genes [34-36]. Some of these regulatory mechanisms involve antagonism of H-NS, which relieves repression and recruits RNA polymerase to these loci to elicit transcriptional activation [37].

Salmonella enterica **Species**

Salmonellae are Gram-negative enterobacteria that infect humans and animals and are a common source of foodborne diseases. These facultative anaerobic bacteria are

generally transmitted as a zoonotic infection through the fecal-oral route and/or direct contact with animal reservoirs. Symptoms in infected organisms range from self-limiting gastroenteritis to severe systemic illness associated with enteric fever [38]. The *Salmonella* genus is believed to have diverged from *E. coli* approximately 120 million years ago [39]. There are two recognized members of the *Salmonella* species: *Salmonella bongori* and *Salmonella enterica*. Of these two species, *Salmonella bongori* (*S. bongori*) is largely associated with reptiles and rarely causes infection in humans [40]. Alternatively, *Salmonella enterica* (*S. enterica*) predominates as an enteropathogen and shows varied host specificity amongst warm-blooded hosts.

The *S. enterica* species is further divided into six subspecies (subsp): *S. enterica* subsp. *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI). These can be further subdivided into serological variants (serovar or sv.) according to their lipopolysaccharide (O-), flagellar (H-), and capsular (Vi-) antigen presentation [39]. Over 2,500 different serovars of *S. enterica* have been identified to date, and nearly all serovars that cause disease in humans and food-producing animals belong to *S. enterica* subspecies I. By convention, *S. enterica* subsp. *enterica* serovars are labelled with capitalized non-italicized text [41]. *S. enterica* serovars have a wide range of host specificity, where some infect a broad range of hosts, such as serovars Typhimurium and Enteriditis, while other serovars have only one primary host target, such as sv. Typhi and sv. Paratyphi in humans, sv. Choleraesuis in pigs, and sv. Pullorum in poultry [42]. *Salmonella enterica* sv. Typhimurium (*S.* Typhimurium) is one of the most common clinical serovars and causes localized gastroenteritis in humans and a

systemic illness in mice, making it a tractable model for enteric infectious disease research [43, 44].

Salmonellosis: Disease and Impact

Salmonella infections have a prominent global impact, with an annual incidence of ~1.3 billion cases world-wide [45]. The majority of these infections are associated with non-typhoidal *Salmonella* (NTS), however typhoidal *Salmonella* infections caused by *S.* Typhi and *S.* Paratyphi are still prevalent in developing countries, with an estimated 21 million cases and up to 600,000 deaths annually [46]. Invasive NTS, a previously underappreciated pathogen is emerging as a clinically important disease in sub-Saharan Africa, associated with a case fatality of 20-25% in infected individuals [47, 48]. Elsewhere, typhoid fever is an important cause of morbidity and mortality in children and adolescents in south-central and south-east Asia, where contaminated water and food sources act as primary vehicles for transmission. Estimates of the costs associated with salmonellosis are high and range from $2.65 billion to $11.39 billion annually in the United States alone [49]. In Canada, *Salmonella* infection incidence is estimated to be greater than 200,000 annually, with an associated cost of $2.24 million [50]. Thus, *Salmonella* infections have a huge global burden, not only due to disease prevalence, but also due to the socioeconomic burden on healthcare, detection, and prevention costs.

In animal models of persistent *S. enterica*-mediated enterocolitis, bacteria survive passage through the stomach and traverse the intestinal mucous layer to localize to the apical surface of the intestinal epithelium. *Salmonella* can specifically target a

specialized epithelial cell type called microfold (M) cells which function to sample intestinal antigens from the gut lumen within the Peyer's patches [51, 52]. At the epithelium, *Salmonella* invade non-phagocytic enterocytes by bacterial-mediated endocytosis using a secretion complex called the type-III secretion system (T3SS). This system translocates bacterial proteins termed effectors into the host cytoplasm to induce actin cytoskeleton rearrangements that cause membrane ruffling, leading to eventual engulfment of the bacteria into the host cell [53, 54]. Additionally, salmonellae disrupt tight junctions to cross the epithelial barrier and induce a pronounced inflammatory response which contributes to diarrheal symptoms of the infection [55]. This inflammatory response involves induction of interleukin (IL)-8 secretion and recruitment of polymorphonuclear lymphocytes (PMNs) which further propagates inflammation [56, 57].

Upon successful invasion of the epithelial barrier, *Salmonella* serotypes that cause systemic illness, such as *S.* Typhi and *S.* Paratyphi can promote their dissemination through the reticuloendothelial system (RES) by surviving within intestinal macrophages. These serovars typically induce a weaker inflammatory response to allow it to discretely colonize systemic organs, such as the mesenteric lymph nodes, spleen, and gallbladder [58, 59]. Salmonellae encounter resident intestinal macrophages once they have crossed the initial epithelial barrier of the gut. Exposure of *Salmonella* to macrophages results in caspase-1 dependent host cell death, however this interaction with the host innate immune system is ultimately beneficial, as uptake by resident phagocytes promotes *Salmonella* survival [60, 61]. Engulfment of salmonellae by macrophages as part of the innate

immune response provides a replicative niche for the bacterium and protects the pathogen from the adaptive immune response. It utilizes a second T3SS to secrete over 20 effectors that manipulate the macrophage intracellular environment. This process transforms the phagosome into a vacuole known as the *Salmonella*-containing vacuole (SCV). T3SS and consequently the development and maintenance of the SCV are vitally important to intramacrophage survival and long-term persistence in mice [62-66]. Overall, *Salmonella* preferentially reside within macrophages for the majority of time during systemic infection [65, 67, 68]. Infected macrophages mobilize from the lamina propria to mesenteric lymph nodes and eventually out to the general circulation, where they encounter organs connected to the RES such as the liver, spleen, and bone marrow [59, 69]. Dissemination to peripheral organs occurs within 24 hours of ingestion following an incubation period of 8-14 days. Shedding of a low level of bacteria back into the bloodstream marks the onset of clinical illness.

Research into how this diversified pathogen can target a multitude of hosts, promote survival in numerous environments, and facilitate consistent self-propagation continues to reveal the nuanced complexity of *Salmonella enterica* infections. Ultimately, many answers reside within the depths of its genome, where through HGT it has collected an assortment of tools encoded within genomic islands that permit its complex lifestyle. Investigation of the function and regulation of the genes within these islands has revealed important features within this broad molecular toolkit.

Salmonella Pathogenicity Islands

Salmonella enterica species possesses a large number of horizontally-acquired genomic islands which encode genes that confer virulence attributes. These islands, named *Salmonella* Pathogenicity Islands (SPI), vary in size and function and are categorized by a numeric designation. Twenty-two SPI have been identified among all *Salmonella* to date, though not all serovars encode all SPI [70, 71]. The two most important, SPI-1 and SPI-2, each encode a type-III secretion system (T3SS) which is a needle-like apparatus that extends through the inner and outer membranes of the bacterium and makes contact with the host membrane to effectively "inject" or translocate bacterial virulence proteins, termed effectors, into the host cell [72]. All *Salmonella* species carry SPI-1, however only *S. enterica* species carry SPI-2, the main genetic feature which distinguishes *S. bongori* and *S. enterica* species. It is the acquisition of SPI-2 that is believed to have allowed *S. enterica* to expand its host range into mammals [40, 73].

SPI-1 is a 40 kb genomic island that encodes the T3SS apparatus, effector genes, and regulators necessary for invasion of non-phagocytic cells [53, 74]. Additional effectors used by this system are encoded outside of SPI-1 in other horizontally-acquired loci. Examples include SopB in SPI-5 and SopE in the SopEΦ bacteriophage [26, 75, 76]. The expression of SPI-1 is induced in the host intestinal environment by activation in high-osmolarity and low-oxygen conditions [77]. Once the needle apparatus is assembled and upon epithelial cell contact, translocated effectors of the SPI-1 T3SS stimulate membrane ruffling of the apical surface of the gut epithelium, ultimately resulting in

engulfment and uptake of *Salmonella* into host cells through bacterial-mediated endocytosis [78]. Expression of this system also results in stimulation of the intestinal inflammatory response and facilitates colonization of the intestine, ultimately producing the disease symptoms associated with *Salmonella*-induced enteritis [79].

In contrast to SPI-1, SPI-2 is a 40 kb genomic island that is essential for intracellular survival in macrophages and the manifestation of systemic disease [62, 63, 80-82]. It is expressed within the phagosome once engulfed by the host cell and mediates intracellular survival by translocating effector proteins into the host cytosol which inhibit phagolysosome maturation [83, 84]. The SCV becomes a replicative niche, circumventing typical pathogen defence mechanisms such as reactive oxygen and nitrogen species [84-86]. Intracellular survival within macrophages ultimately allows for *Salmonella* systemic spread, as activated macrophages travel to peripheral lymph nodes and systemic organs [87].

The twenty remaining SPI contribute in a variety of ways to *Salmonella* virulence and were recently summarized in a review by Sabbagh *et al* [71]. They vary in composition and distribution amongst all *Salmonella*. In brief, SPI-3 to SPI-6 are present in all *S. enterica* subspecies and encode the MgtCB Mg^{2+} uptake system, the large non-fimbrial adhesin SiiE with its cognate type-I secretion system (T1SS), SPI-1 and SPI-2 T3SS effector proteins SopB and PipB, as well as the *saf* fimbrial operon, the adhesin/invasin PagN and a type-VI secretion system (T6SS) [88-93]. SPI-7 is a 133 kb island that is specific for *S.* Typhi, *S*. Dublin and *S*. Paratyphi C. It encodes the *viaB* locus which produces Vi antigen, a capsular exopolysaccharide which has been found to

subvert innate immune responses by the host in preventing recognition by Toll-like receptor-4 (TLR-4) and inhibiting complement-mediated clearance of *S.* Typhi [94-96]. This locus also encodes the SPI-1 associated effector SopE in strains not carrying the SopEΦ bacteriophage [26, 76, 97]. The remaining SPI loci have variable distribution amongst individual *Salmonella* serovars. They encode a variety of virulence associated genes, including additional fimbrial operons, secretion systems, T3SS-associated effectors, and serotype conversion genes amongst others [71]. Overall, the SPIs have far-reaching functions that contribute to the genetic diversity and host specificity of *Salmonella* serovars.

Salmonella Pathogenicity Island-2

SPI-2 is a 40 kb genomic island located between *ydhE* and *pykF* adjacent to the *valV* tRNA locus in all *S enterica* species. It has a G+C content of 44.7%, which is lower than the genome average of 52%, supporting its hypothesized acquisition by HGT [70, 81]. It was first identified in 1996 by the laboratory of David Holden, who used signature-tagged mutagenesis to search for virulence determinants in *Salmonella*. They found that transposon mutants in this locus decreased *S.* Typhimurium virulence in mice by five orders of magnitude compared to wild type strains by either oral or intraperitoneal infection route [80]. At the same time, work by Ochman et al. confirmed the role of this locus in *Salmonella* virulence, showing that SPI-2 mutants were defective for intramacrophage survival [82]. The essential role of SPI-2 during intramacrophage survival was cemented further by studies which showed preferential expression of SPI-2

genes within macrophages as well as demonstrating their importance for maintaining intracellular survival [62, 63]. In addition to many cellular changes evoked by its translocated effectors, it has also been shown to contribute to protection against reactive oxygen and nitrogen species in the phagosomal environment [85, 86, 98].

SPI-2 encodes two distinct regions: a tetrathionate reductase system, and a T3SS. The tetrathionate reductase complex encodes the *ttr* gene cluster which is composed of genes *ttrBCA* that make up the complex, and a two-component regulatory system *ttrRS* that encodes proteins responsible for regulating the *ttr* genes. This gene cluster was acquired in a separate horizontal gene transfer event and is not directly involved in *Salmonella* systemic virulence [99]. The second, larger component of SPI-2 encodes 32 genes that comprise the SPI-2 T3SS and associated factors. Genes within this locus were designated "secretion system" genes, where *ssa* identifies apparatus genes, *ssc* identifies chaperone genes, *sse* identifies secreted effector genes, and *ssr* identifies regulatory genes (Figure 1.1) [100]. SPI-2 encodes a two-component regulatory system, SsrA-SsrB, where the transcriptional regulator SsrB functions to activate the transcription of the remaining genes in SPI-2 and is required for the expression of the T3SS [62, 63, 82].

Overall, there are four predominant operons in SPI-2: a regulatory operon regulated by the *ssrA* promoter, a chaperone/effector operon starting at the *sseA* promoter, and two operons encoding the structural components of the apparatus with promoters upstream of *ssaB* and *ssaG*. Each of these promoters is bound by SsrB *in vitro* and *in vivo* as shown by footprinting and chromatin immunoprecipitation (ChIP) studies. Additional promoter sites for *ssrB, ssaM* and *ssaR* have been reported, though it is

believed that these loci are secondary activation sites to the upstream promoter (Figure 1.1) [101, 102].

The regulatory landscape in place at SPI-2 is quite complex. Along with SsrB, there are several additional transcriptional regulators that positively influence SPI-2 gene expression. These include the OmpR-EnvZ and PhoP-PhoQ two-component systems, the transcription factor SlyA, and the nucleoid-associated factor Fis [103-108]. With the exception of Fis, these regulators act through the *ssrA* promoter and function to activate expression of *ssrA* and *ssrB*, as evidenced by the ability of SsrB to complement deletion mutants of these regulators [105, 109]. Fis was found to also positively regulate the *ssaG* promoter [106, 109]. Additionally, there are several negative regulators of SPI-2 gene expression. These regulators, H-NS, YdgT, and Hha, all function as nucleoid-associated proteins which, in addition to having roles in chromosomal structure, can function to suppress gene expression in loci associated with highly AT-rich DNA [2, 3, 30, 31, 110]. Negative regulation by these factors can be relieved in a number of ways, including SsrB-mediated antagonism of H-NS [102, 111].

The SPI-2 T3SS has been avidly studied since its discovery. While structural characterization of the apparatus has mostly been conducted *in silico* through comparative studies with conserved components in orthologous systems, its translocon and effector proteins have been extensively investigated both *in vitro* and *in vivo*. The structural composition of the secretion apparatus is largely conserved with those encoded by the LEE in EPEC as well as the *ysc* proteins in *Yersinia* [72, 100]. A schematic representation of the predicted SPI-2 T3SS can be found in Figure 1.2. T3SS span the

inner and outer bacterial membranes and extend a hollow needle from the bacterial cell surface in order to make contact with the host membrane. Translocation of effector proteins is mediated through a translocon which is believed to insert into the host membrane [112].

The translocon proteins SseB, SseC and SseD were found to be necessary for translocation of effector proteins into the host cell, and strains carrying deletions in these genes exhibited attenuated virulence [63, 112, 113]. Additionally, SseA was proposed to function as the chaperone to SseB and SseD, though SsaE has also been proposed to function as an SseB chaperone [114-118]. SsaB (also known as SpiC) has been extensively studied although its cellular localization and function remains controversial. In one study it was found to interact with SsaM and SsaL in the bacterial cytosol as a pH-dependent "gate-keeper" believed to be required for sensing the acidified vacuolar environment and for appropriately timing the secretion of the translocon and effector proteins [119, 120]. This finding contrasts with other reports that found that SsaB interacts with host target proteins TassC and Hook3 [121, 122].

As for proposed functions of the remaining apparatus proteins, SsaQ, the protein believed to form the cytoplasmic "C-ring" structure at the base of the secretion apparatus was characterized and found to have two isoforms where the short isoform SsaQ(S) acts as a chaperone to stabilize oligomerization of the long isoform SsaQ(L) [123]. SsaV and SsaRSTU are believed to form part of the inner membrane structure complex, as deletion of these genes results in loss of secretion and virulence attenuation. In fact, *aroC ssaV* mutants have been used as avirulent strains for vaccine development [100, 124-126].

SsaG, SsaH and SsaI are involved in the production of surface structures in *S.* Typhimurium and are necessary for intracellular survival and effector translocation. They have been proposed as putative components of the needle complex, though this has not been experimentally verified [127]. SsaC (also known as SpiA) has been proposed to function in formation of the outer membrane secretin that acts as the base on which the needle filament assembles [128]. A mutant deficient in SsaK was defective in secretion and was avirulent, though its exact function in the secretion apparatus is not known [125, 126, 129]. Finally, SsaN has been identified as the putative ATPase that associates with the secretion system and is hypothesized to energize secretion through ATP hydolysis [130]. SsaN was found to associate with a newly-identified chaperone SrcA in delivery of SPI-2 secreted effectors to the secretion apparatus [131].

Apart from apparatus proteins, SPI-2 encodes several additional chaperones and two secreted effectors, SseF and SseG. SscB was found to function as the chaperone for the effector SseF, and data from our laboratory showed that SscA is the chaperone for SseC [132, 133]. SseF and SseG have been implicated in contributing to reorganization of the microtubule network through microtubule bundling. They localize to *Salmonella*-induced filaments (SIFs) and have been shown to play a role in SIF and SCV formation. SseG was found to influence SCV motility by restricting vacuole speed proximal to Golgi bodies and the microtubule organizing centre (MTOC) [134, 135]. They were also shown to be important for maintaining juxtanuclear positioning of the SCV [136].

In addition to SseF and SseG, the SPI-2 T3SS translocates over 20 additional effectors that are not encoded within SPI-2. These effector genes are distributed

throughout the *Salmonella* genome and their expression is coordinated by the transcription factor SsrB. Together, these effectors mediate a wide variety of activities within the host cell during intracellular growth that help to maintain survival within the SCV. SsrB plays a central and critical role in the expression of these effectors, thereby greatly contributing to intracellular survival and *Salmonella* virulence.

The SsrA-SsrB Two-Component System

The SsrA-SsrB two-component system (TCS) plays an integral role in regulating intracellular survival of *S. enterica*. TCS are signal transduction systems that bacteria use to sense and respond to environmental changes. They mediate rapid adjustments in cellular functions in order to accommodate to their altered surroundings through a phosphotransfer mechanism. In Gram-negative bacteria, TCS are composed of a histidine protein kinase (HK) which resides in the inner membrane of Gram-negative bacteria to sense external stimuli, and a response regulator (RR) which elicits the response to changes relayed by the HK. In many cases the RR functions as a transcription factor (TF) and thus changes to the cell are mediated by altered gene expression [137]. Upon discovery of SPI-2, SsrA (then named SpiR), was identified as a putative sensor kinase, whilst SsrB was found to resemble RRs that function as TFs [80, 82]. Deletion or mutation of either gene prevents expression of the SPI-2 locus and renders *S.* Typhimurium incapable of intracellular replication [62, 63, 138].

Functional characterization of SsrA has been difficult due to its nature as a membrane-bound sensor kinase. Studies relating to its role in SPI-2 regulation have

focused on what signal(s) it recognizes in the vacuolar environment. Its exact stimulatory signal remains elusive, though *in vitro* activation can be achieved by growth in acidic minimal medium at pH ~5.0, containing low amounts of phosphate (PO_4^{2-}) and magnesium (Mg^{2+}) [138, 139]. These stimuli remain controversial, as the role of each component, i.e. low pH, divalent cations such as Mg^{2+} and Ca^{2+}, PO_4^{2-} starvation, as well as low osmolarity has been contentiously debated [138, 140-142]. Temperature has also been shown to influence SPI-2 gene expression, where expression is increased at host body temperature [143].

According to sequence analysis, SsrA is a member of the hybrid HK class of sensor kinases [82]. Generally, HKs possess multiple domains: an input/sensor domain that is located in the periplasmic space, two trans-membrane regions, a dimerization/histidine phosphotransferase (DHp) domain, and a catalytic/ATPase (CA) domain which binds ATP and mediates the autophosphorylation of a conserved histidine residue in the DHp domain. The DHp domain also mediates dimerization [144]. Hybrid HK also have additional domains that mediate an internal phosphorelay system. These typically include a receiver domain which receives the phosphate from DHp on a conserved aspartic acid residue, and an additional histidine phosphotransfer domain that may or may not be attached to the sensor kinase [144, 145]. Most cytoplasmic signalling domains are well conserved amongst sensor kinases. The input domains are thus the primary site of variation, as they have each evolved to recognize specific signals from the periplasmic space [146, 147]. Some HKs have also been found to function as phosphatases that de-phosphorylate the RR in order to inactivate signal transduction,

though SsrA has not been evaluated for this function [148]. In fact, no domains have been investigated thus far in SsrA, and its periplasmic signal remains unknown.

SsrB belongs to the CheY superfamily of RRs, and can be further categorized into the NarL/FixJ subfamily of transcriptional regulators. These proteins are generally composed of two domains connected by a flexible linker: an N-terminal receiver domain and a C-terminal DNA-binding domain [101, 149, 150]. The N-terminal receiver domain contains a conserved aspartic acid residue, Asp56, which is the site of phosphorylation from SsrA. Site-directed mutagenesis of this residue to alanine (D56A) rendered SsrB inactive for transcriptional activation [101]. In some transcription factors such as NarL, the N-terminal domain prevents DNA binding in the unphosphorylated state by physically blocking the DNA-binding domain. Upon phosphorylation, the protein undergoes a conformational change that frees the C-terminus for interaction with DNA [150, 151]. Truncation of SsrB into only its C-terminal DNA binding domain including linker results in a constitutively active form, termed SsrBc, that does not require phosphorylation for mediating transcriptional activation, indicating that the N-terminus likely functions as a regulatory domain [101].

The C-terminal domain of SsrB has been extensively analyzed by site-directed mutagenesis for identification of important functional residues. An NMR structure for SsrBc also provided further insights into how this domain functions to interact with DNA. SsrBc folds into a four-helix bundle, where helices 2 and 3 (7 and 8 in the full-length protein) form a helix-turn-helix (HTH) structure that is common for DNA-binding proteins. Helix 4 (H9 in the full-length protein) was deemed to function as part of a

dimerization domain as mutagenesis of residues conserved between SsrBc and other transcription factors in the same family rendered the protein both unable to activate transcription and bind target sequence *in vitro* [152]. Additionally, a cysteine found in the dimerization helix, Cys203, was found to be a target for nitric oxide (NO), a reactive nitrogen species (RNS) that *Salmonella* encounters in the gut. In functioning as a thiol-based redox sensor, this residue may explain why RNS negatively affects SPI-2 transcription, as exposure to RNS producing agents prevents SsrB from binding DNA *in vitro* [98].

There is at least one additional mechanism by which SsrB function is modulated post-translationally. In addition to S-nitrosylation of Cys203, SsrB was found to be negatively regulated by interaction with $EIIA^{Ntr}$, a component of the nitrogen-metabolic phosphotransferase system [153]. This protein was found to directly interact with SsrB to regulate its role as SPI-2 activator, as it physically prevents SsrB from binding DNA. This mechanism of inactivation is believed to serve as a way to prevent over-expression of SPI-2 genes, as strains deficient in the gene encoding $EIIA^{Ntr}$ had an attenuated phenotype despite increased SPI-2 gene expression [154]. Thus in addition to positive and negative transcriptional regulation of the SPI-2 locus and in turn SsrB, post-translational regulation adds a further level of control to the complex regulatory landscape at SPI-2.

The SsrB Regulon

As a TF, the primary role of SsrB is to regulate gene expression within SPI-2. This includes a positive feedback mechanism where it regulates itself through *ssrA* activation [101]. Additionally, SsrB has been found to regulate a significant number of genes throughout the *S. enterica* genome that play a role in intracellular survival within the host. These include effector proteins and chaperones that work with the SPI-2 T3SS to create a replicative niche within the SCV. SsrB has also been implicated in regulatory evolution, as two ancestral genes, *srfN* and *dalS,* have undergone promoter sequence changes to adapt to SsrB-dependent regulation in the pathogen *S. enterica* as part of divergence from *S. bongori* [7, 8].

Footprinting experiments showed that SsrB physically binds to ten promoters: *ssrA, ssrB, ssaB, sseA, ssaG, ssaM, sseI/srfH, srfN, sifA,* and *sifB*, however there are many more genes that are co-regulated with SPI-2 and are directly dependent on SsrB for transcriptional activation [7, 101, 102, 152, 155]. Overall, there have been 28 SPI-2 T3SS translocated effectors identified, as reviewed recently by Figueira and Holden [156]. Of these, at least 17 are directly regulated by SsrB [155]. Along with effectors, SsrB was also found to regulate a novel virulence chaperone SrcA, which is responsible for aiding secretion of PipB2 and SseL [131]. The complexity and broad range of influence that SsrB has on the *Salmonella* genome merits identification of the entire SsrB regulon, which has been attempted by several groups using a combination of genomic and mutagenesis approaches [105, 157-159].

Salmonella Pathogenicity Island-12

Several members of the SsrB regulon are encoded within other SPI. One member, *sspH2,* is encoded within SPI-12. SPI-12 is a 16 kb pathogenicity island located adjacent to the *pro2* tRNA. It appears to be mostly a *S*. Typhimurium-specific locus, as its size and content varies between *S. enterica* serovars, exhibiting evidence of gene loss in serovars such as *S*. Choleraesuis and *S.* Typhi [94, 160]. In *S.* Typhimurium it contains 15-18 ORFs, the majority of which are genes of bacteriophage origin. This remnant phage region appears to encode genes that would have belonged to the late operon of an infecting prophage, as it contains genes necessary for the final stage of phage particle development. These include genes for a Q antiterminator, a head maturation protease, tail fibers, and holin [161]. In addition to bacteriophage genes, SPI-12 encodes *sspH2* and *oafA*, as well as a few genes of unknown function. SspH2 is a well-characterized SsrB-dependent SPI-2 T3SS-associated effector that functions as an E3 ubiquitin ligase when translocated into the host cell [29, 162-164]. It has been found to localize to the apical membrane surface of polarized epithelial cells through host-mediated palmitoylation and interacts with profilin and filamin, proteins involved in actin cytoskeleton rearrangements, *in vitro* [163, 165, 166]. OafA is an O-antigen acetylase that transfers an acetyl group to O-antigen which confers the O5 serotype to *S.* Typhimurium, and thus plays a role in *Salmonella* antigenic variation [167].

Genome-wide mutagenesis experiments have indicated that SPI-12 contributes to intramacrophage survival as well as systemic virulence in murine infection experiments, however these phenotypes are not attributable to *sspH2* and *oafA* alone. Transposon

mutagenesis studies searching for genes important for intramacrophage survival found insertions within SPI-12 that inhibit intracellular survival [66, 168]. A deletion mutant of SPI-12 (Δ*STM2230-STM2245*) was defective in competitive infection experiments in mice, however this phenotype was not attributable to deletion of *sspH2*, as *sspH2* mutants do not exhibit a fitness defect in the murine model of infection [162, 169]. Therefore, in order to understand the role of this pathogenicity island in *Salmonella* virulence, further research is required to understand the contributions of these additional phage genes.

Bacteriophages and Lysogeny

The bacteriophage gene cluster encoded within *S.* Typhimurium SPI-12 includes a putative transcriptional regulator that was implicated in *Salmonella* fitness within the host. This regulator plays an important role in bacteriophage particle development, and thus necessitates an overview of how phages regulate their life cycle. In general, bacteriophage can employ two different approaches to infect and propagate within their host. The first, termed the lytic state, involves replication and reproduction upon infection, leading to eventual production of phage particles and lysis of the infected bacterium. The second approach, termed the lysogenic state, results in phage genome integration into the bacterial chromosome, resulting in a latent infection where reproduction occurs passively by the bacterium through its own cell division processes. In general, these temperate phage (also known as prophage) remain dormant unless stimulated by host cell stress responses into the lytic state [170]. Activation of prophage into the lytic phase of replication is dependent on a regulatory cascade which is

responsible for coordinated expression of the elements necessary for phage particle synthesis. Most of our understanding of this process is derived from the study of bacteriophage lambda (λ).

The genome of λ and other lambdoid phages is organized into a series of operons under tight control of a number of repressors, activators, and antiterminators. The regulators at play: CI, CII, CIII, Cro, N, and Q function during different stages of phage maintenance and development (Figure 1.3) [171, 172]. CI is considered to be the master regulator of the lysogenic state and functions as a repressor to effectively shut down expression of all other λ genes through binding to operator sites in the right promoter (P_R) [173-175]. CII and CIII function as accessory proteins to the repressor and facilitate lysogeny by promoting expression of genes necessary for integration of the λ genome into the bacterial chromosome [171, 176].

The decision to enter into the lytic or lysogenic state is dependent upon several factors. Upon infection of a bacterium, the λ chromosome circularizes and proceeds with expression of the first genes necessary for replication of the phage. This occurs through two primary promoters, P_L and P_R (left and right promoters, respectively). Here, during early gene expression, the antiterminator N and the regulatory protein Cro are expressed and activate genes necessary for DNA replication and recombination (Figure 1.3). Cro specifically suppresses expression of *cI*, the master repressor [171, 177]. N is an antiterminator that forces RNA polymerase through terminator sites within the P_L and P_R operons, resulting in expression of Q, another antiterminator [178, 179]. Q is responsible for the activation of the late P_R' operon, which encodes all genes necessary for synthesis

and assembly of the phage particle, including genes for the head, tail, and tail fibers, as well as those necessary for lysis of the bacterium and thus ultimate release of synthesized phage [180]. Concurrently, Cro functions to repress transcription at P_L and P_R, shutting down early gene expression including *cII* and *cIII*, however sufficient expression of the CII protein prior to Cro repression can lead to expression of the genes necessary for lysogeny, which can ultimately result in integration of the phage genome into the host chromosome [171, 181-183]. Overall, the regulatory proteins of bacteriophage play a vital role in ensuring appropriate timing and expression of phage genes, which ultimately establish their mode of reproduction.

Bacteriophage-Encoded Antiterminators

Antiterminators are proteins that function to overcome termination sequences within RNA transcripts. These termination sites are typically stem-loop structures that trigger dissociation of RNA polymerase (RNAP) from RNA and DNA. There are generally two classes of terminators: intrinsic and factor-dependent. Intrinsic termination sites are found within DNA or RNA sequence and signal elongation complex dissociation. These generally do not require any factors but can sometimes be enhanced by accessory proteins, whereas factor-dependent termination signals depend on the action of a regulatory protein, such as Rho [184].

Lambdoid bacteriophages encode several terminators and gene transcription is dependent upon the function of the two antiterminators N and Q in the λ phage model. These two antiterminators function in strikingly different ways. N interacts with RNA

transcript at sequences called N utilization sites (NUT) and facilitates transcription by suppressing Rho-dependent terminators. It can act alone or through recruitment of a complex that includes the bacterial proteins NusA, NusB, S10, and NusG [185, 186]. N acts on NUT sites upstream of both the left and right terminators (t_L and t_R) at the left and right promoters (P_L and $P_{R)}$ of the λ genome to transition phage gene transcription between the early and middle stages of phage particle development [187, 188]. The action of N ultimately results in the transcriptional activation of Q through P_R [178].

The Q antiterminator mediates the transcription of the late operon of λ phage. It binds to a DNA sequence known as the Q utilization site (*qut*) that is located just upstream of the transcriptional start site of the P_R' promoter and interacts with RNA polymerase to displace σ factor and allow for processive read-through of downstream genes. The *qut* site for λ was found to overlap with the -10 promoter locus of P_R', though Q proteins from the lambdoid phage family have different specificity and differing *qut* site sequences, and thus are not all interchangeable [189-193]. Q is believed to function as a dimer, and encounters RNAP near the P_R' promoter when bound to the *qut* site on the non-transcribed strand. It is here that it facilitates the read-through process by interacting with RNAP through its β subunit, disrupting β-flap interaction with sigma-70 (σ^{70}) region 4 [194, 195]. This disruption allows Q to gain access to RNAP, displacing σ^{70} and remains associated with RNAP to facilitate transcriptional read-through of the entire late operon, which consists of approximately 22 kb [194]. Interestingly, the Q-like antiterminator in Shiga-toxin producing *E. coli* (STEC) has been found to permit

expression of the Stx2 toxin during infection, as the *stx2* gene is encoded within the late phage operon and thus is expressed upon lytic phase induction [196-198].

Bacteriophages as Agents of Horizontal Gene Transfer

Prophages can encode bacterial virulence determinants in addition to their own phage genes. Transmission of horizontally-acquired DNA by bacteriophages occurs through either generalized or specialized transduction. Generalized transduction involves occasional aberrant packaging of bacterial chromosomal DNA rather than phage DNA during phage particle packing [199]. The resultant phage particle is capable of attaching and delivering its genetic content to a subsequent host, however the DNA itself is not viral and thus bears no harm to the new recipient. This DNA can then be introduced into the chromosome of the recipient. This phenomenon has been observed with several phages specific for *Salmonella* and *Listeria* [200, 201]. Alternatively, specialized transduction involves inclusion of bacterial DNA along with the phage chromosome when temperate phage are activated and excised from the host chromosome. Bacterial DNA is then packaged as a unit with the phage genome and integrates into a novel carrier upon lysogeny of the new host [12, 202]. Transduced genes which convert the host bacterium into a pathogenic variant upon lysogeny are called lysogenic conversion genes [24]. This phenomenon is exemplified by toxin-carrying prophage, such as cholera toxin in *Vibrio cholerae* [203].

Transduced bacterial genes are typically encoded in specific locations within the prophage genome and possess autonomous regulatory inputs independent of phage gene

regulators [204]. In Gram-positive bacteria, these genes can be found near both ends of the prophage in proximity to attachment sites as observed for prophage in *Streptococcus pyogenes* [205]. Lambdoid phages in Gram-negative bacteria have also been found to encode lysogenic conversion genes adjacent to the *N*, *Q*, or lysis genes central to the phage genome [198, 206]. They are also commonly found downstream of phage tail genes in the late operon, as observed with many phage-encoded virulence factors in *Salmonella* [27, 207, 208]. While lysogenic conversion genes have been shown to enhance bacterial virulence, there are only a few instances where bacteriophage-encoding genes have similarly contributed. These include platelet adhesins in *Streptococcus mitis* (*S. mitis*) as well as a subset of bacteriocins in *Erwinia carotovora* and *Pseudomonas aeruginosa*, all of which resemble tail and tail fiber components of phage [209-211]. Nonetheless, these examples indicate that even phage-specific genes may contribute to functions within the bacterial host and thus research investigating adapted functions of all horizontally-acquired genes regardless of origin is necessary to expand our understanding of bacterial evolution.

Purpose and Aims of the Current Study

The purpose of this study was to investigate the role of the transcription factor SsrB in integrating intracellular virulence of *Salmonella* Typhimurium. Previous research had shown that SsrB plays a direct role in regulating many genes associated with intracellular virulence, however the regulatory logic controlling its specificity had remained elusive. Moreover, many SsrB-dependent genes exhibit evidence of horizontal

acquisition. Regulatory evolution would thus be expected to play an important role in assimilation of these genes into the SsrB regulatory network. These changes would be best observed within the SsrB DNA recognition sequence and would provide insights as to how modifications to *cis*-regulatory sequences permit incorporation of new genes into existing regulatory networks.

My hypothesis was that **SsrB not only regulates SPI-2 and its associated translocated virulence effectors, but also regulates other gene classes necessary for fitness within the host environment, thus integrating virulence to enhance *Salmonella* pathogenicity.** The aims of this project were to: 1) Define the SsrB regulon through elucidation of the SsrB DNA recognition motif, and 2) Investigate the role of SsrB in the regulation of horizontally-acquired bacteriophage genes within SPI-12.

The specific goals and findings of this study are discussed in the following two chapters:

1) Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system
    - This study identified the SsrB DNA recognition sequence and defined the members of the SsrB regulon that carry this regulatory sequence in their promoters;
2) The SPI-12 remnant phage island contributes to *in vivo* fitness of *Salmonella* Typhimurium

- This study investigated the role of SsrB and STM2239, a Q-antiterminator, in regulation of bacteriophage genes in SPI-12 and established a role for this locus in *S.* Typhimurium virulence.

Identification of the SsrB recognition motif has provided insights into the flexibility and specificity required by this transcription factor for integration of promoters into the SsrB regulatory network. Conversely, investigation of the unique regulatory cascade at play in SPI-12 has demonstrated how foreign genes can be incorporated into virulence programs. These results have demonstrated that integration of horizontally-acquired genes, including those of bacteriophage origin, may be adapted by pathogenic bacteria to expand their virulence arsenal and improve fitness in the host environment.

Figure 1.1 Schematic diagram of *Salmonella* Pathogenicity Island-2. A graphical representation of the SPI-2 genetic locus encoding the type-III secretion system and associated factors. There are four predominant operons as denoted by promoters for *ssrA* (P_{ssrA}), *ssaB* (P_{ssaB}), *sseA* (P_{sseA}), and *ssaG* (P_{ssaG}). Secondary promoters are indicated with smaller promoter symbols (P_{ssrB}, P_{ssaM}, and P_{ssaR}). Genes are coloured according to function as follows: regulatory (red), effector (yellow), structural (orange), chaperone (green), and unknown (white).

Figure 1.1

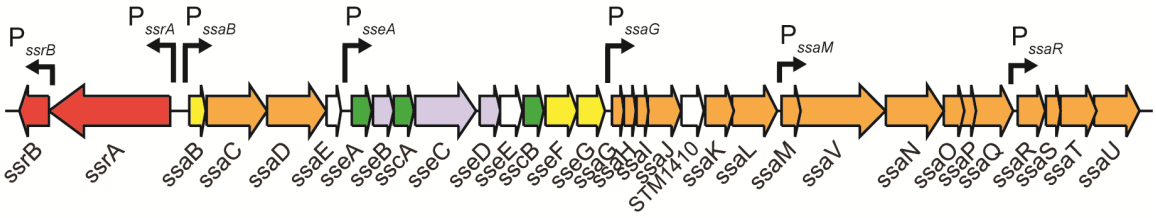

P ssrB
P ssrA
P ssaB
P sseA
P ssaG
P ssaM
P ssaR
ssrB
ssrA
ssaB
ssaC
ssaD
ssaE
sseA
sseB
sscA
sseC
sseD
sseE
sscB
sseF
sseG
ssaG
ssaH
ssaI
ssaJ
STM1410
ssaK
ssaL
ssaM
ssaV
ssaN
ssaO
ssaP
ssaQ
ssaR
ssaS
ssaT
ssaU

Figure 1.2 Schematic of the SPI-2 type-III secretion system. A diagram of the SPI-2 T3SS structure as derived from Kuhle and Hensel (2004) [212]. Proposed locations of selected proteins are indicated. SseB, SseC, and SseD function as the translocon and reside in the host membrane. SsaG, SsaH and SsaI are proposed to compose part of the needle complex. SsaC is the outer membrane secretin. SsaJ connects the secretion apparatus between the inner and outer bacterial membranes. SsaKRSTUV form the inner membrane structure complex. SsaQ forms the cytoplasmic C-ring structure at the base of the secretion apparatus, while SsaN is the proposed ATPase that energizes secretion. SsaBLM function as pH-dependent gate-keepers of secretion. HM: host membrane, OM: outer membrane, PG: peptidoglycan, IM: inner membrane.

Figure 1.2

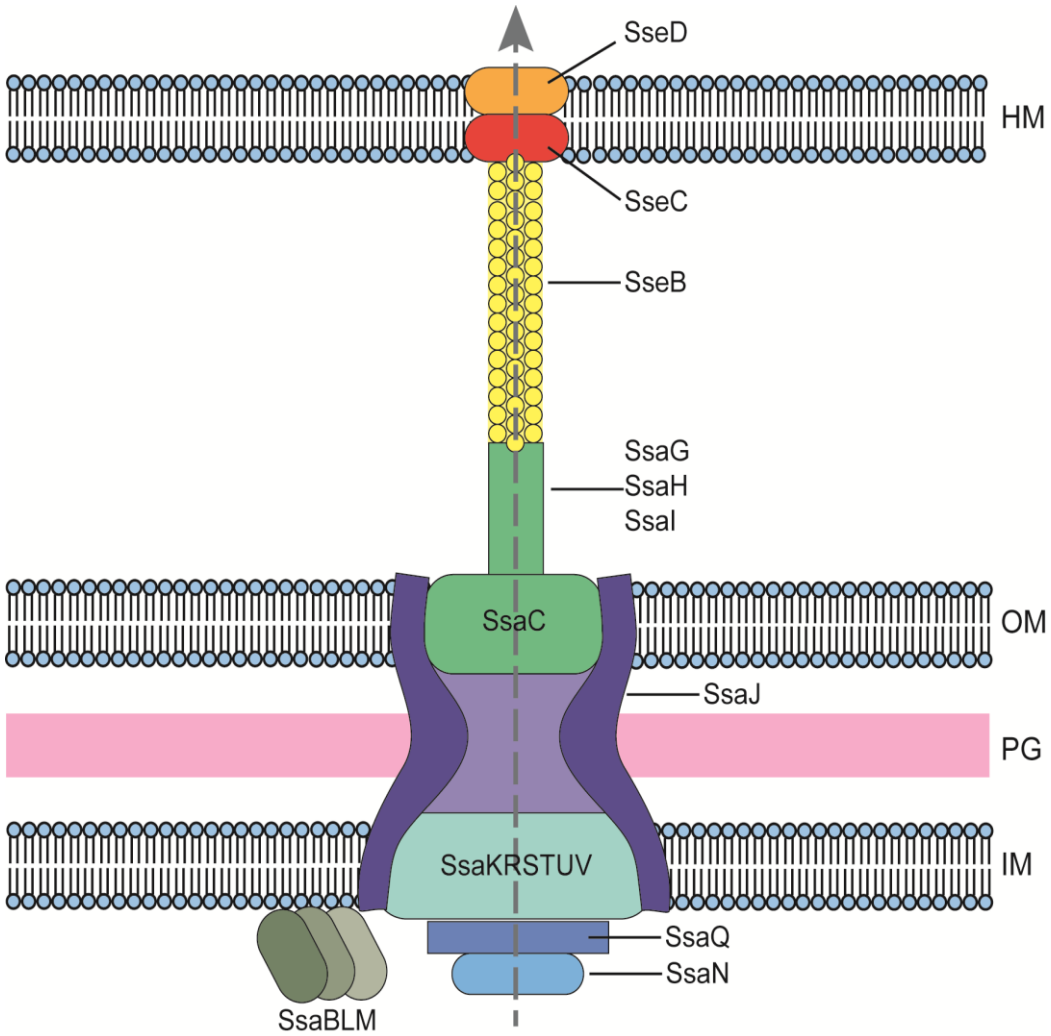
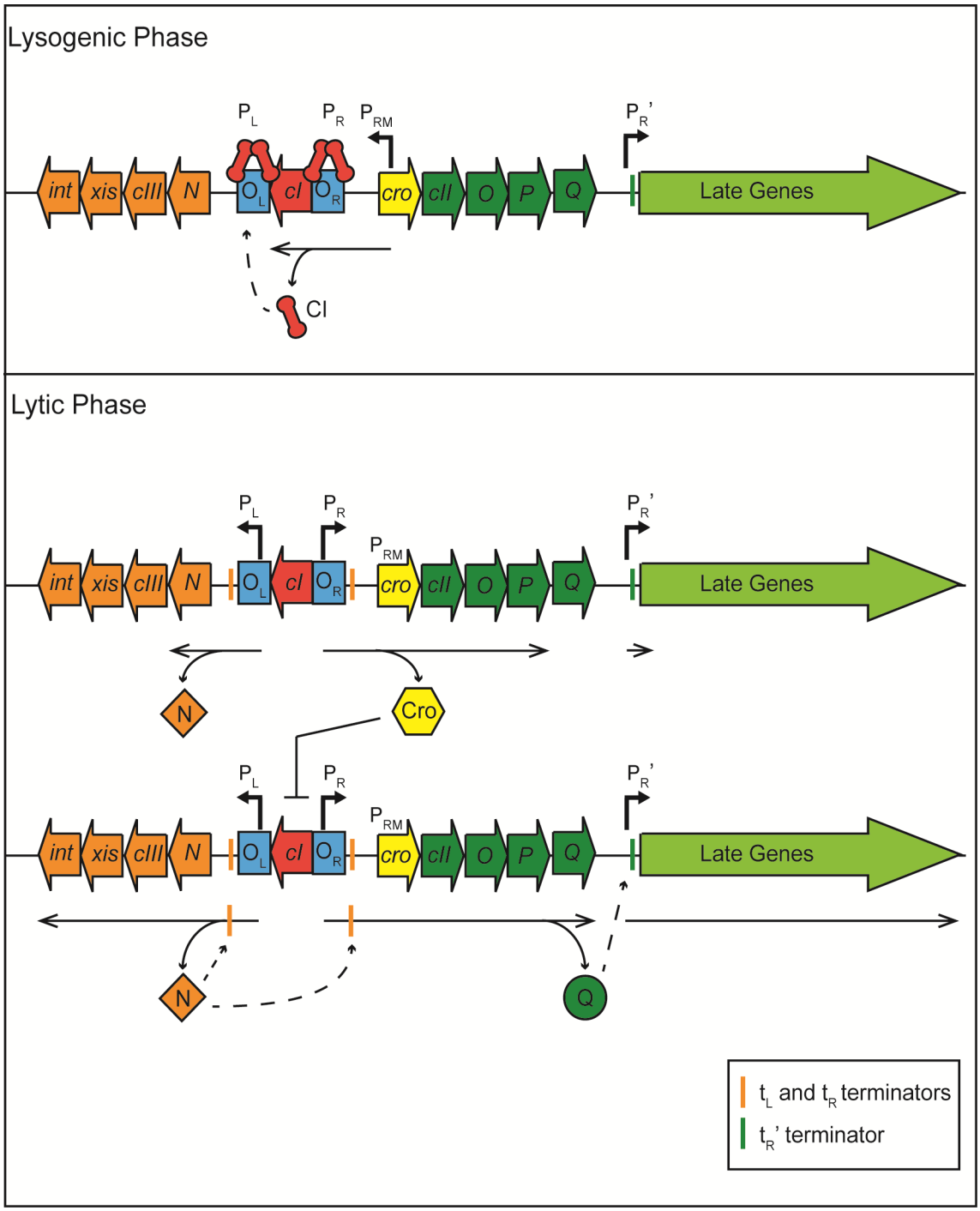

SseD
HM
SseC
SseB
SsaG
SsaH
SsaI
SsaC
OM
SsaJ
PG
SsaKRSTUV
IM
SsaQ
SsaN
SsaBLM

Figure 1.3 Overview of phage λ regulatory mechanisms for lysogenic and lytic states. λ phage gene regulation is tightly controlled during its two reproductive states. During lysogeny, the chromosomally integrated phage λ expresses the CI repressor from the P_{RM} promoter. This repressor binds to operator sites in the left (P_L) and right (P_R) promoters and prevents expression of phage genes, allowing it to remain dormant in its bacterial host. Upon activation of the lytic phase, P_L and P_R become activated and result in expression of the antiterminator N and the repressor Cro. Cro functions to repress expression of the CI repressor which allows the lytic cycle to proceed. Activation of P_L and P_R results in excision and circularization of the phage genome (mediated by Xis) and expression of genes necessary for DNA replication (*O* and *P*), as well as mediates expression of the Q antiterminator. Q is responsible for the activation of the late P_R' operon, which encodes all genes necessary for synthesis and assembly of the phage particle as well as lysis genes that mediate phage particle release. The diagram presented here is a composite from previous publications [171, 213, 214].

Figure 1.3


Lysogenic Phase
P_L
P_R
P_RM
P_R'
int
xis
cIII
N
O_L
cI
O_R
cro
cII
O
P
Q
Late Genes
CI
Lytic Phase
N
Cro
t_L and t_R terminators
t_R' terminator

CHAPTER TWO

Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system

Author's Preface to Chapter Two

In this chapter, I describe the identification of the regulatory logic governing DNA recognition by the transcription factor SsrB. The SsrB DNA-binding motif was found to be a degenerate 18 bp palindrome which is conserved between *Salmonella enterica* and the tse tse fly endosymbiont *S. glossinidius*. Additionally, this study used comparative genomics, transcriptional profiling, and DNA-protein interaction studies to define the SsrB regulon and evaluated how regulatory evolution plays a role in pathogen adaptation. The overall implication of these findings is that the defined palindrome is the minimal architecture needed for SsrB-dependent recognition, and that regulatory evolution has played a role in expansion of the SsrB regulon outside of SPI-2.

The material presented in Chapter 2 was published in the open-access, peer-reviewed journal *PLoS Genetics* in March, 2010. Presented here is the final version of the published manuscript. I shared first-authorship with David T. Mulder for this study. Our individual contributions are outlined in the following co-authorship statement. Of note: There are four additional data sets associated with this publication summarizing the microarray and chromatin immunoprecipitation analyses. These can be freely accessed through the PLoS Genetics website at:
http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000875
All references for this work have been incorporated into the thesis reference list to avoid redundancy.

Chapter Two – Co-authorship statement

Chapter Two consists of the following publication:

Tomljenovic-Berube, A.M., Mulder, D. T. Whiteside, M. D., Brinkman, F. S., and Coombes, B. K. (2010). Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system. PLoS Genet 6: e1000875.

The contributions of each author are outlined as follows:

1) Genome-wide operon and island analysis was conducted by D.T.M.
2) Transcriptional profiling and data analysis was conducted by B.K.C. and D.T.M.
3) Bacterial One-Hybrid experiments were conducted by A.M.T.B. and analyzed by A.M.T.B. and M.D.W.
4) ChIP-on-chip experiments were conducted by A.M.T.B. and bioinformatics analysis was conducted by D.T.M. and A.M.T.B.
5) Transcriptional reporter experiments were conceived and conducted by A.M.T.B. and D.T.M.
6) Sodalis SSR-3 alignments were conducted by D.T.M.
7) Mouse infections were conducted by B.K.C.
8) Manuscript was conceived and written by A.M.T.B., D.T.M., and B.K.C.

Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system

Ana M. Tomljenovic-Berube[1, †], David T. Mulder[1, †], Matthew D. Whiteside[2], Fiona S. L. Brinkman[2], Brian K. Coombes[1]

1 Michael G. DeGroote Institute for Infectious Disease Research and the Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5
2 Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

† These authors contributed equally to this work

Address correspondence to BKC (coombes@mcmaster.ca)

Abstract

Sequence data from the past decade has laid bare the significance of horizontal gene transfer in creating genetic diversity in the bacterial world. Regulatory evolution, in which non-coding DNA is mutated to create new regulatory nodes, also contributes to this diversity to allow niche adaptation and the evolution of pathogenesis. To survive in the host environment, *Salmonella enterica* uses a type III secretion system and effector proteins, which are activated by the SsrA-SsrB two-component system in response to the host environment. To better understand the phenomenon of regulatory evolution in *S. enterica* we defined the SsrB regulon and asked how this transcription factor interacts with the *cis*-regulatory region of target genes. Using ChIP-on-chip, cDNA hybridization and comparative genomics analyses, we describe the SsrB-dependent regulon of ancestral and horizontally acquired genes. Further, we used a genetic screen and computational analyses integrating experimental data from *S. enterica* and sequence data from an orthologous regulatory system in the insect endosymbiont, *Sodalis glossinidius*, to identify the conserved yet flexible palindrome sequence that defines DNA recognition by SsrB. Mutational analysis of a representative promoter validated this palindrome as the minimal architecture needed for regulatory input by SsrB. These data provide a high-resolution map of a regulatory network and the underlying logic enabling pathogen adaptation to a host.

Author Summary

All organisms have a means to control gene expression ensuring correct spatiotemporal deployment of gene products. In bacteria, gene control presents a challenge because one species can reside in multiple niches, requiring them to coordinate gene expression with environmental sensing. Also, widespread acquisition of DNA by horizontal gene transfer demands a mechanism to integrate new genes into existing regulatory circuitry. The environmental awareness issue can be controlled using two-component regulatory systems that connect environmental cues to transcription factor activation, whereas the integration problem can be resolved using DNA regulatory evolution to create new regulatory connections between genes. The evolutionary significance of regulatory evolution for host adaptation is not fully known. We studied the convergence of environmental sensing and genetic networks by examining how the *Salmonella enterica* SsrA-SsrB two-component system, activated in response to host cues, has integrated ancestral and acquired genes into a common regulon. We identified a palindrome as the major element apportioning SsrB on the chromosome. SsrB binding sites have been selected to co-regulate a gene program involved in pathogenic adaptation of *Salmonella* to its host. In addition, our results indicate that promoter architecture emerging from SsrB-dependent regulatory evolution may support both mutualistic and parasitic bacteria-host relationships.

Introduction

Precise gene regulation is crucial to the successful activation and execution of virulence programs for all pathogenic organisms. The acquisition of genes through horizontal gene transfer, a widespread means of bacterial evolution [1], requires a process to integrate new coding sequence into pre-existing regulatory circuitry. Silencing of horizontally-acquired genes by DNA binding proteins like H-NS is one way some incoming genes are initially controlled [2, 3], which can then be subject to regulatory evolution by mutating *cis*-regulatory operator regions to select for optional gene expression. The eventual promoter architecture selected to deploy virulence genes is often modular and should reflect a design that maximizes organismal fitness while limiting fitness trade-offs and antagonistic pleiotropy [215]. Both simulated [216] and functional experiments [217, 218] show that mutation of *cis*-regulatory sequences can be rapid, and that plasticity - the degree to which regulatory mutation can perturb the larger gene network - can be well tolerated in bacterial systems.

Promoter architectures that control quantitative traits such as bacterial virulence are, in fact, modular and evolvable [219]. For instance, *Salmonella enterica* has a multi-faceted pathogenic strategy fine-tuned by several transcriptional regulators. Intracellular survival and persistence of *Salmonella* requires a type III secretion system (T3SS) encoded in a horizontally-acquired genomic island called *Salmonella* Pathogenicity Island-2 (SPI-2) [80, 82]. T3SS are complex secretion machines that deliver bacterial effector proteins directly into host cells through an injectisome during infection [72, 220]. Several ancestral regulators control the genes in the SPI-2 genomic island including the

two-component systems EnvZ-OmpR and PhoQ-PhoP, and the regulatory protein SlyA [221]. SsrA-SsrB is another two-component regulatory system co-inherited by genetic linkage with the SPI-2 locus that is essential for gene expression in SPI-2 [30, 105, 158]. SsrA is a sensor kinase activated in the host environment that phosphorylates the SsrB response regulator to create an active transcription factor needed for spatiotemporal control of virulence genes [221, 222]. In the *Salmonella* genus, the SPI-2 genomic island is found only in pathogenic serotypes of *Salmonella enterica* that infect warm-blooded animals and is absent from *Salmonella bongori*, which colonizes cold-blooded animals [73]. It is generally accepted that SPI-2 was acquired by *Salmonella enterica* after divergence from *S. bongori*, providing a useful pedigree to study regulatory evolution influenced by SsrB.

We recently demonstrated the evolutionary significance of *cis*-regulatory mutations for pathoadaptation of *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium) to an animal host [7]. Our focus was on SsrB because of its broad conservation among the pathogenic Salmonellae and its essentiality for animal infection, suggesting it coordinates fundamental aspects of *Salmonella* pathogenesis beyond the SPI-2 genomic island. In this study we investigated how regulatory evolution assimilates horizontally acquired and ancestral genes into the SsrB regulon on a genome-wide scale using an integrated set of experimental methods. Combining our data with previous biochemical work, along with comparative genomic analyses with an orthologous T3SS-encoding genomic island in the tsetse fly endosymbiont, *Sodalis glossinidius*, we reveal the flexible DNA palindrome that distributes SsrB in the genome to influence

transcriptional activation of the SPI-2 T3SS and almost all of its accessory effector proteins. Our data uncovers the principal SsrB circuitry that appears to have been conserved to support multiple bacterial lifestyles, including parasitic and mutualist symbioses.

Results

Transcriptional profiling of an *ssrB* mutant

To begin to understand regulatory evolution and network expansion of the SsrB response regulator, we profiled the transcriptome of an *ssrB* mutant and compared it to *S.* Typhimurium wild type cells grown in an acidic minimal medium that activates the SsrA-SsrB two-component regulatory system [139]. We identified 133 genes that were significantly down-regulated in the *ssrB* mutant [$z < -1.96$] [223](Table 2.1). This included almost all genes in the SPI-2 genomic island as well as effector genes encoded throughout the genome (Dataset S1). Next, we performed a Clusters of Orthologous Groups (COG) analysis [224] on the 118 genes that had an ortholog in the annotated genome of *S.* Typhimurium strain LT2 [161]. Among these, 45 genes lacked a functional COG assignment and the 73 remaining genes were distributed among 86 COGs (Figure 2.1). The majority of functions in the latter groups are in transport, secretion, and trafficking of cellular components in addition to protein and membrane modification.

The SsrA-SsrB system was acquired by horizontal gene transfer into the *S. enterica* species after divergence from what is now extant *S. bongori*. As such, *S. bongori* has evolved in the absence of SsrA-SsrB and its regulatory architecture has not been influenced by it. Orthologous genes ancestral to both species but regulated by SsrB in *S. enterica* provide evidence for network expansion and regulatory evolution that we previously showed can be mapped to a single *cis*-input location by using functional and comparative genomics [7]. To expand on this, we used a reciprocal BLAST-based analysis and identified 47 orthologs in *S. bongori* among the 133 genes whose

transcription was down-regulated in an *ssrB* mutant (Table 2.1). In Δ*ssrB* cells, the mean fold-change of the orthologous genes (-6.1-fold) was subtler than for the *S. enterica*-specific gene set (mean -21.2-fold), which included the T3SS and associated effector genes (Dataset S1). We also determined the distribution of down-regulated genes among genomic islands [225], including prophages, pathogenicity islands (SPI-islands) and additional regions of difference (ROD) between *S. enterica* and *S. bongori* (Dataset S2). For this we used a BLAST-based comparison of genome-wide synteny between *S.* Typhimurium and *S. bongori* and identified 50 ROD that included 17 previously reported SPI-islands and prophages. Of the 133 down-regulated genes identified, 56 were present within genomic islands (Table 2.1), with a mean change in gene expression of -25.3-fold in Δ*ssrB* cells.

Genome-wide SsrB interactions

To examine SsrB allocation on the chromosome *in vivo*, we isolated functional SsrB-DNA interactions using chromatin immunoprecipitation and examined the bound DNA by chip analysis (ChIP-on-chip) using an *S.* Typhimurium SL1344 array containing 44,021 probes. With this method, we identified 256 significant interaction peaks distributed throughout the genome that were enriched under SsrB-activating conditions and with interaction scores three standard deviations greater than the mean probe score (Figure 2.2A and Table 2.1). Of these 256 peaks, 126 (49%) occurred within coding regions of genes (CDS) and 130 (51%) were in intergenic regions (IGR). Given the strong influence of SsrB on horizontally acquired genes (Table 2.1), we plotted the ChIP-on-chip

data against all genomic islands in *S.* Typhimurium SL1344. From this analysis, 62 of 256 SsrB binding peaks (24.3%) occurred within genomic islands (Figure 2.2B).

SsrB ChIP peaks were observed upstream of previously identified SsrB regulated genes indicating that our ChIP-on-chip data captured functional interactions (Dataset S3). To generate a consensus set of SsrB regulated genes, we performed an analysis to identify operons in the *S.* Typhimurium SL1344 genome that encoded at least one gene down-regulated in Δ*ssrB* cells and that possessed an SsrB binding peak in the upstream regulatory region as defined by our ChIP-on-chip analysis. From this, the 133 down-regulated genes mapped to 86 operons, 49 of which had an SsrB interaction upstream or within the first gene of the operon (Table 2.2). This analysis captured all five reported operons in the SPI-2 genomic island in addition to ten operons outside of this island that encode SPI-2 translocated effectors.

SsrB binds to six promoters within the SPI-2 genomic island

In order to rigorously evaluate our genome-wide functional genomics data, we compared it against traditional biochemical experiments describing SsrB-DNA interactions at the SPI-2 locus. Previous data reported SsrB footprints upstream of 6 genes in SPI-2: *ssrA, ssrB, ssaB, sseA, ssaG,* and *ssaM* [101, 102]. Our ChIP-on-chip data showed discrete SsrB binding at all of these promoters except for the promoter reported to be between *ssrA* and *ssrB* [101] (Figure 2.3). We attempted to verify functional activity at this site, but could not using transcriptional fusions (data not shown). Our data also identified three additional SsrB binding peaks upstream of *sseA*, within the CDS of *ssaJ*,

and in the IGR upstream of *ssaR* (Figure 2.3). Functional interactions were confirmed for *sseA* and *ssaR* in subsequent reporter experiments described below.

Identification of a functional DNA element for SsrB binding

Previous attempts by others to identify a conserved SsrB DNA recognition motif have been unsuccessful. To overcome this, we employed a bacterial one-hybrid screen originally developed to define binding site preferences for eukaryotic transcription factors [226]. We fused the DNA binding domain of SsrB (SsrBc) to the α-subunit of RNA polymerase and screened a prey library of ~10^8 DNA molecules previously counter-selected against self-activation (Figure 2.4A). We used the PhoP response regulator from *E. coli* as a control because a DNA recognition sequence for it was known [227]. Bait-prey combinations surviving selection on medium lacking histidine were purified, and preys were sequenced and analyzed using the motif-finding program MEME [228]. From 189 unique sequences isolated for SsrBc, over 80% contained a degenerate consensus motif, mCCyTA (Figure 2.4B). In control screens with the PhoP-αNTD fusion, the PhoP box sequence (G/T)GTTTA was identified in 11% of sequenced preys (12/109, data not shown) but this sequence was never captured by SsrB-αNTD and vice versa, demonstrating specificity of the bacterial one-hybrid system for prokaryotic regulatory proteins.

Next, we examined our ChIP-on-chip data for the presence of a conserved regulatory motif. We extracted sequence data from the local maxima of the 256 binding peaks and analyzed the sequences with the computational program MDscan [229]. Using

the highest-ranking probes to generate an initial prediction followed by lower-ranking probes for refinement, this analysis identified motifs that represented either the forward or the reverse complement of the consensus sequence ACmTTA, which shares consensus with the motif identified in the bacterial one-hybrid screen (Figure 2.4B). We identified variations of this motif within footprinted regions of SsrB-regulated promoters [101, 102], however sequence degeneracy made it difficult to precisely map the input functions.

Identification of a functional DNA palindrome conserved in *S.* Typhimurium SPI-2 and the *Sodalis glossinidius* SSR-3 region

The analysis of regulatory evolution is particularly challenging because it is difficult to distinguish neutral stochastic change from functional divergence. To solve this problem in the context of mapping the SsrB binding element, we used comparative genomics to search for conserved promoter architecture in another organism with a similar genomic island to *Salmonella* SPI-2. The tsetse fly endosymbiont *Sodalis glossinidius* contains the *Sodalis* Symbiosis Region-3 (SSR-3) that is similar in content and synteny with the *S. enterica* SPI-2 locus [230]. Gene conservation includes the entire T3SS structural module extending to the regulatory genes *ssrA-ssrB* and all other genes except the effectors *sseF* and *sseG*, which are not present in *Sodalis* SSR-3. We aligned the sequences of the five mapped promoters in SPI-2 with the orthologous SSR-3 regions to identify local conservation. Highly conserved sites within the promoters were restricted to regions previously footprinted by SsrB [101, 102], whereas adjacent sequence showed substantial drift (Figure 2.4C). Within the conserved sites we identified a heptameric

sequence in 7-4-7 tail-to-tail architecture that created an 18-bp degenerate palindrome. This palindrome was found in all SPI-2 and SSR-3 T3SS promoters with the exception of the *sseA* promoter that had only one reasonably well-conserved heptamer in the footprinted region (Figure 2.4C and Figure 2.S1). Interestingly, two copies of the palindrome occur upstream of the *ssrA-ssrB* operon in *S.* Typhimurium within the same footprint, and the conservation of either site in *Sodalis* was weak. Evaluation of the heptamer motif in the palindrome showed high similarity to the motifs identified by the bacterial one-hybrid screen and the ChIP-on-chip experiments (Figure 2.4B and 2.4D), giving us confidence that we had identified the major recognition module for transcriptional input by SsrB. In accord with a previous observation [102], there was not a strict requirement in the spacing between the SsrB binding site and the downstream transcriptional start site.

SsrB directs transcriptional input using a flexible palindrome architecture

The presence of a conserved palindrome sequence in SPI-2 promoters and in related sequences from the endosymbiont *S. glossinidius* suggested that regulatory input by SsrB was through a palindrome sequence architecture. However, other lines of evidence suggested that the recognition site architecture was flexible in nature: (*i*) our bacterial one-hybrid screen isolated functional single hexamer sequences, (*ii*) the SsrB footprint at the naturally evolved *sseA* promoter within SPI-2 [102] contained only one reasonably well-conserved heptamer, and (*iii*) degenerate or non-ideal palindromes exist in the genome. In order to deconstruct this architecture, we designed a set of experiments

to test the palindrome's tolerance to mutation. We chose the *ssaG* promoter for these experiments because 16 of 18 bases were identical between SPI-2 and SSR-3 from *S. glossinidius*, differing only in the 4-bp spacer between heptamers (Figure 2.4C). We mutated the palindrome in a series of transcriptional reporters that were otherwise identical to the evolved *ssaG* promoter (Figure 2.5A) and promoter activity was compared to that of a wild type palindrome sequence. Variants in which the first half-site (7') or second half-site (7") was deleted produced similar transcriptional activity to the wild type palindrome, verifying that a single well-conserved heptamer is sufficient for transcriptional input under these experimental conditions (Figure 2.5B). Deletion of the 4-bp spacer sequence between the heptamers - the most degenerate element of the palindrome - also generated wild type promoter activity. However, the orientation of individual heptamers was essential for transcriptional input since rewiring the palindrome in any head-to-head orientation produced negligible promoter activity. However, if the two half sites were swapped front-to-back so that they maintained tail-to-tail orientation (construct labelled "Reverse" in Figure 2.5), wild type promoter activity was restored. Precise deletion of the entire 18-bp palindrome lead to ~10% residual activity in wild type cells, which was reduced to less than 1% in an *ssrB* mutant (Figure 2.5C). To determine whether the remaining 10% transcriptional activity was a result of an SsrB-dependent feed-forward mechanism or transcriptional read-through of our chromosomally integrated reporter, we constructed an ectopic deletion reporter. Assessment of reporter activity for this construct in addition to wild type constructs in wild type and *ssrB* mutant backgrounds showed a similar level of activity to the *ssrB* mutant (Figure 2.5D).

The results for the half-site deletion constructs, which retained activity similar to wild type, were unexpected. Therefore, we compared the sequences generated upon mutation against a consensus palindrome matrix generated from all SPI-2 and other identified putative elements. The 7'-4-X, X-4-7" and 7'-X-7" mutations introduced a number of base transitions and transversions never occurring in the matrix, however the modified 7 base pair heptamer retained 4-5 naturally-occurring bases along with the unchanged wild type sequence in the other heptamer (Figure 2.S2). The possibility existed that this modified heptamer, although now weaker in consensus, could still be sufficient for recruitment of a functional form of SsrB when paired with the other wild type heptamer. To test this, we created an additional ectopic transcriptional fusion construct in which the left half (7') of the palindrome was mutated to bases never occurring in the consensus matrix. When tested in promoter activity experiments, this reporter was unable to activate transcription and was identical to the X-X-X mutant construct (Figure 2.5D).

Salmonella SsrB and the *Sodalis* ortholog (*SG1279*) are 69% identical and 81% similar at the amino acid level. All of the critical residues in the dimerization helix and HTH motif required for specific transcriptional activity by SsrB [152] are conserved in the *Sodalis* ortholog (Figure 2.S3). To demonstrate a functional role for the palindrome identified in *Sodalis*, we engineered luciferase transcriptional reporters that either contained (7'-4-7") or lacked (X-X-X) the identified palindrome from the *Sodalis SG1292* promoter (*ssaG* ortholog) and transformed them into wild type *S.* Typhimurium and an *ssrB* mutant. The transcriptional activity from a wild type *Sodalis* palindrome

sequence was high, but was completely abolished in an *ssrB* mutant and in experiments where only the palindrome sequence was precisely deleted (Figure 2.5D). These experiments demonstrated a functional role for the conserved palindrome in *Sodalis* and the requirement for SsrB for transcriptional activation.

Regulatory evolution of the SPI-2 T3SS effector repertoire

The above results identified the conserved, yet flexible, palindrome sequence defining DNA recognition by SsrB. To examine the extent to which regulatory evolution has been selective for this genetic architecture, we created a position weight matrix (PWM) for the five strongest palindrome sites in SPI-2 and the orthologous sites in *Sodalis* SSR-3. We then searched for representative candidates of this motif in the *S.* Typhimurium genome using the simple scoring algorithm MotifLocator [231, 232]. This analysis recovered the motifs upstream of *ssaB, ssaG, ssaM,* and *ssaR* that were used to generate the PWM. The palindrome in the *ssrA* promoter was not used to create the PWM due to its weaker consensus in the left heptamer, however, it was recovered in the computational analysis in a second group of lower-scoring motifs (Figure 2.6A). We identified 24 palindromes co-occurring with ChIP-on-chip peaks upstream of 24 different SsrB-regulated genes or operons. Applying a stringent threshold to the output allowed us to identify two groups - genes with high-scoring upstream palindromes (*ssaB, ssaG, ssaM, ssaR, sopD2, sifA, sifB, sseK2, sseK3, sseL, sseA', steC,* and *srcA*) and those with medium-scoring palindromes (0.7-0.8 threshold; *ssrA, STM1633, sseI, slrP sspH2, pipB, sseJ, pipB2, srfN, sseA* and *steB*) (Figure 2.6A and Dataset S4) (*sseA'* denotes the SsrB

palindrome sequence upstream of *sseA* that falls within the *ssaE* CDS, while *sseA* refers to the SsrB-footprinted IGR site with only one conserved heptamer defined in Figure 2.4C). Remarkably, this accounted for 17 of 22 SL1344 genome-encoded effectors translocated by the SPI-2-encoded T3SS (exceptions are chromosomal *steA, gogB,* and *sseK*, and plasmid-encoded *spvB* and *spvC*). These genes either lacked an upstream ChIP peak above our 3-standard deviation cut-off (*sseK*) or had such a peak but did not reach statistical significance in our transcriptional profiling experiments (*steA, gogB*).

Our ChIP-on-chip data revealed three additional strong SsrB binding peaks within SPI-2: one in the IGR directly upstream of *ssaR*, a second within the CDS for *ssaJ*, and a third within the CDS for *ssaE* that would be predicted to influence transcription of the downstream effector/chaperone operon beginning with *sseA*. The analysis described above recovered SsrB palindrome sequences at the *sseA*' and *ssaR* locations prompting further validation of these sites. No palindrome was identified for the *ssaJ* interaction peak and so further characterization was not pursued. For the IGR palindrome upstream of *ssaR*, we tested both a chromosome-integrated transcriptional fusion and an autonomous episomal reporter. In wild type cells these reporters were as active as other SPI-2 promoters, whereas promoter activity was abrogated in Δ*ssrB* cells, implicating this IGR as a functional promoter for *ssaR* (Figure 2.6B and Figure 2.S4). We next tested the function of the intragenic palindrome within *ssaE* (*sseA'*). For this, we constructed a single-copy transcriptional reporter that either contained (WT P*sseA*) or lacked (P*sseA* del) the single heptamer site in the *sseA* IGR and integrated this reporter into wild type cells and mutants lacking either *ssrB* or the *ssaE* coding sequence that removed the high-

scoring intragenic palindrome *sseA'*. These experiments showed that the *sseA'* sequence contributes approximately 75% of transcriptional output at the *sseA* promoter (Figure 2.6C), since deleting the single heptamer in the *sseA* IGR had little effect on transcriptional output in any of the strain backgrounds. These reporter data are in line with the respective binding scores for the ChIP-on-chip interaction peaks (Figure 2.3) and the sequence similarity for these elements with respect to the consensus palindrome (Figure 2.6A and Figure 2.4D). Together, these data provide compelling evidence for the identity of the DNA recognition element that has been selected through evolution to co-regulate an SsrB-dependent gene program involved in adaptation to a host.

Discussion

Horizontal gene transfer is a well-recognized mechanism of bacterial evolution that gives rise to new phenotypes due to the coordinated expression of novel genetic components [1]. A good example of this is acquisition of type III secretion by mutualists and pathogenic bacteria enabling new colonization strategies within a host [233, 234]. Evolved changes to regulatory circuitry can also give rise to phenotypic diversity at the species level [7]. In both cases, regulatory evolution is required to correctly deploy gene products during infection, yet the extent to which regulatory evolution contributes to pathogenic adaptation is only beginning to be realized [219]. The SsrA-SsrB two-component regulatory system in *S. enterica* has been the focus of our efforts to understand the significance of regulatory evolution for pathogenic adaptation. This regulatory system was co-acquired with a T3SS encoded in the SPI-2 pathogenicity island and likely contributed to immediate and gradual phenotypic diversity as new regulatory nodes were explored and acted upon by natural selection.

Extensive work has been reported on the characterization of SsrB dependent genes, including functional evaluation of genes encoded within SPI-2 in addition to genome-wide transcriptional studies [105, 158]. In this study we identified genes co-expressed under SsrB-inducing conditions and found those with strong levels of expression localized predominantly to mobile genetic elements, recently acquired genomic islands or other annotated islands. We also identified many weakly co-expressed genes, some of which may represent ancestral *Salmonella* genes recruited into the SsrB

regulon like the previously reported *srfN* [7]. Some of these genes may not be directly regulated by SsrB and will require further experimental investigation.

Direct profiling of SsrB-DNA interactions using ChIP-on-chip was used to identify SsrB binding sites in the genome. This analysis identified many interactions which have not been previously described and interaction sites within coding regions of genes which may represent non-canonical functions for SsrB. Other groups have reported the existence of similar numbers of ChIP-on-chip interactions within intragenic regions for other transcription factors [235, 236] suggesting that this phenomenon is not restricted to SsrB. In light of the disparate number of microarray genes in comparison to ChIP-on-chip peaks we attempted to generate a more comprehensive picture of the SsrB regulon by combining these data sets at the operon level. In doing so we believe that the nineteen operons containing differentially expressed genes determined by microarray and containing a ChIP-on-chip peak three standard deviations above the mean captured by this analysis represent the genes directly activated by SsrB (Table 2.2). Those operons having a ChIP-on-chip peak directly upstream in the IGR region encompass the majority of known SsrB regulated genes while those possessing a ChIP peak within the CDS of the first gene may represent non-functional interactions that deserve follow-up experimental investigation.

The ChIP-on-chip data not only provided information on the identities of SsrB-regulated genes but also gave insight as to the identity of the SsrB recognition element specified by the interaction site sequences. The regulatory architecture governing SsrB input has been elusive despite several SsrB footprints being defined biochemically [7,

101, 102]. Our ChIP-on-chip data further suggested that SsrB binding within SPI-2 was specific, with binding peaks overlapping precisely with regions of the DNA footprinted by SsrB [101, 102]. By using a genetic screening strategy together with functional and comparative genomics, we were able to define the essential SsrB regulatory element as being an 18-bp palindrome with a conserved 7-4-7 internal organization.

In support of the palindrome as the functional entity we showed the loss of SsrB dependence as a result of deletion of this element for the *ssaG* promoter. Evaluation of the 7-4-7 palindrome in the *ssaG* promoter revealed the minimal architecture and sequence orientation required for transcriptional input. Deletion of the entire palindrome resulted in less than 1% activity in wild type cells, an equivalent level of activity to those lacking *ssrB* entirely. A search of the *S. enterica* genome for this palindromic motif revealed candidates upstream of the previously noted SsrB dependent genes, including two additional SPI-2 sites; one IGR site upstream of *ssaR* that until now had been cryptic, and one intragenic palindrome upstream of *sseA* in the *ssaE* CDS. In both cases these input sites were found to be functional.

Although palindrome architecture was conserved upstream of SsrB-regulated genes, degenerate palindromes in which one half-site was more conserved were also functional. As a result of our mutational analyses we conclude that so long as the orientation of a single heptamer of the palindrome is conserved with respect to the downstream gene, SsrB is tolerant of degeneracy in the adjacent spacer and heptamer sequences. While we were able to identify a number of limited palindrome-like sequences from our bacterial one-hybrid screen, this tolerance in addition to the library size required

to pull out an 18-bp palindrome in large numbers may explain why we isolated functional single heptamer sequences and why degenerate palindromes naturally exist in the genome. A recent report by Carroll et al, postulated that SsrB first interacts with DNA as a monomer, followed by dimerization [152]. Our findings also suggest that dimerization is likely required for transcriptional activation however strong recognition by one monomer may stabilize interaction of a second monomer with a less than ideal sequence. The finding that a flexible palindromic sequence can be selective for SsrB input raises many interesting questions around the nature of regulatory evolution. The ability to use a short functional half-site adjacent to an uncharacterized threshold level of tolerated bases would reduce the period of neutral evolution required to generate an inverted repeat sequence twice the length [237], and would limit the loss of intermediate variations to drift while a more desirable palindrome is created by regulatory evolution. For bacteria that make use of horizontal gene transfer, this could increase the tempo with which new DNA is integrated into the regulatory circuitry of the cell.

We showed that the SsrB regulatory palindrome is also present in the orthologous SSR-3 island of the endosymbiont *Sodalis glossinidius* and that the palindrome evolved in *Sodalis* can act as a *cis* regulatory input function in *Salmonella.* Thus, in addition to supporting a pathogenic lifestyle within a host in *Salmonella*, it seems probable that this common promoter architecture may direct the activation of the SSR-3 T3SS of *S. glossinidius* in its endosymbiotic relationship with the tsetse fly host, although we acknowledge this requires experimental validation. The SSR-3 region in *S. glossinidius* is fully conserved in gene synteny and content with that of SPI-2 [230], with the exception

of two effector genes missing in SSR-3 (*sseF* and *sseG*) that are required to localize vacuolar *Salmonella* to the perinuclear Golgi in host cells [238, 239]. The SsrB ortholog in *S. glossinidius* is ~30% divergent with SsrB at the protein level, initially leading us to think that they might have different binding site preferences. To the contrary, high local conservation in the promoters evolved in *Salmonella* and *Sodalis* was the crux in defining the functional SsrB input among stochastic noise. This analysis revealed strong palindrome sequence conservation in five promoters identified in SPI-2 and in the orthologous sequences in *Sodalis* SSR-3.

Among palindrome-containing promoters, the *ssrA* promoter is exceptional for two reasons: a lack of conservation between *Salmonella* and *Sodalis*, and the evolution of tandem palindromes in *Salmonella*. One possible interpretation of this divergent regulatory architecture in front of *ssrA* might relate to bacterial lifestyle. *Salmonella* may have retained or evolved SsrB input here to create a positive feedback loop on the regulatory system to rapidly adapt to the host environment during infection, similar to transcriptional surge described for the PhoP response regulator [240]. The endosymbiotic relationship of *Sodalis* with the tsetse fly - where long-term vertical transmission has ostensibly been formative in shaping regulatory circuitry at certain promoters - may obviate the need for rapid transcriptional surge, leading to regulatory drift or selection against positive feedback. With the structure of SsrB available [152] and its recognized sequence now identified, future studies will be able to build a picture of how SsrB interacts with both its target DNA, RNA polymerase and potentially other transcription

factors including nucleoid associated proteins in order to direct transcription of its regulon.

In summary, this work highlights the evolutionary significance of *cis*-regulatory mutation for the adaptation of *Salmonella* to a host animal. The DNA module that choreographs SsrB-mediated pathogenic behaviour in *Salmonella* appears to have been conserved for mutualism as well, thereby shedding new light on the significance of *cis*-regulatory mutations for bacteria evolving in different ecological settings.

Methods

Ethics statement

All experiments with animals were conducted according to guidelines set by the Canadian Council on Animal Care. The Animal Review Ethics Board at McMaster University approved all protocols developed for this work.

Bacterial strains and growth conditions

The *Salmonella* strain used for microarray and ChIP-on-chip analysis was *Salmonella enterica* serovar Typhimurium strain SL1344. Bacterial strains and plasmids used in this work are described in Table S1. Primer sequences used to generate constructs are available upon request. Bacteria were grown in LB medium unless otherwise indicated. Low-phosphate, low magnesium (LPM) medium was used as bacterial growth medium for microarray, ChIP-on-chip, and transcriptional reporter experiments [139]. Liquid cultures were routinely grown at 37°C with shaking. Antibiotics were added to media as follows when necessary: ampicillin (Amp, 100 μg/mL), chloramphenicol (CM, 34 μg/mL) kanamycin (Kan, 30 or 50 μg/mL), and streptomycin (SM, 50 μg/mL). NM medium was used in the bacterial one-hybrid experiments as described previously [226].

Transcriptional profiling

Microarray experiments were conducted and analyzed as described previously [30]. cDNA was synthesized from RNA harvested from wild type cells and an Δ*ssrB* mutant. cDNA from 2 replicate experiments was hybridized to InGen arrays and analyzed using ArrayPipe version 1.309 [223]. Probe signals underwent a foreground-background correction followed by a printTipLoess normalization by sub-grid. Duplicate spots were

merged followed by averaging of the two replicates. Local intensity z scores were calculated for determination of significance.

Genome-wide operon and island analysis

For operon analysis, *S.* Typhimurium SL1344 operons were defined as groups of genes encoded on the same strand with a maximum intergenic distance of 30-bp. Operons selected for further investigation were those possessing at least one significantly down-regulated gene from the cDNA microarray analysis of an *ssrB* mutant. A cDNA microarray score was assigned based on the average fold-change value of all genes within the operon. For ChIP-on-chip analysis, a top ChIP interaction score was defined as that of the highest scoring probe within the first gene or the intergenic region upstream of the first gene of the operon. For the analysis of regions of difference (ROD) between *S. enterica* serovar Typhimurium and *Salmonella bongori*, a reciprocal-best BLAST analysis was performed to identify orthologous genes between *S.* Typhimurium and *S. bongori*. Orthologs were defined as reciprocal best BLAST pairs with E-values less than 0.005. Comparison of gene synteny between regions encoding orthologous genes was performed to identify regions of low conservation including gene deletions and insertions. The location and names of genes flanking the comprehensive list of genomic islands is provided in Dataset S2 and were compared to those predicted using IslandViewer [241].

Bacterial one-hybrid screen and analysis

The bacterial one-hybrid (B1H) experiments were conducted as outlined previously using a single-step selection procedure [226]. Full-length *phoP* from *E. coli* and the C-terminal domain of *ssrB* (*ssrBc*) from *S*. Typhimurium were cloned into pB1H1 to create a fusion

to the αNTD of RNA polymerase. Each bait vector was transformed into *E.coli* Δ*hisB* Δ*pyrF*, purified, and then cells were transformed again with purified prey library that was previously counter-selected for self-activating preys using 5-fluoro-orotic acid. Transformants were recovered for 1 h in SOC medium, washed with NM medium supplemented with 0.1% histidine (NM + his) and allowed to grow for 2 h in this medium. Cells were washed four times with water, once with NM medium lacking histidine (NM –his), then resuspended in NM –his and plated on 150 x 15 mm dishes containing NM –his media supplemented with either 1 mM (for PhoP screen) or 5 mM (SsrBc screen) 3-aminotriazole. Selection was for ~48 hours at 37°C. Individual clones were selected from plates containing <600 colonies, the prey plasmids were isolated and sent for sequencing (Macrogen USA). Sequences were parsed to extract the 18-bp prey sequence, then inputted into MEME (version 4.1.1) for motif generation [228]. MEME was run with default parameters and included searching for motifs of length 5-17 bp in either forward or reverse direction and with no limit on the number of occurrences within an input string. Motif logos were generated using Weblogo, version 2.8.2 [242].

Chromatin immunoprecipitation-on-chip (ChIP-on-chip)

Chromatin immunoprecipitation-on-chip (ChIP-on-chip) was conducted as described previously using an SL1344 strain containing an *ssrB-3xFLAG* allele on the chromosome [7]. The primer sequences used to generate the DNA for recombination were: 5'GAG TTA CTT AAC TGT GCC CGA AGA ATG AGG TTA ATA GAG TAT GAC TAC AAA GAC CAT GAC GG3' and 5'ATC AAA ATA TGA CCA ATG CTT AAT ACC ATC GGA CGC CCC TGG CAT ATG AAT ATC CTC CTT AG3'. This strain was

generated by an allelic replacement method described previously [243] and causes lethal infection of C57BL/6 mice similar to wild type SL1344 (Figure S5). Immunoprecipitated DNA from three experiments under SsrB-inducing conditions (LPM growth medium) and one experiment under non-inducing conditions (exponential growth in LB medium) was hybridized to a single chip printed with four whole genome arrays designed on *S. enterica* serovar Typhimurium strain SL1344 (Oxford Gene Technology, Oxford UK). Signals for each probe within an experiment were normalized to the median channel signal for the respective array. Signal ratios were obtained for both inducing and non-inducing conditions by calculating the ratio of the control probe value and experimental probe value. A final interaction score was obtained by taking the $\log_2$ value of the ratio between the non-inducing and inducing conditions for each probe to remove SsrB interactions that occur under non-inducing conditions. Parsing and data analyses were performed using the Python scripting language. Genome-wide ChIP-on-chip data was plotted using Circos v.0.51 [244].

Genome-wide motif analysis

ChIP probes were ordered according to their position on the *S.* Typhimurium SL1344 genome and local maxima for ChIP interaction scores were defined as interaction peaks. Peaks with scores greater than three standard deviations from the mean probe signal were considered significant ChIP interaction peaks and were ranked in order of descending interaction score. The sequence of the top-scoring probe for each peak was exported to a text file and used for analysis by MDscan [229]. The background parameter was run with output generated by the included genomebg program from the *S.* Typhimurium SL1344

genome sequence. The initial motif was generated from sequences from the top ten SsrB interaction peaks and refined using the top 25 peak sequences.

To identify instances of the palindrome motif in *S.* Typhimurium, ten 18-bp 7-4-7 palindromic motifs in the promoters of the SPI-2 genes *ssrA*, *ssaB*, *ssaG*, *ssaM*, *ssaR* and their orthologous SSR-3 genes were used as input for MDScan to identify a consensus motif and to determine the position specific scoring matrix (PWM). This PWM was used with MotifLocator to identify instances of this motif in the S. Typhimurium SL1344 genome. A background file specific to SL1344 was generated using the associated script called CreateBackgroundModel. A stringent threshold value of 0.8 was used [231, 232].

Transcriptional reporter experiments

Transcriptional fusions to *lacZ* for the *ssaG* and *sseA* promoter palindrome analysis were generated using chemically synthesized double-stranded DNA (Genscript Corp). Promoter DNA was ligated into pIVET5n, then the plasmid was subsequently conjugated into SL1344 to generate single-copy transcriptional fusions integrated on the chromosome as described previously [139]. Luciferase reporter constructs for the *Sodalis glossinidius SG1292* and *ssaR* promoters were generated by PCR amplification of promoter regions from genomic DNA templates. The luciferase reporter construct for the P*ssaG* scrambled substitution was created by PCR product splicing via overlap extension using the existing *ssaG* promoter cloning primers and two additional internal primers containing the desired mutation sequence (DTM0061R, 5’CGC GAA AGC AAC GAT TAC TCC GGC GCA CG3’ and DTM0061.1F, 5’GAG TAA TCG TTG CTT TCG CGA TAC CGG ATG TTC ATT GCT TTC TA3’). This DNA was ligated into pCS26 and transformed into SL1344

to generate plasmid-based reporters. Overnight cultures of *Salmonella* were sub-cultured into LPM medium and grown with shaking for 7 h. Samples were removed hourly to measure β-galactosidase activity via a chemiluminescence-based assay as described previously [139] or luminescence directly from cultures (EnVision, Perkin-Elmer). Output was relative light units (RLU) normalized to OD600. Each experiment was performed in triplicate then averaged. Reporter activity from mutant and rewired promoters was normalized to that from wild type promoters.

Data deposition

All ChIP-on-chip data can be retrieved from the NCBI Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20192. Data files for viewing in Artemis (http://www.sanger.ac.uk/Software/Artemis/) are available upon request.

Acknowledgments

We thank Michael Lowden for technical assistance with the microarray experiments and Wynne Lock for preliminary bioinformatics analysis. We are grateful to Scot Wolfe and Xiangdong Meng for providing purified prey libraries and technical assistance with the bacterial one-hybrid screens.

Figure 2.1. COG analysis of 133 genes co-regulated with SPI-2. COG categories are indicated to the right of the figure and COG sub-categories are ordered in terms of decreasing gene representation (blue) within each category. Median and average fold change values for each sub-category are indicated in green and red, respectively.

Figure 2.1

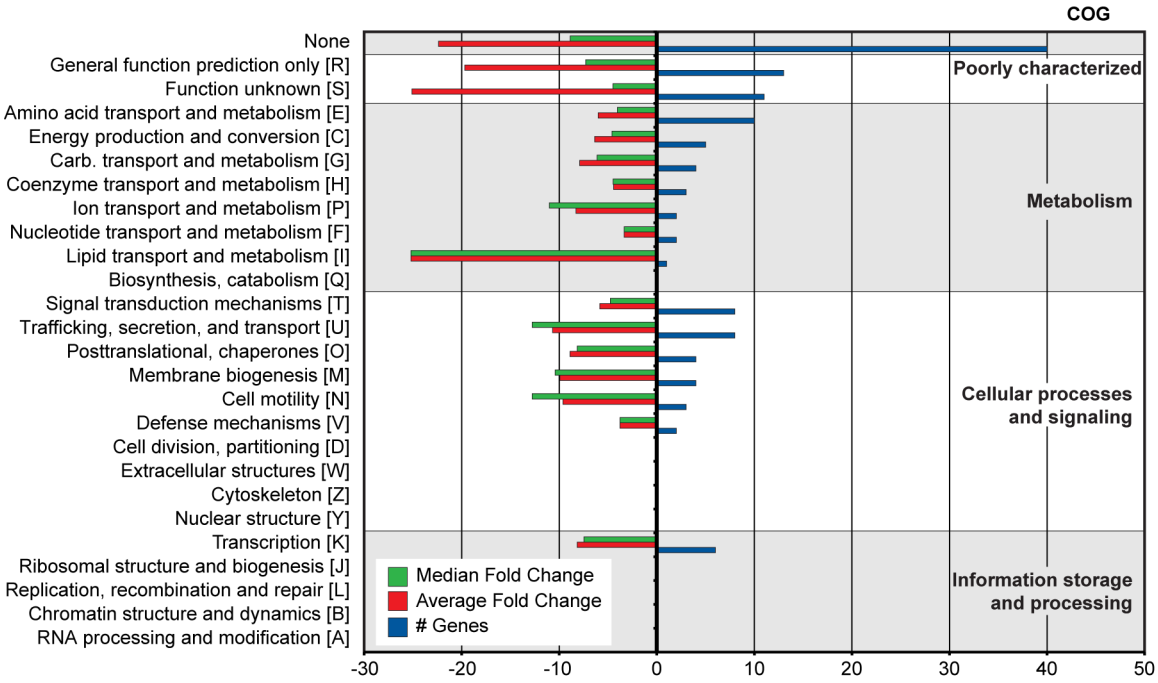

COG
None
General function prediction only [R]
Function unknown [S]
Amino acid transport and metabolism [E]
Energy production and conversion [C]
Carb. transport and metabolism [G]
Coenzyme transport and metabolism [H]
Ion transport and metabolism [P]
Nucleotide transport and metabolism [F]
Lipid transport and metabolism [I]
Biosynthesis, catabolism [Q]
Signal transduction mechanisms [T]
Trafficking, secretion, and transport [U]
Posttranslational, chaperones [O]
Membrane biogenesis [M]
Cell motility [N]
Defense mechanisms [V]
Cell division, partitioning [D]
Extracellular structures [W]
Cytoskeleton [Z]
Nuclear structure [Y]
Transcription [K]
Ribosomal structure and biogenesis [J]
Replication, recombination and repair [L]
Chromatin structure and dynamics [B]
RNA processing and modification [A]
Median Fold Change
Average Fold Change
Genes
Poorly characterized
Metabolism
Cellular processes and signaling
Information storage and processing
-30
-20
-10
0
10
20
30
40
50

Figure 2.2. Genome-wide functional genomics analyses for SsrB

(A) ChIP-on-chip and microarray analysis. Genomic islands are indicated and labelled on track 1 (outermost circle), where prophages are in green, pathogenicity islands are in red and regions of difference (ROD) with *S. bongori* are in blue. ChIP-on-chip interaction scores (log_2) for SsrB-activating conditions are plotted on track 2, where peaks with signals greater than one standard deviation from the mean probe signal are indicated in blue. Transcriptional profiling analysis of SsrB-regulated genes is plotted on track 3. Genes repressed more than 5-fold in an *ssrB* mutant are indicated in green. G+C content is plotted on track 4, where values greater than 53% (genome mean) are indicated in red and less than 53% are indicated in green. (B) Summary of gene expression data and SsrB binding data for pathogenicity islands, prophage islands, *S. bongori* ROD and genome sequence not in islands. The number of SsrB binding peaks identified by ChIP-on-chip for each of the individual islands is shown in the outermost track, along with the number of genes down-regulated in the transcriptional profiling analysis (in brackets). The total number of ChIP-on-chip interaction peaks and genes regulated by SsrB are shown in the innermost circles along with the percent of total (in brackets). The total genomic content of each island or conserved regions are shown below each circle, along with the percent of total genome.

Figure 2.2

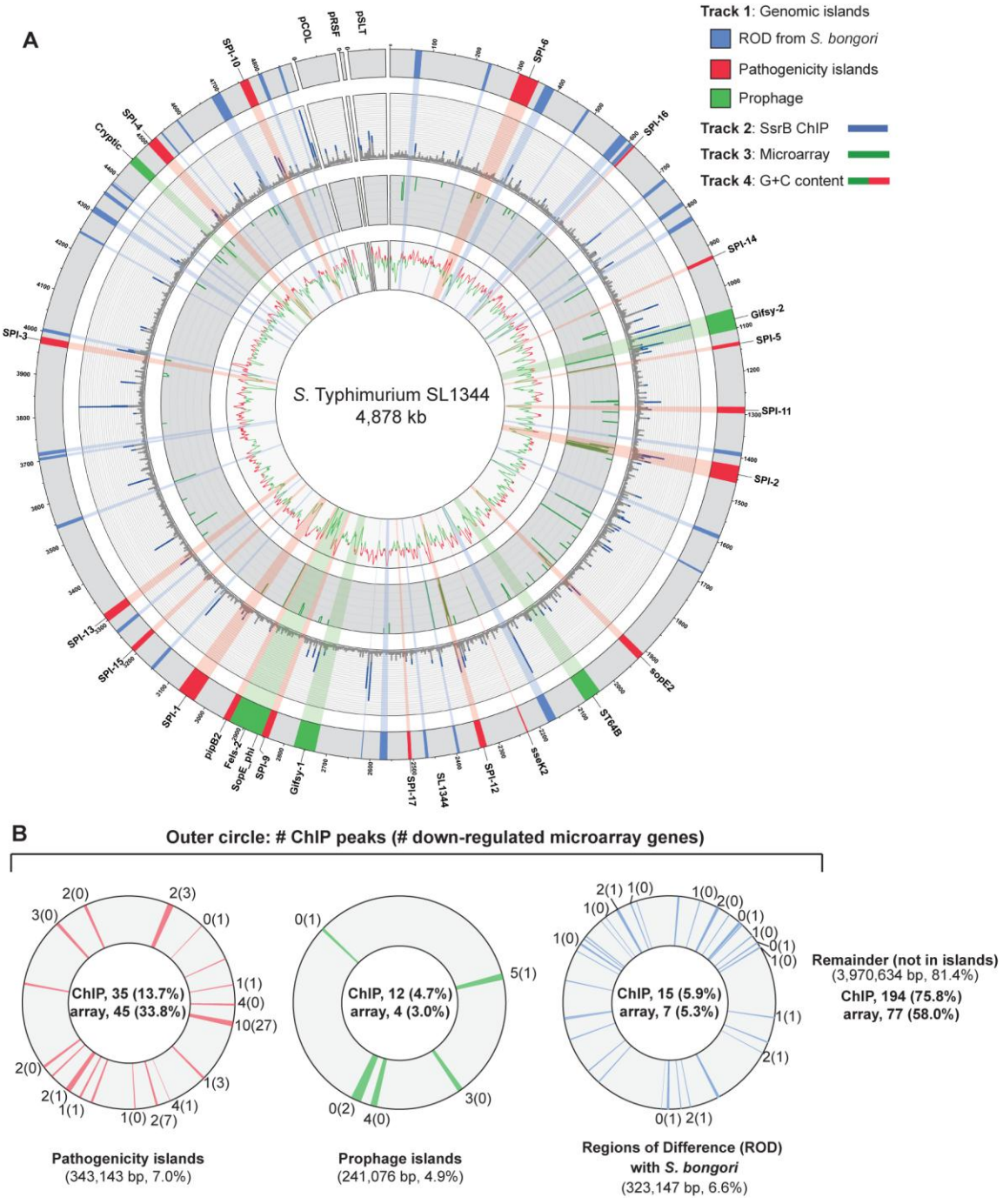

A
Track 1: Genomic islands
ROD from S. bongori
Pathogenicity islands
Prophage
Track 2: SsrB ChIP
Track 3: Microarray
Track 4: G+C content
S. Typhimurium SL1344
4,878 kb
B
Outer circle: # ChIP peaks (# down-regulated microarray genes)
ChIP, 35 (13.7%)
array, 45 (33.8%)
Pathogenicity islands
(343,143 bp, 7.0%)
ChIP, 12 (4.7%)
array, 4 (3.0%)
Prophage islands
(241,076 bp, 4.9%)
ChIP, 15 (5.9%)
array, 7 (5.3%)
Regions of Difference (ROD)
with S. bongori
(323,147 bp, 6.6%)
Remainder (not in islands)
(3,970,634 bp, 81.4%)
ChIP, 194 (75.8%)
array, 77 (58.0%)

Figure 2.3. The SsrB binding profile at the SPI-2 locus. Plotted are the SsrB ChIP interaction scores (black trace) for all probes within the SPI-2 locus. ChIP-on-chip experiments were performed with *S.* Typhimurium grown in SPI-2 inducing (LPM) and non-inducing (LB) conditions. ChIP score was calculated as LB (control/experimental) / LPM(control/experimental) where LB scores were used to normalize the data. Regions of SPI-2 footprinted by SsrB are highlighted with vertical pink bars. G+C content across the region is plotted as a green trace. Genes in the SPI-2 genomic island (top) are plotted to scale and are coloured according to function (see key).

Figure 2.3

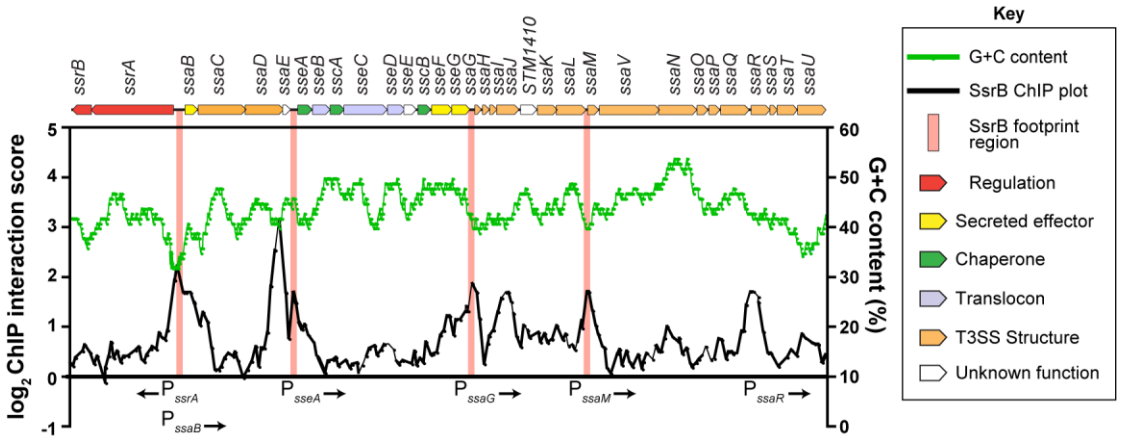
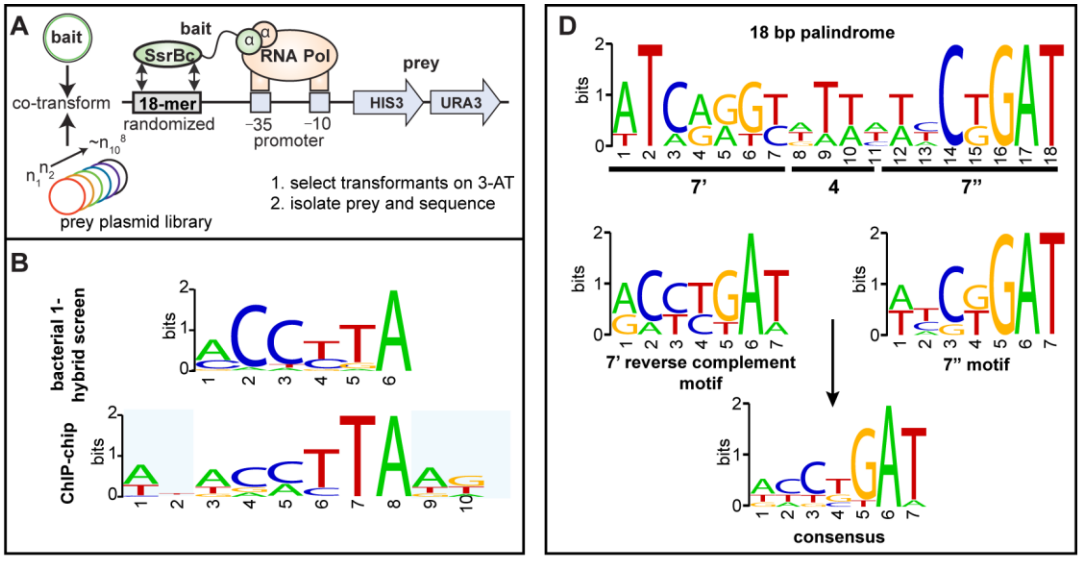

ssrB
ssrA
ssaB
ssaC
ssaD
ssaE
sseA
sseB
sscA
sseC
sseD
sseE
sscB
sseF
sseG
ssaG
ssaH
ssaI
ssaJ
STM1410
ssaK
ssaL
ssaM
ssaV
ssaN
ssaO
ssaP
ssaQ
ssaR
ssaS
ssaT
ssaU
log2 ChIP interaction score
G+C content (%)
P ssrA
P ssaB
P sseA
P ssaG
P ssaM
P ssaR
Key
G+C content
SsrB ChIP plot
SsrB footprint region
Regulation
Secreted effector
Chaperone
Translocon
T3SS Structure
Unknown function

Figure 2.4. Identification of a conserved palindrome in *S.* Typhimurium SPI-2 and *S. glossinidius* SSR-3 promoters. (A) Schematic representation of the bacterial one-hybrid screen for functional SsrB binding sequences. (B) Motif logos generated independently from the bacterial one-hybrid screen and ChIP-on-chip data show similar but degenerate consensus along 6 nucleotides. (C) Aligned promoter regions from the *S.* Typhimurium SPI-2 island and *S. glossinidius* SSR-3 island containing the putative SsrB DNA recognition palindrome. Coordinates indicate position of the conserved palindrome with respect to the translational start site of the downstream gene. The 7'-4-7" tail-to-tail palindrome sequences are underlined in blue (7' site) and red (7" site), aligned and highlighted in yellow. Asterisks indicate conserved nucleotides between *S.* Typhimurium and *S. glossinidius*. The positions of SsrB footprints protected from DNase1 [101, 102] are shown as black text highlighted with a blue horizontal bar. (D) Consensus motif logos for the palindrome sequences from the SPI-2 promoters shown in C.

Figure 2.4



C

Organism	Promoter	Coord.	Sequence alignment (7' 4 7")
S. Typhimurium	*ssrA*	-70..-87	ATTTATCTT-CCTAATGATAAC-ACCATCGATTAATCTTCTGATGAAACTATATGTACTGCGATAGTGATCAAGTGCCAAA
S. glossinidius	*SG1280*	-61..-78	ATACATAAAGCCTCGTGTCATCCATTATTCTATAGCCAACCGGT--AATTTTAAAAAATG-GACATAGCCTGAGTAGCCAT
S. Typhimurium	*ssaB*	-53..-36	AAGAGTATAGTCGAAATAAATGTG-AATCAGGCTTTTTACGGATGTGGTTGTGAGCGAATTTGATAGAAACTCCC--ATTT
S. glossinidius	*SG1281*	-72..-55	ATGTGTCCAATTCCACCCCATGGGTAATCAGGTATTTTACTGATGTACCCTGCTGGGTTTTCAACTGAGTATGACGGAGTT
S. Typhimurium	*sseA*	-65..-48	ACTGCCCAGGTTCAAAATGGCAAGTAAAATAGCCCTAATGGGATAGGCTCTTAGT----TAGCACGTTAATTATCTATCGT
S. glossinidius	*SG1285*	-123..-106	GGCGC-----TACAGGCTG------AAGATATTCTTAATGGCATTTACTTTTGTTATCACAATGCCTCATTGAGTTGCCGT
S. Typhimurium	*ssaG*	-88..-71	AACAACGTGCGCCGGAGTAATCGTT-TTCAGGTATATACCGGATGTTCATTGCTTTCTAAATTTTGCTATGTTGCCAGTAT
S. glossinidius	*SG1292*	-121..-104	GCTAACGT-CGCCGGGCCGGCTGGAATTCAGGTCTACACCGGATGCTGGCAGCGCTGCGG-TGCGGGGATAATCCGCATAT
S. Typhimurium	*ssaM*	-74..-57	CGAAACGCTTCGTGAAGTAAAGATAAATCAGGTTTTATTCTGATACCTGGCTTTCA-ATATTTAGGTAAATTGGCTTTCTG
S. glossinidius	*SG1299*	-74..-57	CGACGTATTGCGGCAGCTGGCGGCAAATCAGGTATAGCCCCGATACCCGCGTCCCGCACCATTTGGGACGCTGTCTGCACT

Figure 2.5. The SsrB binding motif architecture is flexible and conserved in *Sodalis*. (A) The sequences of the wild type SsrB binding motif and mutated permutations from the *ssaG* promoter are shown. The orientation of the 7-bp half sites are indicated, where the blue arrow represents the 7' site and the red arrow represents the 7" site as identified in Figure 4. The orientation of the 4-bp spacer sequence is represented as a black arrow. The colours of the bases are used to distinguish the different deoxyribonucleotides. (B) Transcriptional reporter data for the wild type *cis*-regulatory input (wt) and the seven permutations as defined in A. Promoter activity is shown at 5 h as measured by β-galactosidase assays normalized to promoter activity from the wild type reporter. (C) Comparison of *ssaG* promoter activity between wild type *S.* Typhimurium and an *ssrB* mutant. β-galactosidase assays were conducted for the P*ssaG* X-X-X reporter in both wt and Δ*ssrB* cells. Data represents the promoter activity as a percent of wild type. (D) The conserved palindrome sequence from *Sodalis glossinidius* is functional. Transcriptional reporters that either contained (7'-4-7'') or lacked (X-X-X) the 18-bp palindrome sequence were constructed from evolved orthologous promoters in *Salmonella* (P*ssaG*) and *Sodalis* (P*SG1292*). An additional reporter that mutated the bases in the left heptamer to those never occurring in the consensus matrix (scrambled) was also constructed. Reporters were transformed into wild type *Salmonella* and an *ssrB* mutant and tested for transcriptional activity. Luminescence data was normalized to the OD of the culture. Shown are the means with standard deviation for three separate experiments.

Figure 2.5

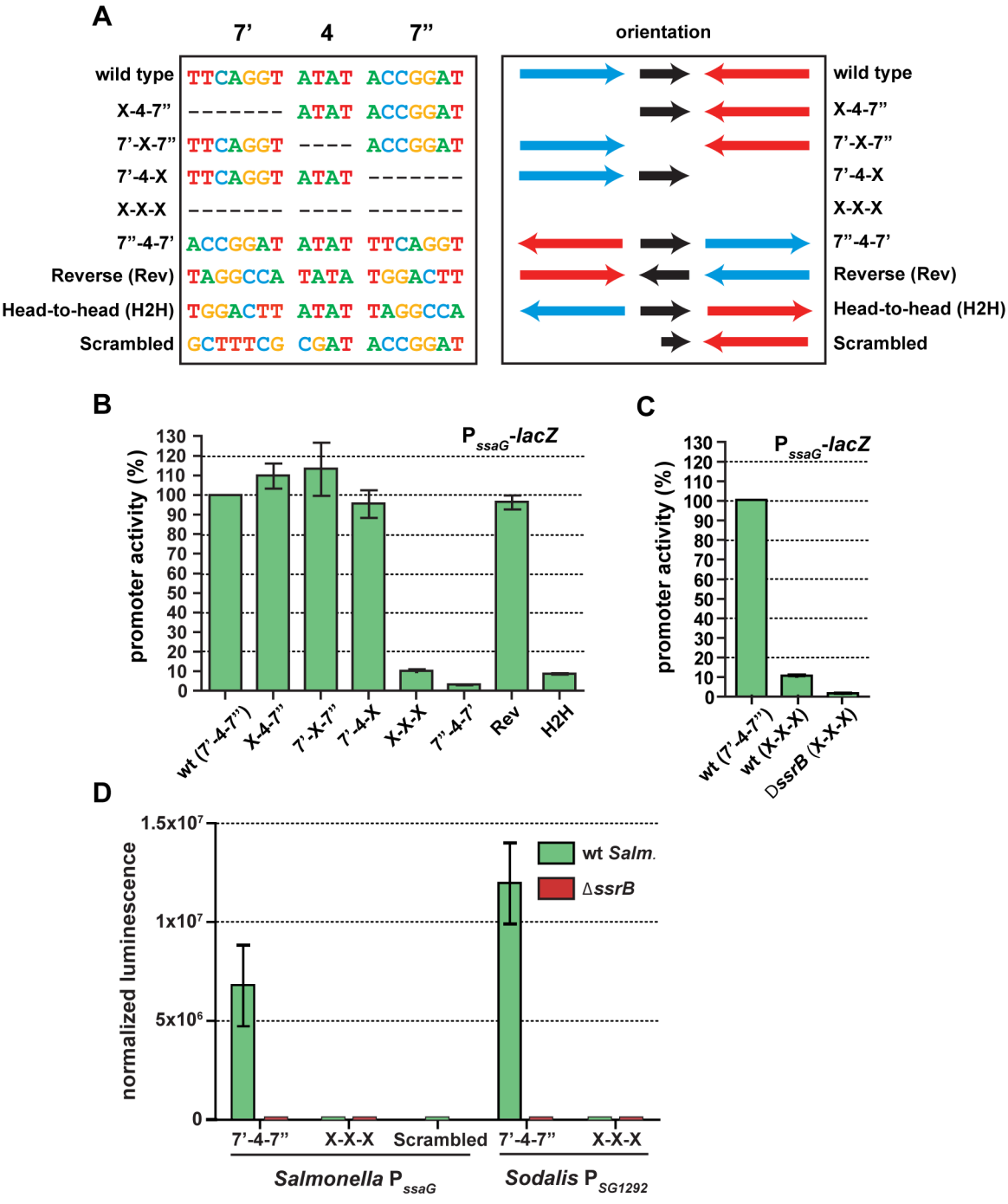

A
7'
4
7"
orientation
wild type TTCAGGT ATAT ACCGGAT
X-4-7" ------- ATAT ACCGGAT
7'-X-7" TTCAGGT ---- ACCGGAT
7'-4-X TTCAGGT ATAT -------
X-X-X ------- ---- -------
7"-4-7' ACCGGAT ATAT TTCAGGT
Reverse (Rev) TAGGCCA TATA TGGACTT
Head-to-head (H2H) TGGACTT ATAT TAGGCCA
Scrambled GCTTTCG CGAT ACCGGAT
wild type
X-4-7"
7'-X-7"
7'-4-X
X-X-X
7"-4-7'
Reverse (Rev)
Head-to-head (H2H)
Scrambled
B
P ssaG -lacZ
promoter activity (%)
wt (7'-4-7")
X-4-7"
7'-X-7"
7'-4-X
X-X-X
7"-4-7'
Rev
H2H
C
P ssaG -lacZ
promoter activity (%)
wt (7'-4-7")
wt (X-X-X)
ΔssrB (X-X-X)
D
normalized luminescence
wt Salm.
ΔssrB
7'-4-7"
X-X-X
Scrambled
7'-4-7"
X-X-X
Salmonella P ssaG
Sodalis P SG1292

Figure 2.6. Genome-wide identification of SsrB palindrome sequences. (A) A position weight matrix was generated from all naturally-evolved palindromes in SPI-2 and used to search the genome for similar sequences. Palindromes identified in regulatory DNA that co-occurred with ChIP peaks upstream of SsrB-regulated genes were selected, aligned and binned according to those scoring >0.8 against the position weight matrix (top box) and those scoring between 0.7-0.8 (lower box). The left (7') and right (7") heptamers and the 4-bp spacer are displayed as a heat-map to show bases of high conservation (dark blue) from degenerate regions (light blue/white). The genes controlled by these promoters are indicated to the left of the sequences and coloured according to function: structural components of the T3SS (orange), effectors (blue), regulatory elements (red), T3SS chaperones and translocon (green) and SsrB-regulated genes of unknown function (black). (B) The high-scoring palindrome in the *ssaR* IGR is functional. A single-copy *ssaR* reporter was integrated on the chromosome and tested for functional activity in wild type cells and in Δ*ssrB*. Promoter activity is shown as the mean with standard error from three separate experiments. (C) Functional validation of the intragenic high-scoring palindrome in the *ssaE* coding sequence. A single-copy transcriptional reporter that either contained (WT P*sseA*) or lacked (P*sseA* del) the single heptamer site in the *sseA* IGR was integrated on the chromosome in wild type cells, or mutants lacking either *ssrB* or the *ssaE* coding sequence that removed the high-scoring intragenic palindrome (P*sseA'*). Transcriptional activity at 6 h is shown as the mean of triplicate determinations with standard error.

Figure 2.6

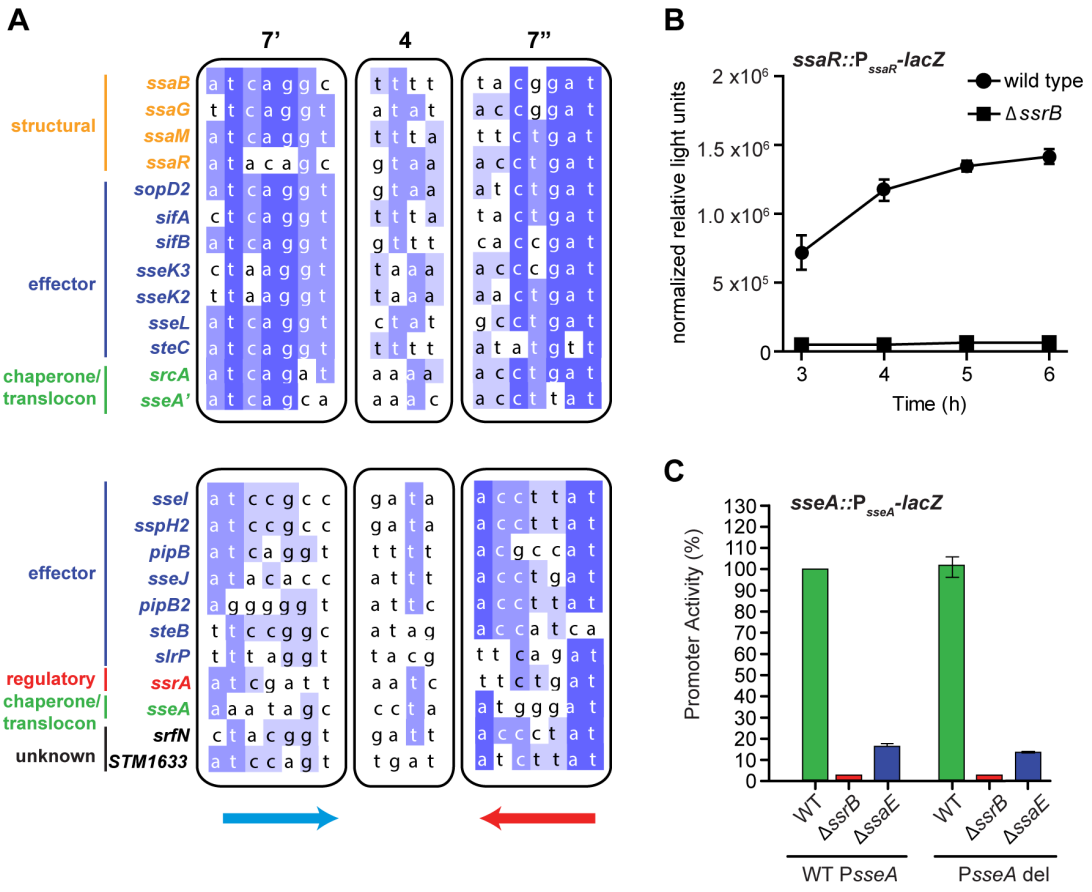

A
7'
4
7"
structural
ssaB
ssaG
ssaM
ssaR
effector
sopD2
sifA
sifB
sseK3
sseK2
sseL
steC
chaperone/ translocon
srcA
sseA'
effector
sseI
sspH2
pipB
sseJ
pipB2
steB
slrP
regulatory
ssrA
chaperone/ translocon
sseA
unknown
srfN
STM1633
B
ssaR::P ssaR -lacZ
wild type
ΔssrB
normalized relative light units
2 x10⁶
1.5 x10⁶
1.0 x10⁶
5 x10⁵
0
3
4
5
6
Time (h)
C
sseA::P sseA -lacZ
Promoter Activity (%)
130
120
110
100
90
80
70
60
50
40
30
20
10
0
WT
ΔssrB
ΔssaE
WT
ΔssrB
ΔssaE
WT PsseA
PsseA del

Table 2.1. Summary of microarray and ChIP-on-chip data

Microarray	
Total significant SsrB-activated genes, *n*	133
Genes with *S. bongori* ortholog, *n*	47
Mean fold change	-15.8 (-6.0*)
Median fold change	-6.5 (-4.5*)
Genes within pathogenicity islands, *n*	45
Genes with *S. bongori* ortholog, *n*	2
Mean fold change	-27.4 (-6.5*)
Median fold change	-13.1 (-6.5*)
Genes within phage islands, *n*	4
Genes with *S. bongori* ortholog, *n*	0
Mean fold change	-20.4 (n/a*)
Median fold change	-9.0 (n/a*)
Genes within *S. bongori* ROD, *n*	7
Genes with *S. bongori* ortholog, *n*	0
Mean fold change	-14.6 (n/a*)
Median fold change	-7.1 (n/a*)
Genes not in genomic islands, *n*	77 [a]
Genes with *S. bongori* ortholog, *n*	45
Mean fold change	-8.9 (-6.0*)
Median fold change	-4.4 (-4.4*)
ChIP-on-chip	
Probes, *n*	42,021
Probe size	60-bp
Average probe separation	57-bp
Genome probe average $\log_2$ signal	-0.02
Standard deviation	0.52
Peaks above 3 st. dev., *n*	256
Peaks within published islands, *n*	47
Peaks within *S. bongori* ROD, *n*	15
Within CDS, *n*	126
Within/Overlapping IGR, *n*	130

n, number

*, (value for genes with an ortholog in *S. bongori*)

[a], includes 21 genes on the microarray that could not be mapped because they were not annotated by the array manufacturer (TIGR).

Table 2.2. Predicted operons encoding SPI-2 associated genes

Gene Range	Operon Size	Fold Change	First Gene	LT2 Ortholog	Annotation	Probe Location[a]	ChIP Score[b]
SL2601-SL2604	4	-1.45	SL2604	STM2640:*rpoE*		CDS	4.12
SL0991-SL0991	1	-57.81	SL0991	STM1051:*sseI*	Effector	IGR	3.37
SL2256-SL2256	1	-43.50	SL2256	STM2287:*sseL*	Effector	IGR	3.02
SL2217-SL2217	1	-7.07	SL2217	STM2241:*sspH2*	Effector	CDS	2.95
SL1561-SL1561	1	-25.20	SL1561	STM1631:*sseJ*	Effector	CDS	2.82
SL1628-SL1628	1	-51.60	SL1628	STM1698:*steC*	Effector	5'-IGR-overlap	2.48
SL1574-SL1576	3	-3.41	SL1576	STM1646:*ydbH*		CDS	2.40
SL1325-SL1326	2	-9.64	SL1326	STM1392:*ssrA*	SPI-2 gene	IGR	2.31
SL1327-SL1330	4	-18.09	SL1327	STM1393:*ssaB*	SPI-2 gene	IGR	2.31
SL1161-SL1161	1	-17.23	SL1161	STM1224:*sifA*	Effector	IGR	2.26
SL1340-SL1340	1	-10.51	SL1340	STM1406:*ssaG*	SPI-2 gene	CDS	1.93
SL0909-SL0909	1	-15.47	SL0909	STM0972:*sopD2*	Effector	5'-IGR-overlap	1.87
SL1027-SL1027	1	-33.97	SL1027	STM1088:*pipB*	Effector	IGR	1.85
SL1331-SL1336	6	-49.46	SL1331	STM1397:*sseA*	SPI-2 gene	IGR	1.81
SL1353-SL1356	4	-12.90	SL1353	STM1419:*ssaR*	SPI-2 gene	5'-IGR-overlap	1.78
SL1347-SL1352	6	-16.59	SL1347	STM1413:*ssaM*	SPI-2 gene	5'-IGR-overlap	1.76
SL1563-SL1566	4	-3.10	SL1563	STM1633		5'-IGR-overlap	1.76
SL2763-SL2763	1	-195.60	SL2763	STM2780:*pipB2*	Effector	IGR	1.75
SL0083-SL0083	1	-8.71	SL0083	STM0082:*srfN*		IGR	1.54
SL1908-SL1910	3	-2.68	SL1908	STM1979:*fliP*		3'-IGR-overlap	1.35
SL2001-SL2011	11	0.66	SL2011	STM2035:*cbiA*		IGR	1.30
SL4242-SL4245	4	-0.98	SL4242	STM4305.S		IGR	1.28
SL0700-SL0700	1	-7.06	SL0700	STM0719		CDS	1.18
SL3259-SL3261	3	-1.29	SL3261	STM3288:*yhbC*		IGR	1.13
SL1992-SL2000	9	-1.57	SL2000	STM2024:*cbiL*		CDS	0.96
SL2114-SL2114	1	-19.35	SL2114	STM2138:*srcA*	Effector	CDS	0.94
SL1026-SL1026	1	-2.64	SL1026	STM1087:*pipA*	Effector	IGR	0.90
SL1159-SL1160	2	-5.29	SL1160	STM1223:*potC*		CDS	0.86
SL1341-SL1346	6	-42.72	SL1341	STM1407:*ssaH*	SPI-2 gene	3'-IGR-overlap	0.84
SL1785-SL1785	1	-11.81	SL1785	STM1856		IGR	0.81
SL0037-SL0038	2	-4.00	SL0037	STM0036		5'-IGR-overlap	0.78
SL1170-SL1171	2	-2.84	SL1171	STM1233:*ycfC*		5'-IGR-overlap	0.74
SL1337-SL1339	3	-8.51	SL1337	STM1403:*sscB*	SPI-2 gene	3'-IGR-overlap	0.72
SL1878-SL1878	1	-7.65	SL1878	STM1949:*yecF*		IGR	0.72
SL1705-SL1705	1	-6.08	SL1705	STM1777:*hemA*		5'-IGR-overlap	0.69
SL2193-SL2196	4	-2.43	SL2193	STM2216:*yejA*		CDS	0.68
SL4478-SL4479	2	-1.83	SL4478	STM4547:*yjjQ*		IGR	0.67
SL1763-SL1763	1	-21.12	SL1763	STM1834:*yebN*		IGR	0.62
SL0758-SL0760	3	-6.10	SL0758	STM0781:*modA*		5'-IGR-overlap	0.60

SL1200-SL1200	1	-10.18	SL1200	STM1265		CDS	0.58
SL1238-SL1242	5	-1.54	SL1238	STM1303:*argD*		CDS	0.56
SL1958-SL1961	4	-26.21	SL1961	-		IGR	0.56
SL2210-SL2211	2	-19.29	SL2211	STM2235		CDS	0.56
SL0701-SL0703	3	7.94	SL0701	STM0719		5'-IGR-overlap	0.55
SL0547-SL0549	3	-5.97	SL0549	STM0559:*rfbI*		IGR	0.55
SL0393-SL0393	1	-3.20	SL0393	STM0398:*phoR*		CDS	0.53
SL1689-SL1692	4	0.78	SL1692	STM1764:*narG*		IGR	0.53
SL0160-SL0160	1	-3.75	SL0160	STM0159		IGR	0.52
SL1373-SL1375	3	-1.29	SL1375	STM1443:*ydhI*		IGR	0.52

[a], CDS (within coding sequence); IGR (in intergenic region); 5'-IGR overlap (probe overlaps IGR and 5' region of CDS)
[b], $\log_2$

Figure 2.S1. *S.* Typhimurium SL1344 SPI-2 and *S. glossinidius str. 'morsitans'* SSR-3 motifs. Sequence logos of the half-site (7' and 7") and full-length (7'-4-7") palindrome motif identified within the promoter regions of SPI-2 in *S.* Typhimurium and SSR-3 T3SS in *S. glossinidius*. 'Combined' refers to the consensus sequence based on the single left and right heptamer sequences within a given organism. 'Merged' refers to the consensus sequence for individual left and right heptamers from both *Salmonella* and *Sodalis*.

Figure 2.S1

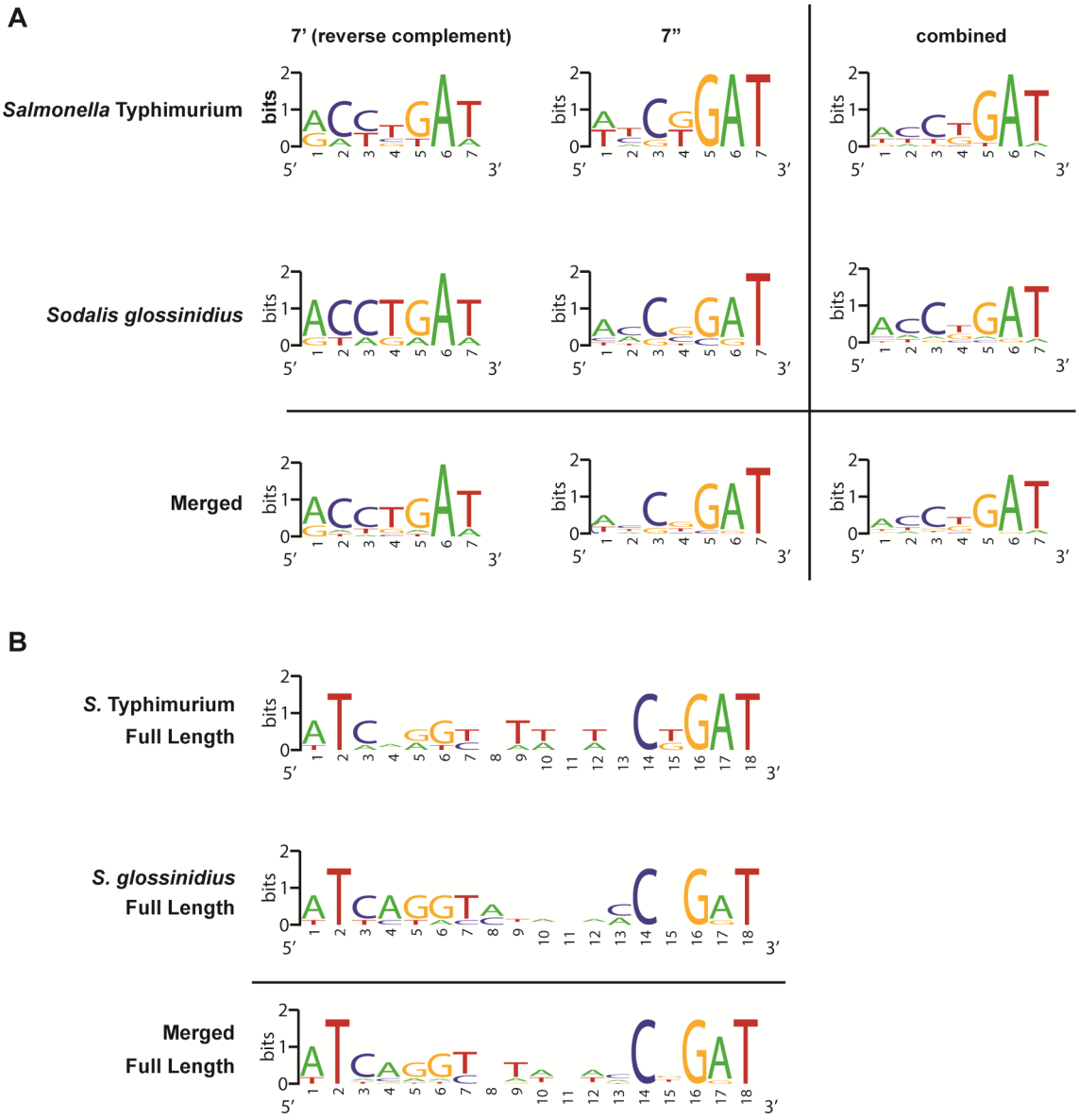
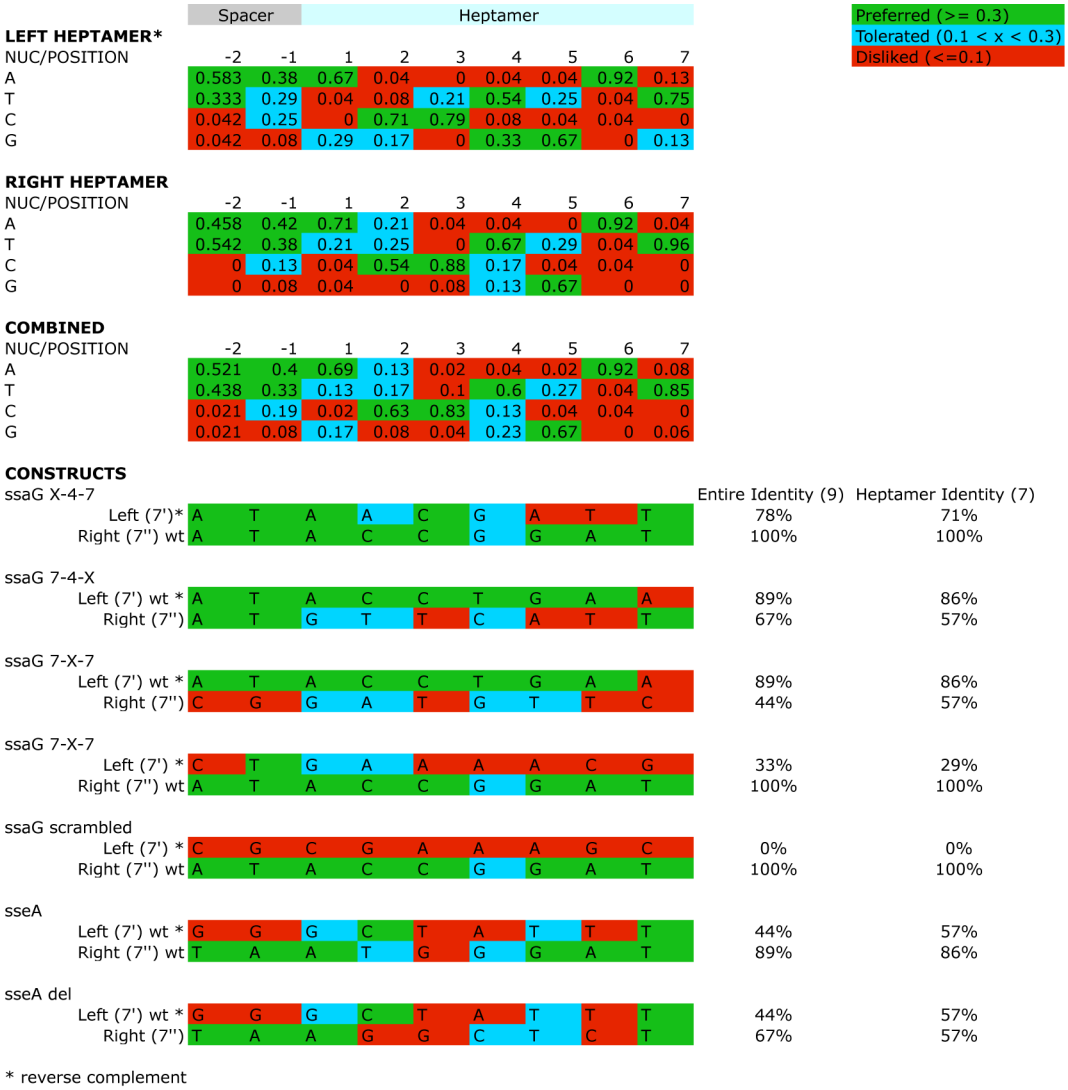

A
7' (reverse complement)
7"
combined
Salmonella Typhimurium
Sodalis glossinidius
Merged
B
S. Typhimurium Full Length
S. glossinidius Full Length
Merged Full Length
bits
5'
3'

Figure 2.S2. Frequency matrix analysis of SsrB palindromes. Presented is a frequency matrix of nucleotides occurring at each position of the 18-bp position in the 24 identified palindromes in S. Typhimurium SL1344. The 18-bp palindrome is shown as two 9-bp sequences with positions (-2) and (-1) referring to the spacer and positions 1-7 referring to the heptamer as indicated at the top of the figure. Matrices are shown for the left and right components of the palindrome in addition to a combined matrix.
Preferred, tolerated and disliked nucleotides are assigned based on frequency thresholds of 0.1 and 0.3 and are indicated by green, blue, and red respectively. Sequences for generated constructs are shown at the bottom of the figure and sequence identities with respect to the combined matrix are noted to the right of each construct sequence. The two possible permutations of the mutated palindrome are shown for the *ssaG* 7-X-7 construct.

Figure 2.S2



	Spacer		Heptamer						
LEFT HEPTAMER*									
NUC/POSITION	-2	-1	1	2	3	4	5	6	7
A	0.583	0.38	0.67	0.04	0	0.04	0.04	0.92	0.13
T	0.333	0.29	0.04	0.08	0.21	0.54	0.25	0.04	0.75
C	0.042	0.25	0	0.71	0.79	0.08	0.04	0.04	0
G	0.042	0.08	0.29	0.17	0	0.33	0.67	0	0.13
RIGHT HEPTAMER									
NUC/POSITION	-2	-1	1	2	3	4	5	6	7
A	0.458	0.42	0.71	0.21	0.04	0.04	0	0.92	0.04
T	0.542	0.38	0.21	0.25	0	0.67	0.29	0.04	0.96
C	0	0.13	0.04	0.54	0.88	0.17	0.04	0.04	0
G	0	0.08	0.04	0	0.08	0.13	0.67	0	0
COMBINED									
NUC/POSITION	-2	-1	1	2	3	4	5	6	7
A	0.521	0.4	0.69	0.13	0.02	0.04	0.02	0.92	0.08
T	0.438	0.33	0.13	0.17	0.1	0.6	0.27	0.04	0.85
C	0.021	0.19	0.02	0.63	0.83	0.13	0.04	0.04	0
G	0.021	0.08	0.17	0.08	0.04	0.23	0.67	0	0.06

Preferred (>= 0.3)
Tolerated (0.1 < x < 0.3)
Disliked (<=0.1)

CONSTRUCTS

		Entire Identity (9)	Heptamer Identity (7)
ssaG X-4-7			
Left (7')*	A T A A C G A T T	78%	71%
Right (7'') wt	A T A C C G G A T	100%	100%
ssaG 7-4-X			
Left (7') wt *	A T A C C T G A A	89%	86%
Right (7'')	A T G T T C A T T	67%	57%
ssaG 7-X-7			
Left (7') wt *	A T A C C T G A A	89%	86%
Right (7'')	C G G A T G T T C	44%	57%
ssaG 7-X-7			
Left (7') *	C T G A A A A C G	33%	29%
Right (7'') wt	A T A C C G G A T	100%	100%
ssaG scrambled			
Left (7') *	C G C G A A A G C	0%	0%
Right (7'') wt	A T A C C G G A T	100%	100%
sseA			
Left (7') wt *	G G G C T A T T T	44%	57%
Right (7'') wt	T A A T G G G A T	89%	86%
sseA del			
Left (7') wt *	G G G C T A T T T	44%	57%
Right (7'')	T A A G G C T C T	67%	57%

* reverse complement

Figure 2.S3. Alignment of *Salmonella* SsrB and the ortholog from *Sodalis glossinidius*. Full length protein sequences of SsrB from *S.* Typhimurium strain SL1344 (SL1325) and the *Sodalis* ortholog (SG1279) were aligned with ClustalW. The helices are labeled according to the structure of SsrBc as described by Carroll et al. 2009. Helices H2 and H3 make up the helix-turn-helix DNA binding motif (H168-K191) and the dimerization helix is H4. The conserved aspartic acid residue (D56) and other residues shown to be important for transcriptional activity by Carroll et al. 2009 are indicated in red [152].

Figure 2.S3

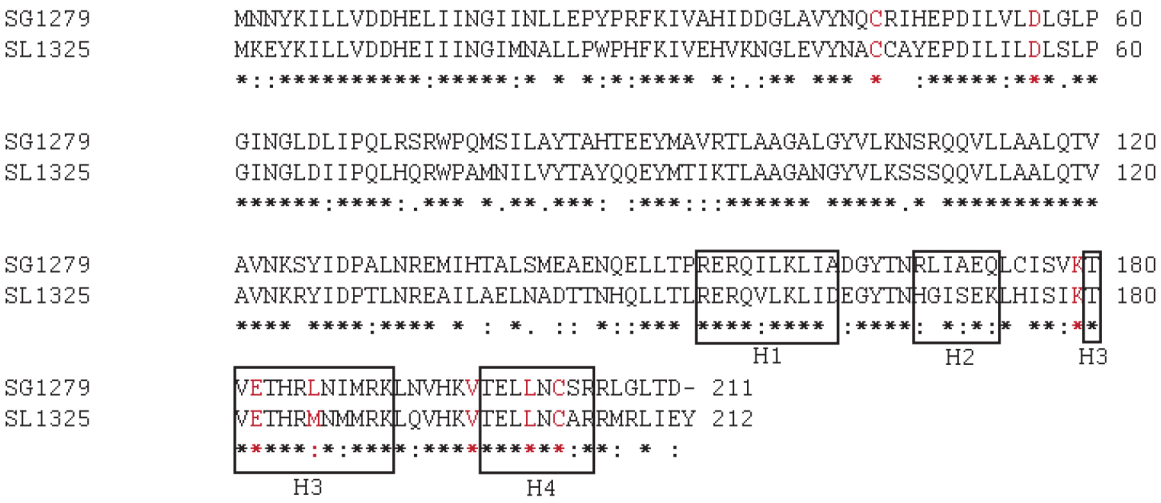

SG1279 MNNYKILLVDDHELIINGIINLLEPYPRFKIVAHIDDGLAVYNQCRIHEPDILVLDLGLP 60
SL1325 MKEYKILLVDDHEIIINGIMNALLPWPHFKIVEHVKNGLEVYNACCAYEPDILILDLSLP 60
SG1279 GINGLDLIPQLRSRWPQMSILAYTAHTEEYMAVRTLAAGALGYVLKNSRQQVLLAALQTV 120
SL1325 GINGLDIIPQLHQRWPAMNILVYTAYQQEYMTIKTLAAGANGYVLKSSSQQVLLAALQTV 120
SG1279 AVNKSYIDPALNREMIHTALSMEAENQELLTPRERQILKLIADGYTNRLIAEQLCISVKT 180
SL1325 AVNKRYIDPTLNREAILAELNADTTNHQLLTLRERQVLKLIDEGYTNHGISEKLHISIKT 180
H1
H2
H3
SG1279 VETHRLNIMRKLNVHKVTELLNCSRRLGLTD- 211
SL1325 VETHRMNMMRKLQVHKVTELLNCARRMRLIEY 212
H3
H4

Figure 2.S4. Episomal transcriptional reporter for the *ssaR* palindrome.

Luminescence production from a P*ssaR*-*lux* transcriptional fusion indicates an active, SsrB-dependent promoter at this location, in accord with SsrB binding to this location *in vivo*. Shown are data (mean with standard deviation) from triplicate determinations from three separate experiments. Luminescence was normalized to the optical density of the culture.

Figure 2.S4

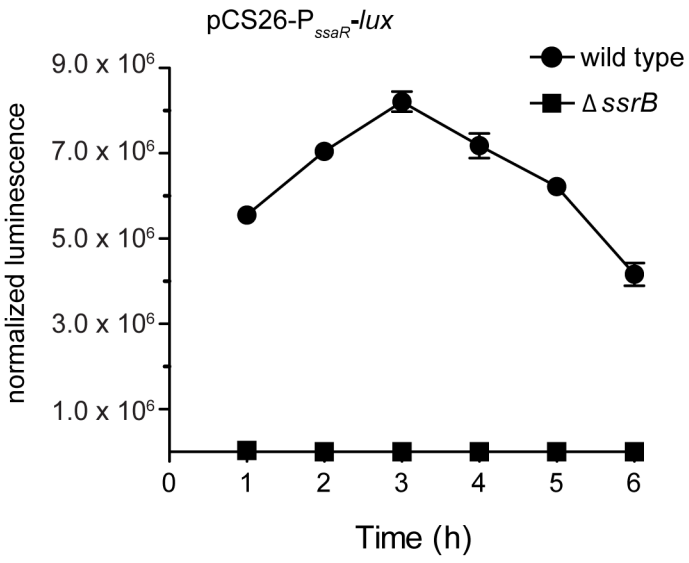

pCS26-P$_{ssaR}$-*lux*
wild type
Δ *ssrB*
normalized luminescence
9.0 x 10^6
7.0 x 10^6
5.0 x 10^6
3.0 x 10^6
1.0 x 10^6
0
1
2
3
4
5
6
Time (h)

Figure 2.S5. Mouse virulence data for the *ssrB-FLAG* SL1344 strain. Groups of C57BL/6 mice were infected by oral gavage with 10^6 colony forming units of wild type *S.* Typhimurium strain SL1344 or SL1344 containing an allelic replacement of *ssrB-FLAG*. Mice were monitored for endpoint and sacrificed when they had lost 20% of their initial body weight. The percent of mice surviving on each day after infection is shown. There is no statistical difference between the groups.

Figure 2.S5

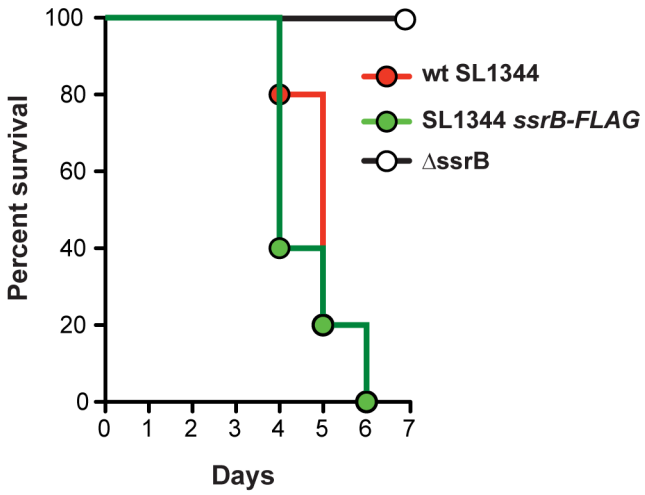

100
80
60
40
20
0
Percent survival
0 1 2 3 4 5 6 7
Days
wt SL1344
SL1344 ssrB-FLAG
ΔssrB

Table 2.S1 List of strains and plasmids used in this study

Strain or plasmid	Genotype or description	Reference
Strains		
DH5α	*supE44* Dl*acu169 (f80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Lab strain
DH5α λ*pir*	*supE44* Dl*acu169 (f80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1 λpir*	Lab strain
E. coli SM10 λ*pir*	*thi recA thr leu tonA lacY supE* RP4–2–Tc::*Mu λpir*	Lab strain
E. coli B1H	uso F' *pyrF hisB*; bacterial one-hybrid host strain	[226]
SL1344	Wild-type *S. enterica* sv. Typhimurium, SmR	[245]
SL1344 *ssrB*::*aphT*	Marked deletion of *ssrB*, KanR	[90]
SL1344 Δ*ssrB*	Unmarked, in-frame deletion of *ssrB*	This work
SL1344 Δ*ssaE*	Unmarked, in-frame deletion of *ssaE*	This work
SL1344 *ssrB::3FLAG-kan*	Encodes chromosomally FLAG-tagged *ssrB,* KanR	[143]
SL1344 P*ssaG*::pIVET-P*ssaG*	Merodiploid containing integrated P*ssaG-tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-X47	Merodiploid containing integrated P*ssaG*-X47-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-7X7	Merodiploid containing integrated P*ssaG*-7X7-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-74X	Merodiploid containing integrated P*ssaG*-74X-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-XXX	Merodiploid containing integrated P*ssaG*-XXX-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-H2H	Merodiploid containing integrated P*ssaG*-H2H-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-Rev	Merodiploid containing integrated P*ssaG*-Rev-*tnpR-lacZ* reporter	This work
SL1344 P*ssaG*::pIVET-P*ssaG*-7"-4-7'	Merodiploid containing integrated P*ssaG*-7"-4-7'-*tnpR-lacZ* reporter	This work
SL1344 P*sseA*::pIVET-P*sseA*	Merodiploid containing integrated P*sseA-tnpR-lacZ* reporter	This work
SL1344 P*sseA*::pIVET-P*sseA*del	Merodiploid containing integrated P*sseA*del-*tnpR-lacZ* reporter	This work
SL1344 Δ*ssrB* P*sseA*::pIVET-P*sseA*	*ssrB* mutant, merodiploid containing integrated P*sseA-tnpR-lacZ* reporter	This work
SL1344 Δ*ssrB* P*sseA*::pIVET-P*sseA*del	*ssrB* mutant, merodiploid containing integrated P*sseA*del-*tnpR-lacZ* reporter	This work
SL1344 Δ*ssaE* P*sseA*::pIVET-P*sseA*	*ssaE* mutant, merodiploid containing integrated P*sseA-tnpR-lacZ* reporter	This work
SL1344 Δ*ssaE* P*sseA*::pIVET-P*sseA*del	*ssaE* mutant, merodiploid containing integrated P*sseA-tnpR-lacZ*del reporter	This work
SL1344 P*ssaR*::pIVET-P*ssaR*	Merodiploid containing integrated P*ssaR-tnpR-lacZ* reporter	This work
Plasmids		
pB1H1	Bacterial one-hybrid bait plasmid, CmR	[226]
pB1H1-*ssrBc*	Bait plasmid carrying the C-terminal domain of *ssrB* (*ssrBc*), CmR	This work
pB1H1-*phoP*	Bait plasmid carrying *E. coli phoP*, CmR	This work
pH3U3	Bacterial one-hybrid prey plasmid, KanR	[226]
pIVET5n	*tnpR-lacZ*, *sacB*, R6K ori, *bla*, AmpR	[222]
pCS26	pSC101 *ori, luxCDABE,* KanR	[246]
pIVET-P*ssaG*	*ssaG* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-X47	*ssaG* promoter with "X-4-7" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-7X7	*ssaG* promoter with "7-X-7" mutation fused to *tnpR*-	This work

	lacZ in pIVET5n, AmpR	
pIVET-P*ssaG*-74X	*ssaG* promoter with "7-4-X" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-XXX	*ssaG* promoter with "X-X-X" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-H2H	*ssaG* promoter with "H2H" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-Rev	*ssaG* promoter with 'Rev" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaG*-7"-4-7'	*ssaG* promoter with "7"-4-7'" mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*sseA*	*sseA* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*sseA*del	*sseA* promoter with motif deletion mutation fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-P*ssaR*	*ssaR* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pCS26-P*ssaG*	*ssaG* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-P*ssaG*-XXX	*ssaG* promoter with “X-X-X” deletion fused to *luxCDABE* in pCS26, KanR	This work
pCS26-P*SG1292*	*Sodalis glossinidius SG1292* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-P*SG1292*-XXX	*Sodalis glossinidius SG1292* promoter with “X-X-X” deletion fused to *luxCDABE* in pCS26, KanR	This work
pCS26-P*ssaR*	*ssaR* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-PssaG mut(1-9)	Scrambled *ssaG* promoter fused to *luxCDABE* in pCS26, KanR	This work

CHAPTER THREE

The SPI-12 remnant phage island contributes to *in vivo* fitness of *Salmonella* Typhimurium

Author's Preface to Chapter Three

In this chapter I describe the investigation of bacteriophage genes as fitness factors for *Salmonella* Typhimurium in the context of the host environment. This work evaluated the individual fitness contributions of six phage genes from SPI-12 using *in vivo* and *in vitro* infection experiments, and identified how SsrB regulates this locus. It also identified STM2239, a Q-like antiterminator, as the transcriptional regulator of the *STM2237* phage gene operon through gene expression and chromatin immunoprecipitation studies. Notably, this study brings together evidence for the role of SPI-12 in *Salmonella* intra-host survival from published work over the last 20 years and provides an explanation for how this locus is transcriptionally regulated in the host environment. The overall implication of these findings is that the SsrB regulon includes genes outside of T3SS function to facilitate intracellular survival, and that genes of bacteriophage origin contribute to pathogen fitness in the host.

Note: All references for this work have been incorporated into the thesis reference list to avoid redundancy.

Chapter Three – Co-authorship Statement

Chapter Three consists of the following manuscript:

The SPI-12 remnant phage island contributes to *in vivo* fitness of *Salmonella* Typhimurium.

At the time of this thesis publication, this manuscript was intended for submission in October, 2012 to the journal *Infection and Immunity*.

The following individuals are intended to be co-authors on this manuscript: Ana M. Tomljenovic-Berube, Brandyn Henriksbo, Steffen Porwollik, Michael McClelland, and Brian K. Coombes

I, Ana M. Tomljenovic-Berube, conceived and conducted all experiments contained herein with the exception of the following contributions:

1) Brandyn Henriksbo conducted the STM2239 chromatin immunoprecipitation experiments
2) Steffen Porwollik from the Michael McClelland laboratory (San Diego) conducted the STM2239 microarray hybridization and statistical analysis.

Additionally, Dr. Brian Coombes provided comments on the manuscript prior to inclusion in this document.

The SPI-12 remnant phage island contributes to *in vivo* fitness of *Salmonella* Typhimurium

Ana M. Tomljenovic-Berube [1, 2], Brandyn Henriksbo [2], Steffen Porwollik [3], Michael McClelland [3, 4], and Brian K. Coombes [1, 2]

[1] Michael G. DeGroote Institute for Infectious Disease Research,

[2] Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario Canada L8S 4K1,

[3] Vaccine Research Institute of San Diego, San Diego, California, United States of America,

[4] Department of Pathology and Laboratory Medicine, University of California Irvine, Irvine, California, United States of America

* Correspondence and request for materials should be addressed to BKC.

Dr. Brian Coombes
Dept. Biochemistry & Biomedical Sciences
McMaster University
1280 Main Street West
Health Sciences Centre Room 4H17
Hamilton, Ontario Canada L8S 4K1
Phone: 905-525-9140 ext. 22159
E-mail: coombes@mcmaster.ca

Running title: *Role of SPI-12 in Salmonella fitness*

Abstract

Bacteriophages can mediate horizontal gene transfer in bacteria and often carry genes that are selective under certain environmental conditions. This study investigated the role of a remnant phage island encoded within *Salmonella* Pathogenicity Island 12 (SPI-12) of *Salmonella enterica* serovar Typhimurium in its contribution to bacterial virulence and revealed a unique regulatory mechanism at play. Both *in vivo* murine infection experiments as well as *in vitro Salmonella* replication assays determined that four phage genes in SPI-12 contribute to survival in the host. In particular, STM2239, a putative Q-antiterminator showed a prominent virulence contribution and functional role at this locus. Transcriptional reporter experiments, quantitative RT-PCR, and immunoblotting demonstrated that the virulence regulator SsrB as well as STM2239 contribute to transcriptional activation of phage genes in SPI-12. SsrB was found to regulate this locus indirectly by way of transcriptional read-through from the *sspH2* promoter, resulting in down-stream transcription of *STM2240* and *STM2239*. Chromatin immunoprecipitation (ChIP) determined that STM2239 interacts with the *STM2237* promoter, indicating that it may function as an antiterminator to directly activate its adjacent genes. These results have demonstrated that horizontally-acquired genes including those from bacteriophage may be adapted by pathogenic bacteria to improve fitness in the host and thus expand their virulence arsenal in specific niche environments.

Introduction

Horizontal gene transfer (HGT) is a prominent driver of evolution in bacterial pathogens, allowing rapid adaptation to new environments [1]. These integration events have shaped the genomic landscape of many organisms, and in some cases have contributed to the emergence of new species, such as the divergence of *Salmonella bongori* and *Salmonella enterica* (*S. enterica*) species through the acquisition of *Salmonella* Pathogenicity Island (SPI)-2 [73, 81, 161, 247, 248]. While there are several mechanisms by which HGT can occur, bacteriophage-mediated transduction has contributed significantly to the mosaic nature of bacterial genomes, introducing substantial diversity not only amongst different species, but within subspecies as well [249].

In *Salmonella enterica* serovar Typhimurium (*S.* Typhimurium), prophages have introduced a number of virulence factors that have contributed to the pathogenic nature of this organism. These genes, called lysogenic conversion genes, are integrated into new bacterial genomes through passive carriage by bacteriophages that accidentally package bacterial genes from one host during a lytic cycle, then introduce this new genetic content into a subsequent host when they undergo lysogeny [24]. *S.* Typhimurium has acquired an extensive arsenal of virulence genes in this manner, including *sopE* encoded within the SopEΦ prophage, *gogB* encoded within Gifsy-1, *sseI* encoded within Gifsy-2, as well as the superoxide dismutases *sodC1* and *sodC3* encoded within Gifsy-2 and Fels-1, respectively [26, 27, 29, 157, 207, 250]. Additionally, several other virulence genes are encoded within cryptic prophages. These genetic loci contain bacteriophage genes but

have lost subsets of their viral genome over time and thus are unable to form viable phage, but appear as remnants in the bacterial genome. Frequently, these remnant phage islands are associated with genes that confer a virulence or fitness advantage to the bacterial host as part of adaptation to specific environments [251].

The incorporation of horizontally-acquired genes into existing regulatory networks and bacterial processes has been the subject of much research, particularly in *Salmonella*. Transcriptional regulators involved in the activation of virulence factors have been extensively studied, with regulators such as PhoP and SsrB incorporating large gene sets into their regulatory networks [36, 155]. We recently identified the regulatory logic controlling SsrB-DNA interactions and showed that SsrB has evolved to regulate about 4% of the *Salmonella* genome [252]. SsrB is the response regulator in the SsrA-SsrB two-component regulatory system that is required for the expression of genes in SPI-2, a horizontally-acquired region in *S. enterica* species that encodes a type-III secretion system (T3SS) necessary for intracellular survival within host macrophages [62, 63]. SsrB also regulates the expression of almost all SPI-2-associated T3SS effector cargo proteins and other fitness factors encoded throughout the genome [102, 155]. Transcriptional profiling experiments looking for SsrB-regulated genes identified a cluster of bacteriophage genes encoded in a remnant phage island, SPI-12, that included a known SsrB-dependent SPI-2 T3SS-secreted effector SspH2 [94, 162, 252]. Previous work had shown that this island contributes to *S.* Typhimurium systemic survival in a murine infection model [169].

This study sought to investigate the mechanism of SsrB-dependent regulation of this locus and assess whether the remnant bacteriophage genes themselves contribute to *S.* Typhimurium pathogenesis. Our findings revealed a unique example of a regulatory cascade involving incorporation of bacteriophage genes into the regulon of a virulence transcriptional regulator. This was found to occur by way of transcriptional read-through from a virulence factor promoter, resulting in activation of a phage-encoded transcriptional regulator which mediated further activation of downstream phage genes in the remnant phage island at SPI-12.

Materials and Methods

Bacterial strains and growth conditions

Bacterial strains and plasmids used in this work are described in Table 3.S1. Strains were propagated in Luria-Bertani (LB) medium unless otherwise indicated. Liquid cultures were routinely grown at 37^{o}C, 225 rpm unless otherwise noted. Antibiotics were supplemented to media in the following concentrations: ampicillin (100 μg/mL), kanamycin (50 μg/mL), and streptomycin (50 μg/mL). Low-phosphate, low magnesium (LPM) medium [139] was used for microarray, RT-PCR, chromatin immunoprecipitation, β-galactosidase, and immunoblot experiments.

Cloning and mutant construction

Primers used for construction of plasmid inserts and mutant or tagged strains are summarized in Table 3.S2. All *S.* Typhimurium deletion mutant strains were generated using the lambda red recombination method as described by Datsenko and Wanner (2000)

[253], where pKD4 (KanR) was used as the template plasmid. Similarly, chromosomal -HA or -3FLAG tagged constructs were generated using pSUB315 (KanR) and pSUB11 (KanR) as template, respectively. Resistance cassettes were resolved using pCP20 where indicated in Table 3.S1.

For generation of β-galactosidase reporter plasmids, ~1,000 bp of intergenic sequence upstream of the indicated gene were cloned into pIVET5n using XhoI/MfeI restriction sites and confirmed by sequencing. Plasmids were then transformed into *E. coli* SM10 λpir and transferred to *S.* Typhimurium via conjugation. Merodiploids were selected for based on both ampicillin and streptomycin resistance. pWSK129 plasmids expressing -2HA tagged constructs were generated by amplification of ~600-800 bp upstream of STM2237 or STM2240 followed by each subsequent gene in the operon. For instance, pWSK129-STM2234-2HA contains the promoter for *STM2237*, and genes *STM2237*, *STM2236*, *STM2235*, and *STM2234*, with the -2HA tag fused to the 3' end of *STM2234*. Inserts were introduced using SalI/BglII, where the sequence encoding the -2HA tag was previously cloned into the vector backbone. For construction of the pWSK129-P*sspH2::STM2239* complementation vector, the promoter from *sspH2* was joined together with the coding sequence for *STM2239* using SOE PCR as described by Horton et al. [254]. pFLAG-STM2239 was generated by ligating the coding sequence of *STM2239* in-frame with the FLAG epitope sequence in pFLAG-CTC vector (Sigma) using XhoI/BglII restriction sites. Finally, for generation of the pCS26 luminescence reporter plasmids, ~600 bp of intergenic sequence upstream of the indicated genes were

cloned into pCS26 [246] using XhoI/BamHI restriction sites and confirmed by sequencing.

Chromatin immunoprecipitation

Chromatin immunoprecipitation (chIP) was conducted as described previously [7]. Modifications to this protocol are as follows: Wild type, SsrB-3FLAG and STM2239-3FLAG strains were grown in LPM, pH 5.8 to mid-log. Formaldehyde was added to 1% final concentration, then quenched with glycine (125 mM). Cells were pelleted then resuspended in a 1:1 mix of ChIP lysis buffer and 2X RIPA solution in a volume 1/10 the original culture volume. Suspension was incubated at 37^{o}C for 30 minutes to lyse bacteria, then aliquoted to 0.5 mL and subjected to sonication using a Misonix 4000 sonicator (Fisher Scientific) at 40% intensity for 10-15 second pulses to shear chromosomal DNA to a size range of 500-1000 bp. Lysates were clarified via centrifugation and supernatants were used as input for immunoprecipitation. Immunoprecipitation was conducted overnight at 4°C on a rolling shaker using prewashed α-FLAG-M2 magnetic bead resin (Sigma).

IP resin was subjected to multiple wash treatments (as described in [7]), then treated with RNase A (20 μg/mL) for 30 min at 37^{o}C and subjected to a final wash in TE buffer followed by resuspension in 100 μL ChIP elution buffer. Samples were incubated for 10 min at 65^{o}C, then Proteinase K was added (final concentration 0.8 mg/mL) and samples were incubated for 2 h at 42^{o}C, following with overnight incubation at 65^{o}C. The eluted IP DNA and input DNA were purified using phenol/chloroform extraction, precipitated with ethanol, and resuspended in nuclease-free water (Life Techonologies).

Immunoprecipitated DNA from wild type (untagged) and STM2239-3xFLAG tagged strains was purified and used as template for PCR. 1 μL of DNA was used as template in a 50 μL PCR reaction using Platinum Taq polymerase (Life Techonologies) as directed by the manufacturer. PCR was allowed to proceed for 30 cycles and was resolved via electrophoresis on 1% agarose gel.

β-galactosidase assays

β-galactosidase activity was assayed as described previously [30]. Briefly, washed *Salmonella* cultures were inoculated 1:50 from overnight cultures into LPM, pH 5.8 liquid medium and grown with shaking for 7 hours. Hourly samples were taken to measure β-galactosidase activity using a chemiluminescence-based assay (Applied Biosystems). Luminescence was measured on an EnVision® multi-label reader (Perkin-Elmer) and values were determined based on relative light units (RLU) normalized to OD_{600}. Each experiment was performed in triplicate, then averaged.

Immunoblotting

Bacterial cultures grown overnight were washed then sub-cultured 1:50 into LPM, pH 5.8 and grown to OD_{600}~0.6. Cells were pelleted, then resuspended in SDS-PAGE sample buffer (100 mm Tris-HCl (pH 6.8), 20% (v/v) glycerol, 4% (w/v) SDS, 0.002% (w/v) bromophenol blue, and 200 mm DTT) with volume normalized according to optical density as described previously [139]. Resuspended samples were heat-treated at 100^{o}C for 10 minutes, then 10 μL of each was loaded onto 15% SDS-PAGE gels and subjected to electrophoresis. Gels were transferred to PVDF membrane, then probed with the following antibodies: mouse anti-HA (1:1,000 or 1:2,000, Covance), mouse anti-DnaK

(1:5,000 or 1:10,000, Stressgen), rabbit anti-SseC (1:10,000, gift from M. Hensel), goat anti-mouse:HRP (1:5,000, Sigma), and goat anti-rabbit:HRP (1:5,000, Sigma). Conjugated HRP was detected using the Western Lightning Plus-ECL chemiluminescence reagent (PerkinElmer) as per manufacturer's directions.

Gentamicin protection assays

RAW264.7 macrophage cell line was propagated at 37^{o}C, 5% CO_2 in DMEM supplemented with 10% (v/v) FBS (Gibco) throughout all experiments. Bacterial cultures of strains of interest were grown overnight in LB, then centrifuged and resuspended in opsonisation medium (DMEM, 20% human serum) and incubated at 37^{o}C for 30 minutes to opsonise bacteria. RAW264.7 cells were seeded to 2 x 10^5 cells/well in 24-well tissue culture plates, then infected with a 1:100 dilution of opsonised bacteria to a multiplicity of infection of ~100. Infected cultures were incubated for 30 minutes to allow for uptake, then wells were washed three times with PBS, followed by addition of DMEM supplemented with 100 μg/ml gentamicin (Bioshop). Cultures were incubated in gentamicin medium for 90 minutes, then washed twice with PBS and either lysed or cultured for an additional 18 hours in medium supplemented with 10 μg/ml gentamicin. At 2 and 20 hours macrophages were lysed in 250 μL of lysis buffer (1% Triton X-100, 0.1% SDS), and lysates were serial diluted and plated to determine bacterial CFU counts. Fold replication was determined as CFU (20 h) / CFU (2 h), and was normalized to fold replication of wild type for each independent experiment. All experiments were repeated 5 times unless otherwise noted. Statistical analysis was performed using a student paired *t* test.

Competitive infection of animals

All animal protocols were performed in accordance with the Canadian Council on the Use of Laboratory Animals and approved by the McMaster University Animal Ethics Committee. Competitive infection experiments were conducted as follows: ~2 x 10^6 *Salmonella* containing a 1:1 mix ratio of chloramphenicol-resistant wild type and unmarked mutant were orally administered to female C57BL/6 mice (Charles River). At minimum, 5 mice per strain were used. At 3 days post-infection, mice were sacrificed by cervical dislocation and spleen, liver, and caecum were harvested, homogenized, serial diluted, and plated on selective media (LB streptomycin – 50 or 100 μg/mL) to determine total CFU, then replica-plated onto LB containing streptomycin (50 μg/mL) and chloramphenicol (10 μg/mL) to determine the ratio of chloramphenicol-resistant to -sensitive colonies. Competitive index was calculated as (mutant/wild type)$_{output}$/(mutant/wild type) $_{input}$. Statistical analysis was performed using a student *t* test.

RNA experiments

For isolation of bacterial RNA, *Salmonella* strains were grown in either LB or LPM, pH 5.8 liquid culture to mid-log phase, OD_{600}~0.5. 1 mL of 95% ethanol/5% acidic phenol (pH 4.3) (v/v) solution (Sigma) was added per 5 mL of culture and this mixture was incubated at 4^{o}C for 20 minutes to stabilize RNA. Bacteria were isolated by centrifugation, then resuspended in 200 μL 10 mM Tris, pH 8.0 with lysozyme (1 mg/mL) and incubated for 10 minutes at 37^{o}C. RNA was isolated using the Roche High Pure RNA Isolation Kit (Roche) as per manufacturer's instructions for bacterial RNA,

including on-column DNase I treatment. Purified RNA was subjected to two additional rounds of DNase I treatment using the TURBO DNA-Free kit (Ambion).

For qRT-PCR, RNA was treated with DNase Inactivation Reagent (Ambion), then assessed for DNA contamination using PCR prior to reverse transcription (RT). Total RNA was equalized to identical concentrations prior to RT to normalize between isolates. RT was conducted using Superscript III reverse transcriptase as per manufacturer's instructions using random hexamers as primers (Life Techonologies). First-strand cDNA was diluted 1:10 prior to qRT-PCR. qRT-PCR experiments were conducted using the Roche LightCycler® 480 (Roche) in 96-well format with Roche SYBR®-Green I Master reagents (Roche). Normalized ratio was calculated via the pFAFFL equation method, using the *rsmC* transcript as the internal control across all strains. Each experiment was conducted in triplicate.

For 5' Random Amplification of cDNA Ends (RACE) analysis, RNA was isolated from wild type *S.* Typhimurium carrying an IPTG-inducible pFLAG-STM2239 vector as described above. This strain was grown in LB supplemented with 1 mM IPTG in order to over-express STM2239 while maintaining environmental conditions suitable for purification of stable RNA necessary for 5' transcript mapping. First-strand cDNA template was generated using primer GSP1 P2237, then purified using a Qiaquick PCR purification column (Qiagen). Poly-C tailing and cDNA amplification was achieved using the 5' RACE System for Rapid Amplification of cDNA Ends, Version 2.0 from Invitrogen (Life Technologies). PCR products from primer pair GSP3 P2237-AUAP were

purified and analyzed via sequencing to determine transcriptional start site. Experiment was conducted in duplicate to ensure reproducibility.

For microarray analysis meant to identify putative STM2239-dependent genes, RNA was isolated as described above with the following adjustments. *Salmonella* strains carrying either pFLAG-CTC (empty vector control) or pFLAG-STM2239 were grown in 10 mL LB ampicillin for 1.5 hours at 37^{o}C, 225 rpm, then induced with IPTG (1 mM) for 1.5 hours to induce expression of STM2239. RNA was purified and DNase-treated as described, then pooled and concentrated following TURBO DNase treatment on an RNeasy RNA purification column (Qiagen) to achieve concentrations suitable for microarray analysis (~1 μg/ μL). RNA was then used as template to generate Cy3 and Cy5-labelled first-strand cDNA using Superscript II (Life Technologies) and Cy3 or Cy5 dye-labelled dCTP. 20 μg total RNA was pre-incubated with 2.4 μg random hexamer primers (Life Technologies) at 70^{o}C for 10 minutes, then placed on ice for 2 minutes. The RNA mix was then combined with the Superscript II mastermix which was composed of the following: 4μL dye labeled dCTP (Cy3/Cy5), 6 μL 0.1M DTT, 12 μL 1st Strand Buffer (Life Technologies),1.2 μL nucleotide mix (25 mM dATP, dTTP, dGTP / 10mM dCTP), 4 μL Superscript II reverse transcriptase (Life Technologies), 2 μL Rnasin (Roche), and 0.8 μL nuclease-free water. The reaction was incubated at 42^{o}C for 1 h, followed by addition of 2 μL more of Superscript II and a second incubation at 42^{o}C for 1 h. 3 μL of fresh 1 M NaOH was added, then incubated at 70^{o}C for 10 minutes, followed by reaction neutralization with 3 μL of 1 M HCl. cDNA was purified using the Qiagen PCR Purification Kit according to manufacturer's directions (Qiagen). Labelled cDNA

were hybridized to *S*. Typhimurium 14028 high-density oligo arrays (Nimblegen) by the McClelland Laboratory. Experiments were performed in triplicate with dye-swapping. Output data was analyzed using WebarrayDB and LIMMA software packages [255, 256].

Results

Bacteriophage genes within SPI-12 contribute to *in vivo* fitness of *Salmonella* Typhimurium

Using competitive infection experiments in mice, previous work identified SPI-12 as a pathogenicity island, where an *STM2230-STM2245* deletion mutant was found to be attenuated for virulence [169]. We later identified SPI-12 in microarray and ChIP-on-chip genomic studies to identify SsrB-regulated genes in *Salmonella* [252]. We sought to further define the contribution of the bacteriophage genes encoded in this island for the fitness advantage associated with the presence of SPI-12. A multi-gene mutant , Δ*STM2234-STM2240*, was generated and assessed for its ability to compete with wild type cells during systemic infection in mice. The competitive index (CI) in the spleen, liver, and caecum at 3 days post-infection was 0.38 ± 0.27 ($p < 0.001$), 0.46 ± 0.35 ($p < 0.05$), and 0.58 ± 0.32, respectively, indicating that the mutant strain was attenuated for systemic dissemination compared to wild type (Figure 3.1A). These results were in agreement with those of Haneda et al. (2009), indicating that one or more virulence factor(s) may be encoded among the bacteriophage genes in SPI-12 [169].

To examine this further, we made single gene mutants for *STM2234, STM2235, STM2236, STM2237, STM2239*, and *STM2240* and tested these mutants individually against wild type cells in competitive infections of mice. Of the six mutants tested, four (Δ*STM2234*, Δ*STM2235*, Δ*STM2236*, and Δ*STM2239*) showed a fitness defect (Figure 3.1B). The *STM2239* mutant had a significant defect in systemic sites compared to wild type, with a CI of 0.10 ± 0.097 ($p < 0.0001$) and 0.120 ± 0.126 ($p < 0.0001$) in the spleen

and liver, respectively. These mutants exhibited no general growth defects when grown *in vitro* under various growth conditions (Figure 3.S1).

In vitro gentamicin protection assays were also conducted with each of the mutant strains to test their ability to replicate in tissue culture cells. Of all seven mutant constructs tested, the Δ*STM2234-STM2240* and Δ*STM2239* strains were unable to replicate like wild type cells by 20 hours post-infection (Figure 3.1C). This replication defect was able to be complemented back to wild type levels by providing STM2239 *in trans* from a plasmid (Figure 3.1D). These results indicate that four bacteriophage genes from the SPI-12 remnant phage island contributed to *S.* Typhimurium fitness in a host, and that Δ*STM2239* showed the most profound fitness virulence defect both systemically *in vivo* and during intracellular replication *in vitro.*

SsrB influences transcriptional regulation of SPI-12 bacteriophage genes

Having established a virulence contribution for phage genes within SPI-12, we next sought to determine how this locus was being regulated. Our previous work showed that gene expression in SPI-12 was decreased following deletion of the SPI-2-encoded transcription factor, SsrB [252]. In order to assess whether SsrB was directly involved in transcriptional regulation of this region, chromosomal β-galactosidase transcriptional reporters were generated for three intergenic regions (IGR) in SPI-12 (designated I, II, and III) (Figure 3.2A). In these experiments, transcriptional reporters for IGR-1 and IGR-3 showed decreased signal in an SsrB mutant compared to wild type, whereas the IGR-2 region showed similar activity in wild type and Δ*ssrB* cells (Figure 3.2B).

To confirm that SsrB influenced the transcript levels of the phage genes within SPI-12, we conducted quantitative real time RT-PCR experiments comparing RNA from wild type versus Δ*ssrB* cells grown in SPI-2 inducing conditions. The ribosomal RNA methyltransferase gene, *rsmC*, was used as an internal control across both strains, and *sspH2* was used as a positive control as it is known to be directly regulated by SsrB [140, 162, 257]. All six phage genes tested showed decreased transcript levels comparable to the *sspH2* control, while mRNA levels for *STM2238* were unaffected by the loss of *ssrB* (Figure 3.2C).

To determine protein levels for these genes, we created chromosomal hemagglutinin (HA)-fusions for each of the six phage genes. Of the six phage proteins assessed, only four were found to be expressed under these conditions (STM2234, STM2235, STM2236, and STM2239). Expression of each of these four phage proteins was found to be diminished in an *ssrB* mutant compared to wild type, in agreement with the reporter and RT-PCR data (Figure 3.2D). These findings together indicate that SsrB plays a role in regulating the expression of STM2234, STM2235, STM2236, and STM2239.

SsrB regulates SPI-12 phage gene expression by way of transcriptional read-through from the *sspH2* promoter

Having established that SsrB influences transcriptional regulation of the phage genes within SPI-12, we wanted to determine whether SsrB acted directly or indirectly on this locus. Our previous work mapping SsrB binding sites using ChIP-on-chip

experiments revealed that despite transcriptional and translational data supporting a direct role for SsrB-dependent regulation at the *STM2237* promoter, there was no SsrB binding site directly upstream of STM2237 (Figure 3.3A, [252]). There was, however, a prominent interaction peak found at the *sspH2* promoter, implying possible transcriptional read-through from this promoter to *STM2240* and *STM2239*. To investigate this possibility, cDNA generated from RNA isolated from wild type, Δ*ssrB* and Δ*STM2239* strains was used as template for a PCR experiment designed to amplify the IGR that resides between *STM2240* and *sspH2*. Primers were designed to amplify from within *sspH2* 21 bp from the 3' end of the coding sequence into the first 7 bp of the 5' coding sequence of *STM2240*. PCR products were observed for the wild type and Δ*STM2239* cDNA but there was no discernable product observed from the Δ*ssrB* cDNA (Figure 3.3B). This data demonstrates that a transcript spanning between *sspH2* and *STM2240* was produced, and that SsrB was necessary for production of this transcript. Additionally, we evaluated the expression of STM2239 in an *sspH2* promoter deletion mutant strain (ΔP_{sspH2}). Deletion of P_{sspH2} resulted in loss of STM2239 expression (Figure 3.3C). These data together indicate that STM2240 and STM2239 are likely regulated by SsrB through transcriptional read-through from the *sspH2* promoter.

STM2239 is a transcriptional regulator of the *STM2234-STM2237* operon

Having established a role for STM2234, STM2235, STM2236, and STM2239 in *S.* Typhimurium pathogenesis as well as a dependence on SsrB for transcriptional activity, we next explored possible functions for these proteins. BLAST analysis showed

that these genes are most similar to other bacteriophage genes typically encoded in the late operon of lysogenic phage, summarized in Table 3.1. Notably, STM2239 is similar to Q-like antiterminators of lambdoid phages. Q antiterminator proteins facilitate transcriptional activation of the late operon in temperate phage during the final stages of bacteriophage assembly after activation of the lytic cycle [258]. We investigated this potential function for STM2239 in the context of the phage genes encoded within SPI-12. Given that *STM2234-STM2237* were most similar to genes encoded in the late promoter of lambdoid phages, we evaluated whether STM2239 directly influenced transcriptional activation at the *STM2237* promoter. $P_{STM2237}$ β-galactosidase transcriptional reporter activity was assessed in an *STM2239* mutant background which showed a loss of transcriptional activity compared to wild type. This level of decrease was similar to that seen in the *ssrB* mutant (Figure 3.4A, Figure 3.2B-I). Evaluation of transcript levels of SPI-12 genes between wild type and the *STM2239* mutant showed a decrease for the genes in the *STM2237* operon, but no difference in the other genes tested (Figure 3.4B). Moreover, phage protein levels were diminished in an *STM2239* mutant, similar to that seen in the *ssrB* mutant (Figure 3.4C). These data indicate that STM2239 is a transcriptional regulator of the *STM2237* operon.

STM2239 directly interacts with the *STM2237* promoter

In order to establish whether STM2239 functions as a Q antiterminator protein, chromatin immunoprecipitation (ChIP) was performed to determine whether STM2239 physically interacts with the *STM2237* promoter. PCR amplification of part of the

STM2237 promoter using STM2239-bound ChIP DNA demonstrated that STM2239 was able to specifically bind to the *STM2237* promoter when *S.* Typhimurium was grown under inducing conditions (Figure 3.5A). Mapping of the transcriptional start site of *STM2237* using 5' random amplification of cDNA ends (5' RACE) indicated that the transcriptional start site resided 198 bp upstream of the translational start (Figure 3.5B). Examination of the sequence surrounding this +1 site identified putative -10 and -35 σ^{70} elements as defined by Hawley and McClure (1983), though no definitive Q utilization (*qut*) site was found [259]. These data indicate that STM2239 functions as a DNA-binding protein that influences transcriptional activation and expression of the genes within the *STM2237* operon.

STM2239 regulates only the genes within SPI-12

Given the prominent virulence defect observed in the *STM2239* mutant in the context of host infection, we hypothesized that STM2239 may play a larger role in regulating additional genes in the *S.* Typhimurium genome. In order to address this possibility, we conducted a microarray experiment evaluating transcript levels of Δ*STM2239* strains grown in LB over-expressing STM2239 from an IPTG-inducible plasmid. A strain carrying an empty vector in the presence of IPTG was used as a control. The results from this study indicated that the genes most influenced by overexpression of STM2239 were indeed those within SPI-12, followed by a subset of genes involved in various metabolic pathways (Table 3.2). Four promoters outside of the SPI-12 locus, P*ackA*, P*kdgT*, P*metK*, and P*marR* were selected for additional luminescence-based

transcriptional reporter experiments to assess whether STM2239 influenced their transcriptional activation. Reporters were assessed over a time course in wild type and *STM2239* mutant backgrounds under SPI-2 inducing conditions. In these experiments, none of the four promoters selected demonstrated any prominent dependence upon STM2239 as was observed for the STM2237 promoter (Figure 3.6, Figure 3.3A). Based on these observations, we conclude that their behaviour from the microarray may be an artefact of the experimental conditions, given that the strains were grown in rich media supplemented with IPTG.

Discussion

In this study, we characterized the virulence contribution and regulatory mechanisms associated with the bacteriophage genes encoded within *Salmonella* Typhimurium SPI-12. Additionally, we established that SsrB plays an indirect role in activating genes within this locus through transcriptional read-through from the *sspH2* promoter. SspH2 itself played no part in contributing to the avirulent phenotypes associated with these genes. In fact, previous *sspH2* mutant studies established that this mutant does not exhibit a virulence defect in the murine model of infection [162].

Of the bacteriophage genes which exhibited a direct SsrB-dependent phenotype, *STM2239* demonstrated a prominent contribution to *S.* Typhimurium fitness and had a function consistent with a Q-like phage antiterminator. Q-antiterminators are transcriptional regulatory proteins that facilitate the activation and elongation at the late operon in lambdoid phages [180]. Their expression is tightly controlled by other transcriptional regulators that are only inactivated when the phage initiates the lytic cycle and activates another antiterminator N to read through terminators just upstream of the Q gene. Q itself then interacts with a Q utilization site (*qut*) on the chromosome, which generally overlaps with the -10 site of the P_R' late operon promoter [190, 213]. It also interacts directly with RNA polymerase holoenzyme through its β-flap to de-stabilize σ^{70} and allow for processive transcription through the terminator that normally prevents transcriptional activation of the phage late operon and facilitates transcript elongation through this entire region [195, 260]. The SPI-12 locus in *S.* Typhimurium is a remnant prophage which has undergone substantial genome degradation. As such, the intrinsic

terminators that may have been present to regulate this locus have been lost over time, as there is no evidence of terminator sequences upstream of either *STM2240* or *STM2237*. Loss of the terminators at this locus has permitted integration of the phage genes *STM2240* and *STM2239* into the SsrB regulon. Additionally, our data suggests that STM2239 has maintained its function as a Q-antiterminator protein in that it facilitates transcriptional activation of the adjacent gene cluster *STM2234-STM2237*.

Interestingly, this remnant phage locus has been linked to intracellular survival in previous work. Transposon mutagenesis studies investigating genes important for *S.* Typhimurium intra-macrophage survival found a Tn10 insertion at this locus that was defective for survival in macrophages [66, 168]. This Tn10 insertion, MS5076, was mapped to *STM2234* and *STM2235*, which likely encode tail fiber assembly and tail fiber proteins, respectively. A similar mutagenesis experiment by Worley et al. (2000) looking for SsrB-dependent genes also identified an insertion at this same site, MJW707, denoted as SsrB-regulated factor K or *srfK* [157]. These studies originally indicated that this locus shares homology with a DNA B-*pin* inversion system in *Shigella boydii* [261]. Inversion systems are common in bacteriophage tail fiber proteins to alter host target specificity of phages. They are characterized by inverted repeats within two adjacent tail fiber genes and an invertase that mediates inversion of the DNA between the two repeat sites [262]. While no specific function has been determined for STM2234 and STM2235, it is possible that this system may play a role in an adapted function in *S.* Typhimurium, though the SPI-12 locus does not encode a putative invertase gene. Phage tail fiber proteins have been found to contribute to pathogen-specific functions in other organisms;

PblA and PblB in *Streptococcus mitis* have been found to function as platelet adhesins that allow *S. mitis* to adhere to human platelets during infective endocarditis [209-211]. They have also been adapted for use as the target-specific component of a subset of bacteriocins whose protein structures resemble bacteriophage tails. Examples include the carotovoricins of the plant pathogen *Erwinia carotovora* and the R- and F-type pyocins of *Pseudomonas aeruginosa* [263-265]. The additional phage genes in this locus, STM2236, STM2237 and STM2240, exhibit some interesting features. STM2236 is orthologous to a head maturation protease, while STM2237 appears by BLAST analysis to be a hybrid of a holin protein and a portal protein, which may explain why it is not expressed properly. STM2240 and its putative function is largely a mystery, belonging to a large family of proteins of unknown function of phage origin (DUF122). Interestingly, this gene was found to be duplicated in a multi-drug resistant strain of *Salmonella* serovar [4,5,12:i:-] [266], though what bearing this may have on its function is unknown.

The SPI-12 locus described in this study is largely S. Typhimurium-specific, with only a few other sequenced serovars carrying the genes described. These include *S.* Paratyphi B and *S.* Heidelburg. In *S.* Typhi, the causative agent of typhoid fever, this region has undergone further degradation, having lost the six phage genes investigated in this study [94]. Evaluation of other serovars has also demonstrated a loss of this region, though interestingly most have retained the *sspH2* gene. While it remains unclear as to whether this variation amongst serovars is a by-product of inefficient prophage decay at this locus or whether it reflects an intended outcome remains to be seen. The current school of thought is that remnant phage islands are simply a snapshot of the continual

process of genome decay and thus it is believed these loci will eventually be lost by the host organism [24, 249, 267]. While this is likely true to some extent, there is the additional possibility that remnant phage genes may confer some employable function to their host, similar to their lysogenic conversion gene counterparts, and are thus maintained within the bacterial genome for a reason.

The integration of these genes into the SsrB regulon suggests that these bacteriophage gene products may contribute in some way to the pathogenicity of the strains that carry them, as they do demonstrate an *in vivo* virulence defect when deleted in *S.* Typhimurium (Figure 3.1). Further studies in *S.* Heidelburg and *S.* Paratyphi B would be required to establish whether they exhibit a similar regulatory cascade involving SsrB. The inclusion of the Q-like antiterminator protein STM2239 into the SsrB regulatory cascade is also intriguing. Prophages encoding Shiga-toxin proteins have been shown to be regulated by their own Q proteins, though transcriptional activation is mediated by induction of the phage lytic cycle in enterohemorrhagic *E.* coli (EHEC) during host infection [196, 268]. Whether STM2239 plays any additional role in *S.* Typhimurium gene regulation outside of this locus was not able to be determined, however perhaps its function on the SPI-12 locus itself is the desired outcome in specific host environments. All together, this study highlights a unique instance of the integration of bacteriophage genes into the regulatory network of a transcription factor in the context of *S.* Typhimurium virulence, and demonstrates that genes of bacteriophage origin may be adapted by pathogenic bacteria to improve fitness in the host environment.

Figure 3.1. Four bacteriophage genes within SPI-12 contribute to *Salmonella* Typhimurium virulence. (A and B) Competitive infection experiment between wild type and the strain indicated. Each data point represents an individual animal, and the horizontal bars represent the geometric means. (A) Wild type versus Δ*STM2234-STM2240* in C57BL/6 mice (n = 9). (B) Wild type competed against six individual phage gene mutants as indicated (n = 5). Mutants for *STM2234*, *STM2235*, *STM2236*, and *STM2239* showed a C.I. < 1.0. (C and D) Gentamicin protection assays conducted to assess gene contribution to intracellular replication in RAW264.7 cells. Macrophages were infected with the strains indicated, and the fold change was determined as the ratio of replication between 2 and 20 hours post-infection and normalized to wild type. (C) Δ*STM2239* exhibited a replication defect to half of wild type. Δ*ssrB* was used as a replication defect control (n = 5). (D) The replication defect observed in Δ*STM2239* can be complemented by providing the gene *in trans* from a plasmid (n = 5). (*, $P < 0.05$, **, $P < 0.01$, ***, $P < 0.001$).

Figure 3.1

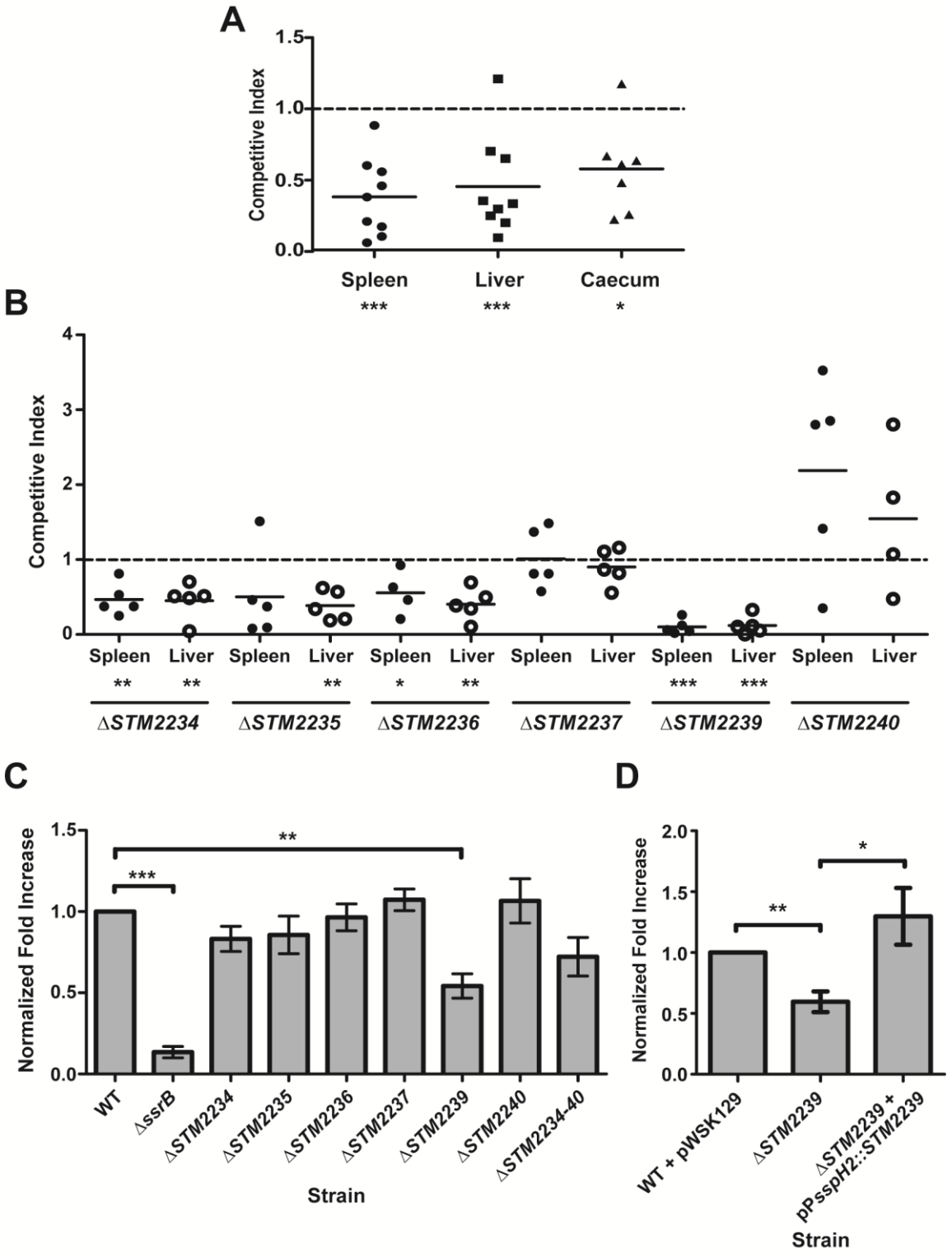

A
Competitive Index
1.5
1.0
0.5
0.0
Spleen
Liver
Caecum
***
***
*
B
Competitive Index
4
3
2
1
0
Spleen
Liver
**
**
**
*
**
***
***
ΔSTM2234
ΔSTM2235
ΔSTM2236
ΔSTM2237
ΔSTM2239
ΔSTM2240
C
Normalized Fold Increase
1.5
1.0
0.5
0.0
***
**
WT
ΔssrB
ΔSTM2234
ΔSTM2235
ΔSTM2236
ΔSTM2237
ΔSTM2239
ΔSTM2240
ΔSTM2234-40
Strain
D
Normalized Fold Increase
2.0
1.5
1.0
0.5
0.0
**
*
WT + pWSK129
ΔSTM2239
ΔSTM2239 + pPsspH2::STM2239
Strain

Figure 3.2 The SPI-12 locus is regulated by the virulence transcription factor SsrB. (A) Schematic diagram of the gene arrangement at SPI-12. Black arrows indicate genes of bacterial origin, blue arrows indicate genes of bacteriophage origin, and the red arrow identifies the lysogenic conversion gene *sspH2*. Roman numerals identify the intergenic regions evaluated for transcriptional activation. (B) Transcriptional activation at three intergenic loci in SPI-12 as determined by β-galactosidase assays in wild type and Δ*ssrB* backgrounds. The *sspH2* promoter (III) was used as a positive control. Data points with closed circles (●) denote wild type while closed squares (■) denote Δ*ssrB* background. Reporter strains were grown in LPM, pH 5.8 and β-galactosidase was quantified by chemiluminescence. Relative light units (RLU) were normalized to OD_{600} at each time point (n = 3). (C) Quantitative RT-PCR experiments evaluating the transcript levels of SPI-12 genes in wild type versus Δ*ssrB* backgrounds. *rsmC* was used as an internal control across strains while *sspH2* served as a positive control as it is known to be SsrB-dependent. All phage genes demonstrate a decrease in transcript levels in the *ssrB* mutant. (D) Evaluation of phage protein expression by immunoblotting. Phage proteins were tagged with –HA on the chromosome in wild type and Δ*ssrB* backgrounds and cell lysates of strains grown in LPM, pH 5.8 were detected with antibody directed against this tag. DnaK antibody was used as a loading control. Data is representative of three experiments with identical results. All four phage proteins tested showed decreased expression levels in the *ssrB* mutant. SseC blot was included as a control showing a typical SsrB-dependent expression phenotype.

Figure 3.2

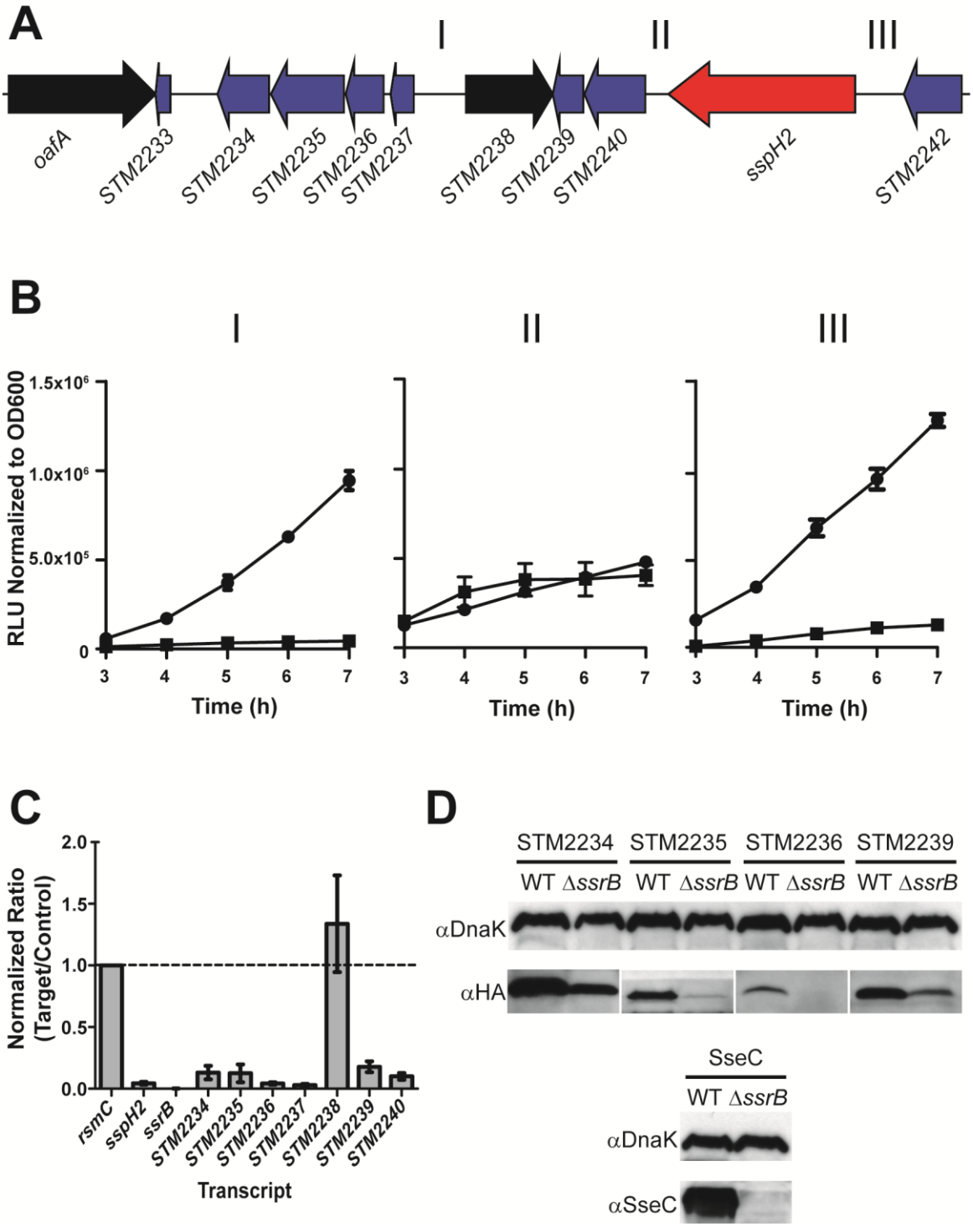

A
I
II
III
oafA
STM2233
STM2234
STM2235
STM2236
STM2237
STM2238
STM2239
STM2240
sspH2
STM2242
B
I
II
III
RLU Normalized to OD600
1.5x10⁶
1.0x10⁶
5.0x10⁵
0
3
4
5
6
7
Time (h)
C
Normalized Ratio (Target/Control)
2.0
1.5
1.0
0.5
0.0
rsmC
sspH2
ssrB
STM2234
STM2235
STM2236
STM2237
STM2238
STM2239
STM2240
Transcript
D
STM2234
STM2235
STM2236
STM2239
WT ΔssrB
αDnaK
αHA
SseC
αSseC

Figure 3.3 SsrB regulates STM2239 expression through transcriptional read-through from the *sspH2* promoter. (A) ChIP interaction profile of SsrB at the SPI-12 locus. The sole, prominent interaction peak is located at the *sspH2* promoter. This data is derived from the SsrB ChIP-on-chip experiment described in [252]. (B) Verification of a transcript that spans between *sspH2* and *STM2240*. First-strand cDNA derived from wild type, Δ*ssrB,* and Δ*STM2239* strains grown in LPM, pH 5.8 was used as template for PCR using primers specific for the 3' end of *sspH2* and the 5' end of *STM2240* genes. Amplification of product indicates that a transcript spanning that locus exists in wild type and Δ*STM2239* strains but not in Δ*ssrB*. (C) Immunoblot demonstrating STM2239 expression is dependent on the *sspH2* promoter (P_{sspH2}). STM2239 was tagged with –HA on the chromosome in wild type and ΔP_{sspH2} backgrounds and cell lysates of strains grown in LPM, pH 5.8 were detected with antibody directed against this tag. DnaK antibody was used as a loading control. Data is representative of two experiments with identical results.

Figure 3.3

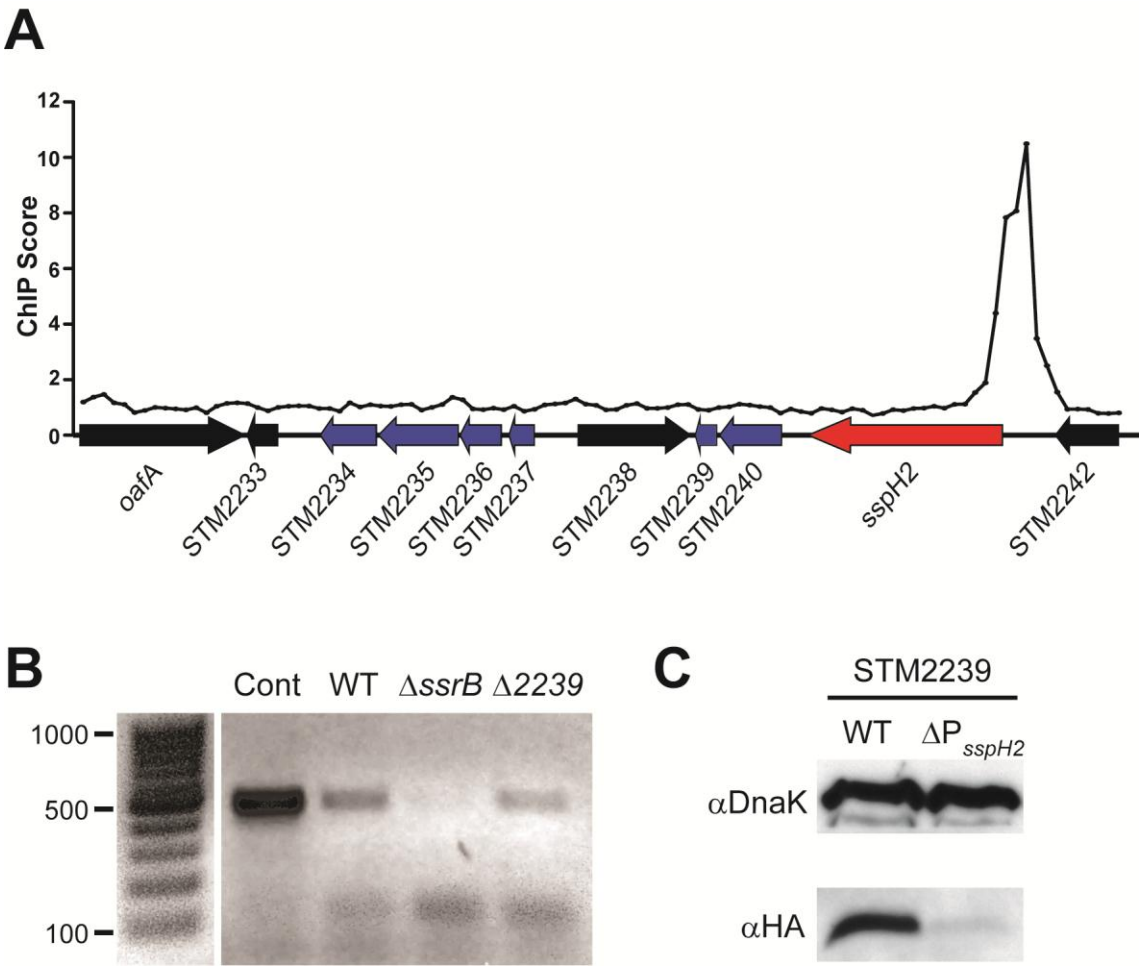

A
ChIP Score
12
10
8
6
4
2
0
oafA
STM2233
STM2234
STM2235
STM2236
STM2237
STM2238
STM2239
STM2240
sspH2
STM2242
B
Cont
WT
ΔssrB Δ2239
1000
500
100
C
STM2239
WT
ΔP sspH2
αDnaK
αHA

Figure 3.4 STM2239 regulates the *STM2237* operon. (A) Transcriptional activation at the *STM2237* promoter as determined by β-galactosidase assays in wild type and Δ*STM2239* backgrounds. Reporter strains were grown in LPM, pH 5.8 and β-galactosidase was quantified by chemiluminescence. Data points with closed circles (●) denote wild type while closed squares (■) denote Δ*STM2239* background. Relative light units (RLU) were normalized to OD_{600} at each time point (n = 3). (B) Quantitative RT-PCR experiments evaluating the transcript levels of SPI-12 genes in wild type versus Δ*STM2239* backgrounds. *rsmC* was used as an internal control across strains. Only the phage genes within the *STM2237* operon demonstrate a decrease in transcript levels in the *STM2239* mutant. (C) Evaluation of phage protein expression by immunoblotting. Phage proteins were expressed with an –HA tag under the control of the STM2237 promoter from a plasmid in wild type, Δ*ssrB,* and Δ*STM2239* strains and cell lysates of strains grown in LPM, pH 5.8 were detected with antibody directed against this tag. DnaK antibody was used as a loading control. Data is representative of three experiments with identical results. Decreased expression in Δ*STM2239* background is comparable to Δ*ssrB* background.

Figure 3.4

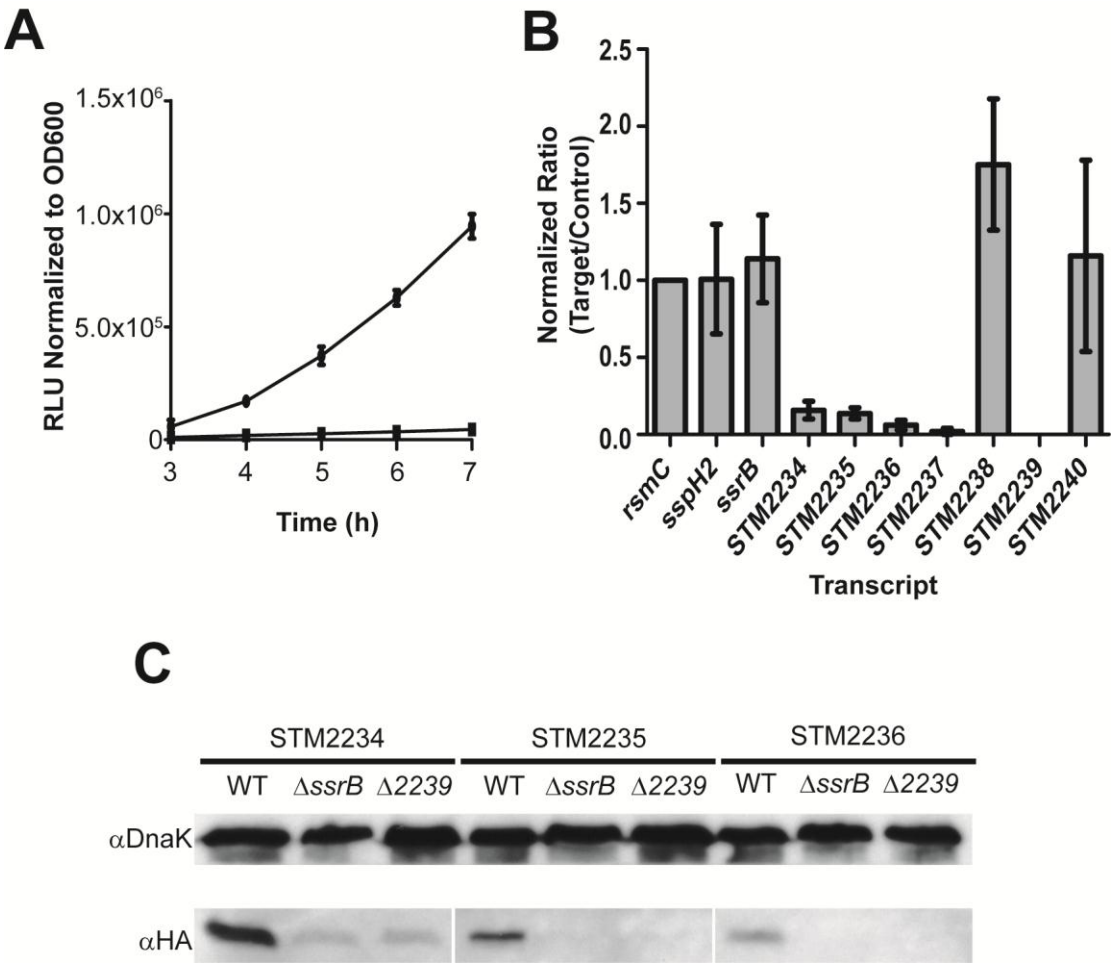

A
RLU Normalized to OD600
1.5x10⁶
1.0x10⁶
5.0x10⁵
0
3
4
5
6
7
Time (h)
B
Normalized Ratio (Target/Control)
2.5
2.0
1.5
1.0
0.5
0.0
rsmC
sspH2
ssrB
STM2234
STM2235
STM2236
STM2237
STM2238
STM2239
STM2240
Transcript
C
STM2234
STM2235
STM2236
WT
ΔssrB
Δ2239
αDnaK
αHA

Figure 3.5 STM2239 physically interacts with the *STM2237* promoter. (A) Chromatin immunoprecipitation (ChIP) paired with PCR demonstrated amplification of the *STM2237* promoter from chromosomal DNA immunoprecipitated with STM2239-3xFLAG, indicating binding of STM2239 at the *STM2237* promoter in *S*. Typhimurium SPI-12 when grown in LPM, pH 5.8. W = whole cell lysate, U = unbound, E = eluted DNA, + = positive control using genomic DNA. *xerD* and P*sseI* are negative and positive controls for ChIP, respectively. (B) 5' RACE mapping of the *STM2237* translational start site. Start site "g" is in bold, predicted -10 and -35 sites are highlighted in blue, and translational start site is in green.

Figure 3.5

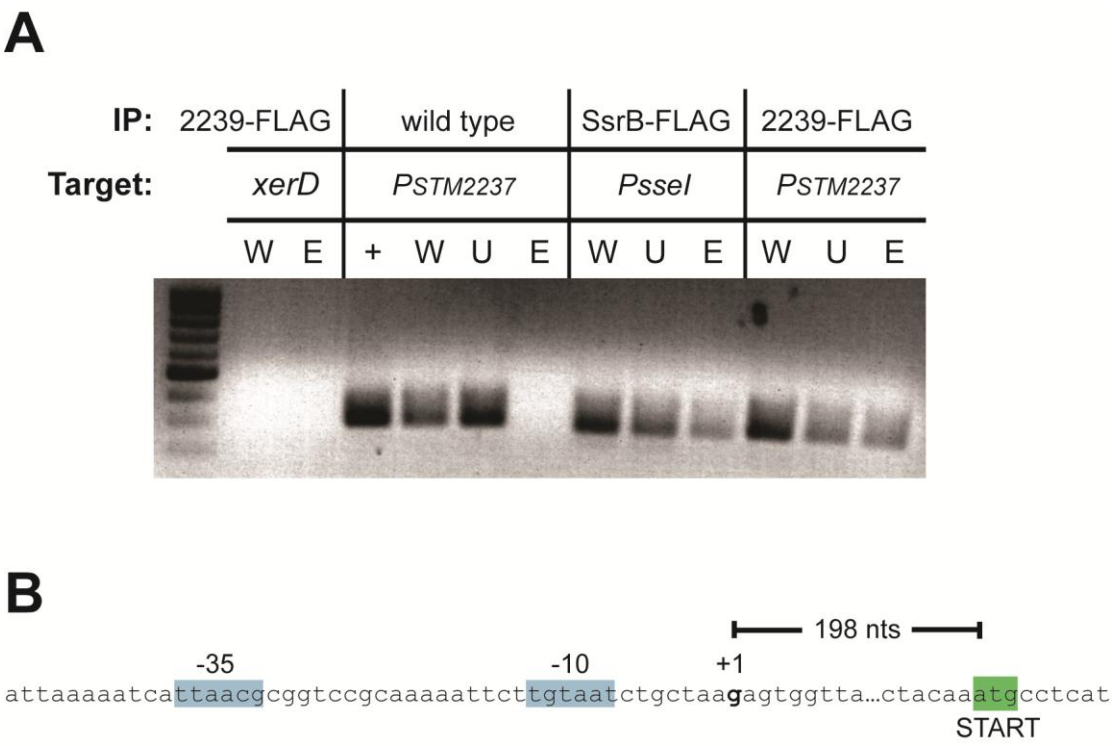

A
IP:
2239-FLAG
wild type
SsrB-FLAG
2239-FLAG
Target:
xerD
PSTM2237
PsseI
PSTM2237
W E
+ W U E
W U E
W U E
B
198 nts
-35
-10
+1
attaaaaatcattaacgcggtccgcaaaaattcttgtaatctgctaagagtggtta…ctacaaatgcctcat
START

Figure 3.6 Transcriptional reporter experiments investigating promoters of interest from the STM2239 microarray. pCS26 luminescence-based transcriptional reporters for four promoters of interest were tested for their dependence upon STM2239 for activation in LB (left column) and LPM, pH 5.8 media (right column). Promoters of interested are indicated along the right-hand side. Data points with closed circles (●) denote wild type while closed squares (■) denote Δ*STM2239* background.

Figure 3.6

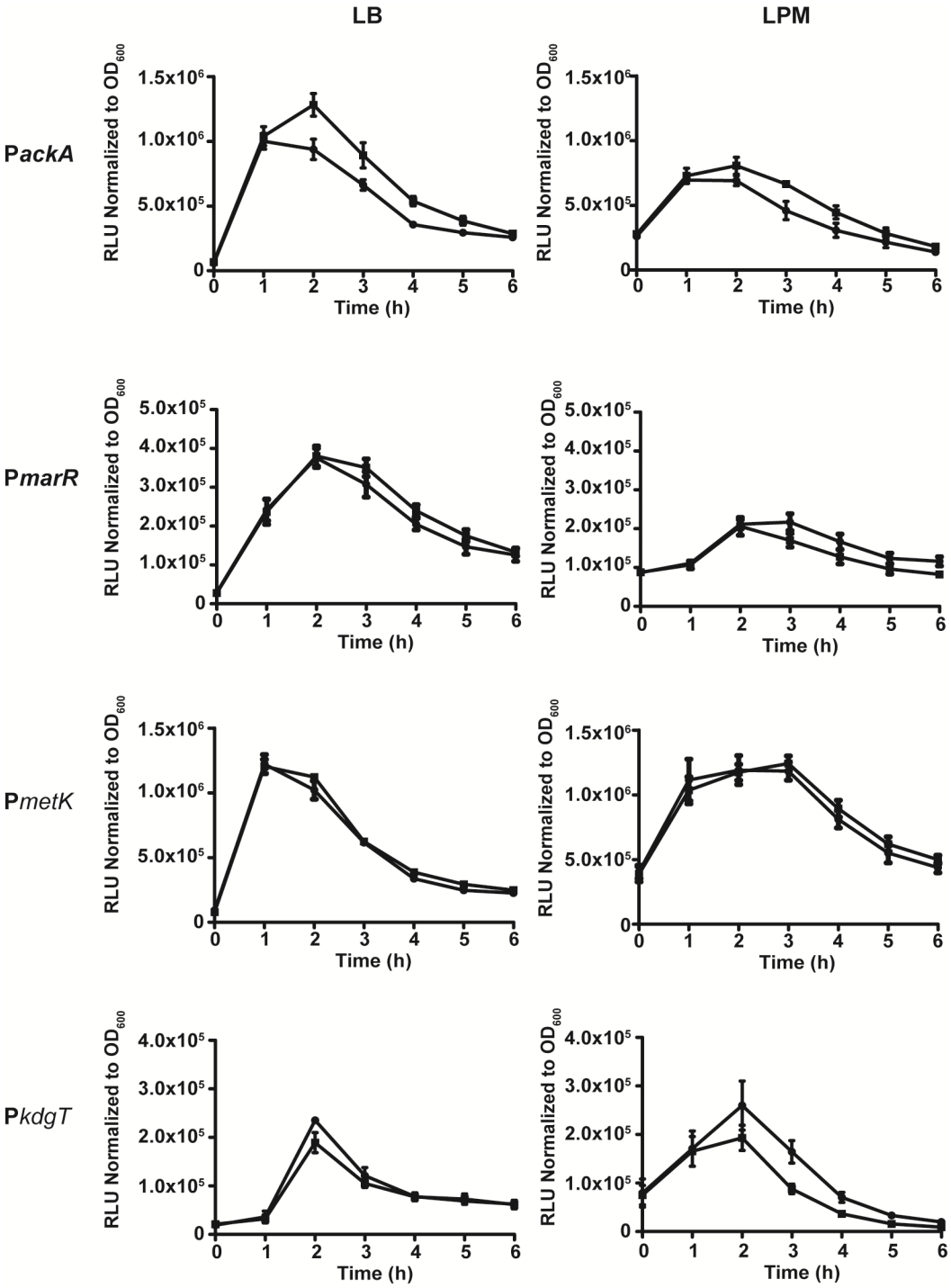

LB
LPM
PackA
PmarR
PmetK
PkdgT
RLU Normalized to OD600
Time (h)
1.5x10^6
1.0x10^6
5.0x10^5
4.0x10^5
3.0x10^5
2.0x10^5
1.0x10^5
0
0 1 2 3 4 5 6

Table 3.1 Summary of Phage Gene Characteristics and Phenotypes

Gene ID	Putative Function Description	BLAST-p Results	Virulence Phenotype		*In vitro* Expression	SsrB Microarray Fold Change[1]
			in vitro	*in vivo*		
STM2234	putative phage tail fiber assembly protein; pfam02413, Caudo-TAP superfamily	All S. Typhimurium; S. Heidelburg; S. Paratyphi B; S. 4,[5],12:i:; S. Javiana; S. Agona; S. Dublin	no	Yes	yes	-6.0
STM2235	hypothetical protein, putative phage tail fiber protein	S. Typhimurium LT2; S. Heidelburg; S. St.Paul; S. Paratyphi B; S. Javiana; S. Agona; S. Dublin	no	Yes	yes	-14.0
STM2236	hypothetical protein, phage head maturation protease; Peptidase U35 superfamily	All S. Typhimurium; S. Heidelburg; S. Paratyphi B; S. 4,[5],12:i:; S. Enteriditis; S. Javiana; S. Agona; S. Dublin; Phage SfV	no	Yes	yes	-30.1
STM2237	putative inner membrane protein; similar to lambda phage holin; looks like a fusion protein - N-term holin & C-term portal protein	All S. Typhimurium; S. Heidelburg; S. Paratyphi B; S. 4,[5],12:i:; S. Javiana; S. Agona; S. Dublin; Phage SfV	no	No	no	-8.5
STM2239	putative phage antiterminator; similar to antiterminator protein Q; pfam06530, Phage AntitermQ superfamily	All S. Typhimurium; S. Heidelburg; S. Paratyphi B str. SPB7; Escherichia coli ETEC H10407 and other E. coli; E. fergusonii; Phage cdtI; Shigella boydii; Pantoea sp.; various Yersinia spp.	yes	Yes	yes	-24.7
STM2240	putative cytoplasmic protein, putative phage protein; proteins of unknown function DUF122 superfamily, pfam06914	All S. Typhimurum; S. Heidelburg, S. St.Paul; S. Cholerasuis; S. Dublin; S. Weltevreden; S. Enteriditis; S. Schwarzengrund; Escherichia coli O103:H2 and other E. colis; Shigella boydii	no	No	no	-26.2

[1]Obtained from Tomljenovic-Berube et al. [252].

Table 3.2 STM2239 Microarray Genes of Interest

S. Typhimurium LT2 locus ID	Gene Name	log2 Median (n=3)	p by t test (n=3)	Predicted Function
STM2239[a]		-6.45732	1.63E-22	putative phage protein; homology to antiterminator protein Q of phage P5
STM2237*		-5.09231	1.83E-15	putative inner membrane protein
STM2236		-4.94473	2.07E-20	hypothetical protein
STM2235		-4.78025	4.71E-43	hypothetical protein
STM2234		-4.49065	3.41E-23	putative phage tail fiber assembly protein
STM2233		-4.02237	1.05E-16	putative cytoplasmic protein
STM3881	**rbsD**	-3.52919	2.49E-12	D-ribose pyranase; catalyzes the conversion between beta-pyran and beta-furan forms of D-ribose
STM1214	**ycfR**	-3.18236	3.53E-09	putative outer membrane protein
STM3828	**dgoA**	-3.17315	2.88E-39	2-oxo-3-deoxygalactonate 6-phosphate aldolase/galactonate dehydratase
STM0701	**speF**	-3.12397	2.87E-76	ornithine decarboxylase
STM0152	**aceE**	-3.06009	1.88E-108	pyruvate dehydrogenase subunit E1; E1 component; part of pyruvate dehydrogenase; forms a complex with DlaT and LpdC
STM0153	**aceF**	-2.89769	2.14E-04	dihydrolipoamide acetyltransferase; E2 component of pyruvate dehydrogenase multienzyme complex
STM0161	**kdgT***	-2.79983	1.28E-37	2-keto-3-deoxygluconate permease; transports degraded pectin products into the bacterial cell
STM3828.1N		-2.79733	3.47E-21	2-dehydro-3-deoxyphosphogalactonate aldolase
STM3090	**metK***	-2.71529	5.13E-36	S-adenosylmethionine synthetase; catalyzes the formation of S-adenosylmethionine from methionine and ATP
STM2338	**pta**	-2.69939	1.53E-63	phosphate acetyltransferase; catalyzes the synthesis of acetylphosphate or propionylphosphate from acetyl-CoA or propionyl-CoA and inorganic phosphate
STM2337	**ackA***	-2.65978	5.27E-47	acetate kinase; AckA utilizes acetate and can acetylate CheY which increases signal strength during flagellar rotation; also involved in conversion of acetate to aceyl-CoA
STM1519.S	**marA**	-2.60731	9.14E-19	DNA-binding transcriptional activator MarA; transcriptional activator of genes involved in the multiple antibiotic resistance (Mar) phenotype
STM3882	**rbsA**	-2.59391	7.30E-47	Ribose ABC transport system, ATP-binding protein RbsA; involved in D-ribose utilization
STM0162		-2.53135	2.36E-37	putative inner membrane protein
STM1518	**marB**	-2.46216	9.44E-08	multiple antibiotic resistance protein marB
STM3829	**dgoK**	-2.41968	1.42E-19	2-oxo-3-deoxygalactonate kinase
STM3512	**gntT**	-2.39647	9.62E-38	high-affinity gluconate permease
STM3361	**yhcN**	-2.35993	5.16E-13	putative outer membrane protein
STM1520	**marR***	-2.34775	1.15E-12	DNA-binding transcriptional repressor MarR; involved in the activation of both antibiotic resistance and oxidative stress genes
STM1211	**ndh**	-2.32043	1.48E-36	respiratory NADH dehydrogenase 2; cupric reductase (2nd module)
STM1423	**valW**	-2.31198	1.31E-02	valine tRNA 2A
STM0587	**ybdZ**	-2.25141	1.45E-06	putative cytoplasmic protein
STM0585	**fepA**	-2.25028	3.44E-57	outer membrane receptor FepA; outer membrane receptor of ferric enterobactin and colicins B and D; interacts with the TonB-ExbBD
STM3883	**rbsC**	-2.23703	1.57E-22	ribose ABC transporter permease protein; functions to transport ribose at high affinity; complexes with RbsA

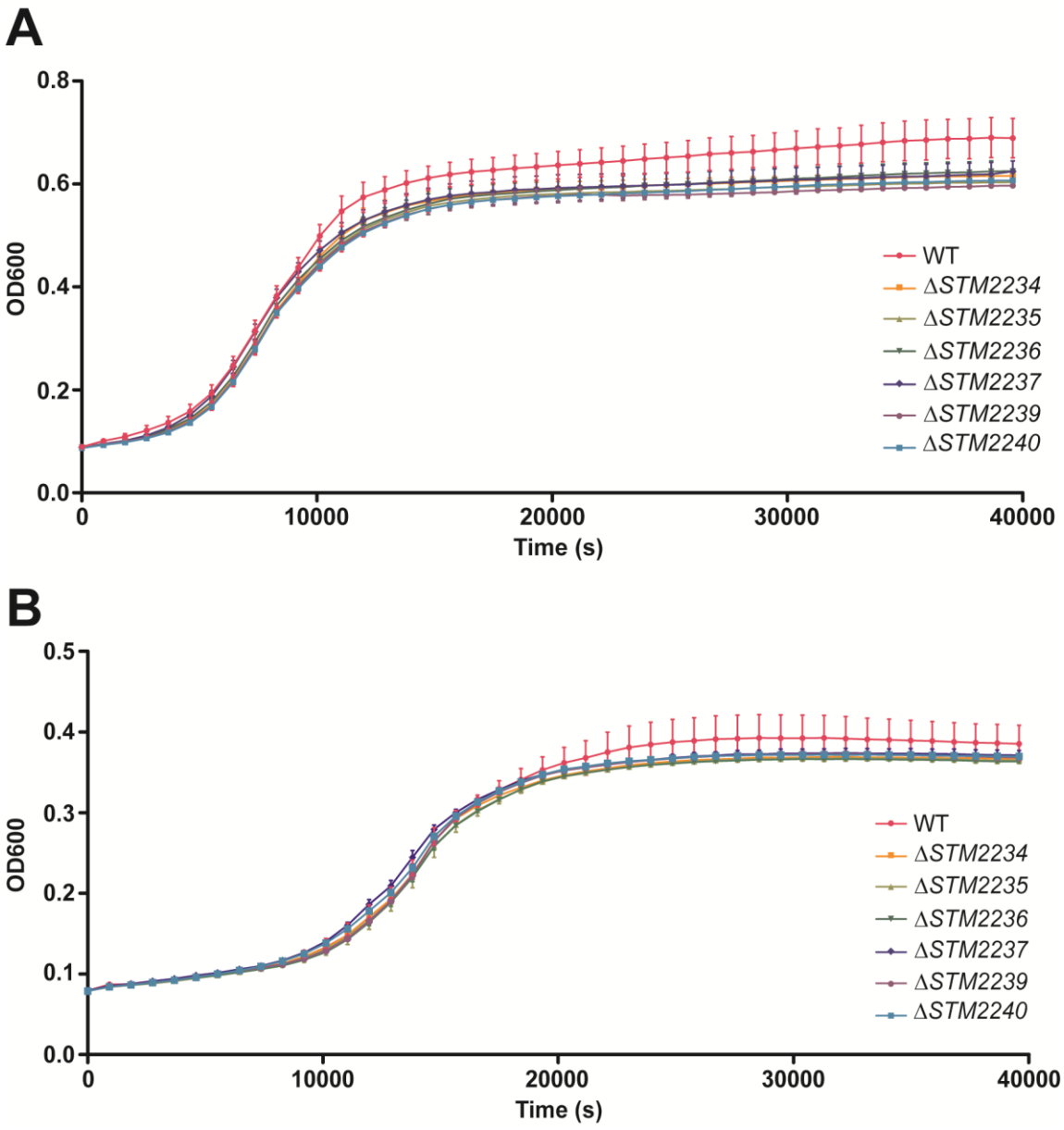
STM3808.S	**ibpB**	-2.23062	3.90E-15	heat shock chaperone IbpB; associates with aggregated proteins, together with ibpA
STM2068	**yeeF**	-2.14677	1.95E-39	putative amino acid transport protein
STM0586	**fes**	-2.1164	7.85E-31	enterobactin/ferric enterobactin esterase
STM0509		-2.09816	1.13E-27	putative outer membrane protein
STM0700	**potE**	-2.07471	9.27E-41	putrescine transporter
STM1424	**valV**	-2.0371	2.83E-06	valine tRNA 2B
STM3830	**dgoR**	-2.0264	3.69E-19	galactonate operon transcriptional repressor

[a] Overexpressed from pFLAG-STM2239

* Indicates gene promoters that were evaluated with transcriptional report

Figure 3.S1 Growth profiles of SPI-12 phage gene mutants over time. Phage gene mutant strains were grown in (A) LB medium and (B) LPM medium over the course of 11 hours (40,000 s) and OD_{600} was measured every 15 minutes. Data represents the average of three experiments.

Figure 3.S1


A
0.8
0.6
0.4
0.2
0.0
OD600
0
10000
20000
30000
40000
Time (s)
WT
ΔSTM2234
ΔSTM2235
ΔSTM2236
ΔSTM2237
ΔSTM2239
ΔSTM2240
B
0.5
0.4
0.3
0.2
0.1
0.0
OD600
0
10000
20000
30000
40000
Time (s)
WT
ΔSTM2234
ΔSTM2235
ΔSTM2236
ΔSTM2237
ΔSTM2239
ΔSTM2240

Supplementary Methods

S. Typhimurium growth curves

S. Typhimurium phage gene mutant strains were sub-cultured from washed overnight cultures into 200 μL LB (1:100) or LPM medium (1:50) in triplicate into a clear flat-bottomed 96-well plate (VWR). Cultures were grown for ~11 hours (40,000 s) at 37^{o}C with medium shaking in a Tecan microplate reader (Tecan Group Ltd.) set to take readings every 15 minutes. Data represents the average of three independent growth experiments.

Table 3.S1 List of strains and plasmids used in this study

Strains or Plasmids	Genotype or description	Reference
Strains		
DH5α	*supE44* Dlac*u169 (f80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*	Lab strain
DH5α λ pir	*supE44* Dlac*u169 (f80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1 λpir*	Lab strain
E. coli SM10 λ pir	*thi recA thr leu tonA lacY supE* RP4–2–Tc::*Mu λpir*	[269]
SL1344	Wild-type *S. enterica* sv. Typhimurium, SmR	[245]
SL1344 *ushA::Cm*	Wild-type SL1344 derivative containing a Cm cassette insertion at pseudogene *ushA*	[246]
SL1344 *ssrB::3FLAG-kan*	Encodes chromosomally FLAG-tagged *ssrB,* KanR	[143]
SL1344 *STM2239::3FLAG-kan*	Encodes chromosomally FLAG-tagged *STM2239*, KanR	This work
SL1344 Δ*ssrB*	Unmarked, in-frame deletion of *ssrB*	[252]
SL1344 Δ*STM2234-40*	Unmarked, in-frame deletion of *STM2234-40*	This work
SL1344 Δ*STM2234*	Unmarked, in-frame deletion of *STM2234*	This work
SL1344 Δ*STM2235*	Unmarked, in-frame deletion of *STM2235*	This work
SL1344 Δ*STM2236*	Unmarked, in-frame deletion of *STM2236*	This work
SL1344 Δ*STM2237*	Unmarked, in-frame deletion of *STM2237*	This work
SL1344 Δ*STM2239*	Unmarked, in-frame deletion of *STM2239*	This work
SL1344 Δ*STM2240*	Unmarked, in-frame deletion of *STM2240*	This work
SL1344 *STM2234::2HA-kan*	Encodes chromosomally HA-tagged *STM2234*, KanR	This work
SL1344 *STM2235::2HA-kan*	Encodes chromosomally HA-tagged *STM2235*, KanR	This work
SL1344 *STM2236::2HA-kan*	Encodes chromosomally HA-tagged *STM2236*, KanR	This work
SL1344 *STM2237::2HA-kan*	Encodes chromosomally HA-tagged *STM2237*, KanR	This work
SL1344 *STM2239::2HA-kan*	Encodes chromosomally HA-tagged *STM2239*, KanR	This work
SL1344 *STM2240::2HA-kan*	Encodes chromosomally HA-tagged *STM2240*, KanR	This work
SL1344 Δ*ssrB STM2234::2HA-kan*	Encodes chromosomally HA-tagged *STM2234* in *ssrB* mutant, KanR	This work
SL1344 Δ*ssrB STM2235::2HA-kan*	Encodes chromosomally HA-tagged *STM2235* in *ssrB* mutant, KanR	This work
SL1344 Δ*ssrB STM2236::2HA-kan*	Encodes chromosomally HA-tagged *STM2236* in *ssrB* mutant, KanR	This work
SL1344 Δ*ssrB STM2237::2HA-kan*	Encodes chromosomally HA-tagged *STM2237* in *ssrB* mutant, KanR	This work
SL1344 Δ*ssrB STM2239::2HA-kan*	Encodes chromosomally HA-tagged *STM2239* in *ssrB* mutant, KanR	This work
SL1344 Δ*ssrB STM2240::2HA-kan*	Encodes chromosomally HA-tagged *STM2240* in *ssrB* mutant, KanR	This work
SL1344 Δ*STM2239 STM2234::2HA-kan*	Encodes chromosomally HA-tagged *STM2234* in *STM2239* mutant, KanR	This work
SL1344 Δ*STM2239 STM2235::2HA-kan*	Encodes chromosomally HA-tagged *STM2235* in *STM2239* mutant, KanR	This work
SL1344 Δ*STM2239 STM2236::2HA-kan*	Encodes chromosomally HA-tagged *STM2236* in *STM2239* mutant, KanR	This work
SL1344 Δ*STM2239 STM2237::2HA-kan*	Encodes chromosomally HA-tagged *STM2237* in *STM2239* mutant, KanR	This work
SL1344 Δ*STM2239 STM2240::2HA-kan*	Encodes chromosomally HA-tagged *STM2240* in *STM2239* mutant, KanR	This work
SL1344 P*STM2237::*pIVET-PSTM2237	Merodiploid containing integrated P*STM2237-tnpR-lacZ* reporter	This work
SL1344 P*STM2240::*pIVET-PSTM2240	Merodiploid containing integrated P*STM2240-tnpR-lacZ* reporter	This work
SL1344 P*sspH2::*pIVET-PsspH2	Merodiploid containing integrated P*sspH2-tnpR-lacZ* reporter	This work

Plasmids		
pIVET5n	*tnpR-lacZ*, *sacB*, R6K ori, *bla*, AmpR	[222]
pCS26	pSC101 *ori, luxCDABE,* KanR	[246]
pWSK129	pSC101 *ori*, *lacZ*α, KanR	[270]
pFLAG-CTC	pBR322 *ori*, *tac* promoter, *lacI*, AmpR	Sigma
pIVET-PSTM2237	*STM2237* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-PSTM2240	*STM2240* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pIVET-PsspH2	*sspH2* promoter fused to *tnpR-lacZ* in pIVET5n, AmpR	This work
pWSK129-STM2234-HA	Expresses HA-tagged STM2234 controlled by STM2237 promoter	This work
pWSK129-STM2235-HA	Expresses HA-tagged STM2235 controlled by STM2237 promoter	This work
pWSK129-STM2236-HA	Expresses HA-tagged STM2236 controlled by STM2237 promoter	This work
pWSK129-P*sspH2::STM2239*	Expresses STM2239 under the control of the *sspH2* promoter	This work
pFLAG-STM2239	Expresses STM2239 under the control of the *tac* promoter (IPTG-inducible)	This work
pCS26-PackA	*ackA* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-PmarR	*marR* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-PmetK	*metK* promoter fused to *luxCDABE* in pCS26, KanR	This work
pCS26-PkdgT	*kdgT* promoter fused to *luxCDABE* in pCS26, KanR	This work

Table 3.S2 List of primers used in this study

Primer Name	Sequence
Deletion Mutants:	
STM2234 red F	gaatggccagtggtatacggtgggagcataATGAGGCATTTTgtgtaggctggagctgcttcg
STM2234 red R	gaacgtttcgagcacatgatttttatCTATTTATTCAGcatatgaatatcctcctta
STM2235 red F	gttgtttttgcttatccggAAGGAAAAGAGAATGCACAGAATTgtgtaggctggagctgcttcg
STM2235 red R	gtgaaatttttaaaatgcctcatttTTATGCTCCCACcatatgaatatcctcctta
STM2236 red F	ccaaaaccacaatctaggagaacACTCAAGATGAGTGAACGTgtgtaggctggagctgcttcg
STM2236 red R	caaccgtatgtgacggacaaagcttatcaaTCACACATTCAGcatatgaatatcctcctta
STM2237 red F	ctgtatcagcgaggtgagagataACTACAAATGCCTCATAACgtgtaggctggagctgcttcg
STM2237 red R	cgttcactcatcttgagtgttctcCTAGATTGTGGTcatatgaatatcctcctta
STM2239 red F	cgtgctggcctgatttgGTGGAGAAAGTTGATGCGTGATgtgtaggctggagctgcttcg
STM2239 red R	tacgcccctgcatggggcgtattaTTAAACGTCCATcatatgaatatcctcctta
STM2240 red F	gctcgataccgttcattGAGGTGATCGTGAGAGCGTTAgtgtaggctggagctgcttcg
STM2240 red R	gaatatcacgcatcaactttctccaccaaaTCAGGCCAGCACcatatgaatatcctcctta
dPsspH2-rev red F	ACTGATGGTGGCGGGAAGACATCCGCTTCCAATATGAAAGGGCATgtgtaggctggagctgcttcg
dPsspH2-rev red R	ATCGTGGAAAGCGACCCTGAAAAAATCGACACTTTAGCTGTAAAATGAcatatgaatatcctcctta
Chr HA & FLAG Tag Constructs:	
STM2234 BossHA F	ctggcgctagatgttgcacctgatattaactggccggaatatctgaataaaTATCCGTATGATGTTCCTGA
STM2234 BossHA R	cccgttgtgcggccgatttcaattgaacgtttcgagcacatgatttttatATATGAATATCCTCCTTAG
STM2235 BossHA F	gacgcctgttacagcttaaagtgaatggccagtggtatacggtgggagcaTATCCGTATGATGTTCCTGA
STM2235 BossHA R	gaacaagggttaattccgttgttttagtgaaatttttaaaatgcctcatttATATGAATATCCTCCTTAG
STM2236 BossHA F	cacctgcgattctgcgatcagaacgtggggcgatactgttgctgaatgtgTATCCGTATGATGTTCCTGA
STM2236 BossHA R	cttccggataagcaaaaacaacaaccgtatgtgacggacaaagcttatcaaATATGAATATCCTCCTTAG
STM2237 BossHA F	gatgaaatgctggtcagcgtcaacgcctcccggccagccaaaaccacaatcTATCCGTATGATGTTCCTGA
STM2237 BossHA R	acctcgccgctgtaacagcgaatttcacgttcactcatcttgagtgttctcATATGAATATCCTCCTTAG
STM2239 BossHA F	gatttattgatggctgcctttcaatgctggatctaacattagatatggacgttTATCCGTATGATGTTCCTGA
STM2239 BossHA R	caaatgtcattcatccagtaaataatacgcccctgcatggggcgtattaATATGAATATCCTCCTTAG
STM2239 BossFLAG3 R	gatttattgatggctgcctttcaatgctggatctaacattagatatggacgttGACTACAAAGACCATGAC
STM2240 BossHA F	gagttgatatttcgttttatcgatcgcgcgctggcaattggcgtgctggccTATCCGTATGATGTTCCTGA
STM2240 BossHA R	ccaacgctcaagaaccatctgaatatcacgcatcaactttctccaccaaaATATGAATATCCTCCTTAG
pWSK129 constructs	
pWSK-STM2234-37 F	ACGC**GTCGAC**cttaatgaagttttgtttttcctcctcc
pWSK-STM2234 R	ggaagatcttttattcagatattccggccag
pWSK-STM2235 R	ggaagatcttgctcccaccgtatacc
pWSK-STM2236 R	ggaagatctcacattcagcaacagtatcgc
pWSK-STM2237 R	ggaagatctgattgtggttttggctgg
pWSK-STM2239-40 F	ACGC**GTCGAC**gtcagaacgctggctctcg
pWSK-STM2239 R	ggaagatctaacgtccatatctaatgttagatcc
pWSK-STM2240 R	ggaagatctggccagcacgccaattgc
PsspH2-STM2239 F	ACGC**GTCGAC**gtttgtgcgtcgtatcggtg
PsspH2-STM2239 M	gaaccatctgaatatcacgcatcataacaaaaaaccctttataaatttac
PsspH2-STM2239 M2	gtaaatttataaaggttttttgttatgatgcgtgatattcagatggttc
PsspH2-STM2239 R2rc	ggaagatctttaaacgtccatatctaatgttagatccagcattg
pIVET constructs	
STM2237-IVET F	ccgctcgaggatgcttttagttgttcagtg

STM2237-IVET R	caccaattgttatgggttatgaggcatttgtag
STM2240-IVET F	ccgctcgagaggcgttaagagcaaacac
STM2240-IVET R	caccaattgttacagtaacgctctcacgatcac
sspH2-IVET F	ccgctcgaggtttgtgcgtcgtatcggtg
sspH2-IVET R	caccaattgttatccaatatgaaagggcataac
pFLAG construct	
STM2239 FLAG F	ccgCTCGAGatgcgtgatattcagatggttctt
STM2239 FLAG R	ggaagatctaacgtccatatctaatgttagatccag
5'RACE for PSTM2237	
GSP1 P2237	tgaccagcatttcatcaagg
GSP2 P2237	Caaaggtcagtgtcagagcg
GSP3 P2237	tctcgagcatttttcgcttc
RT-PCR Primer Sets	
rsmC RTF2	gaaaagcagccgcagtttag
rsmC RTR2	cagttggctaccaacatcca
2234 RTF	attgctgcggataaagatgc
2234 RTF	aataacgccgttgctgtagg
2235 RTF2	cgataccctggttacggttg
2235 RTR2	cggatagccaggtcttctga
2236 RTF	gggtatggttcggtctttga
2236 RTR	cttctgcgacccaggataaa
2237 RTF	cgctctgacactgacctttg
2237 RTR	tgaccagcatttcatcaagg
2238 RTF	gggggacagggaaatcagaa
2238 RTR	cggcaaaatgctcgcaatag
2239 RTF	gacacgtgtggcttgttctg
2239 RTR	cacgcacctttgcaatctta
2240 RTF	tctatcgtgctggtggtctg
2240 RTR	gcatggttgtcatctgttcg
sspH2 RTF	gtaatcgccgcatttatcgt
sspH2 RTR	aaatcgtccagatgcactcc
ssrB RTF	gcgttggccagcaatgaata
ssrB RTR	ttgcaatgccgctaacagaa
pCS26 Constructs	
pCS26-PackA F	ccgctcgagcgtaaaaccagttaaggattg
pCS26-PackA R	cgc**ggatcc**CTTACTCGACATGGAAGTACC
pCS26-PmarR F	ccgctcgagcattctgaagccgataaacg
pCS26-PmarR R	cgc**ggatcc**GGTGCTTTTCAAAATGGTTGC
pCS26-PkdgT F	ccg**ctcgag**GATTACGAATTTCTGCGCGA
pCS26-PkdgT R	cgc**ggatcc**TTTGATGTTCATGGTCTTACC
pCS26-PmetK F	ccgctcgagcacgaagaaaacgtcgtgt
pCS26-PmetK R	cgc**ggatcc**GTGTTTTGCCATATTTAATTTCAC

CHAPTER FOUR

DISCUSSION

DISCUSSION

Overview of Major Findings

The studies presented here provide insight into the role that the transcription factor SsrB plays in *Salmonella* Typhimurium pathogenesis. Previously, it had been well established that the SPI-2 T3SS was vital to intracellular survival within macrophages and facilitation of systemic disease in the mouse model of typhoid [62, 63, 80, 82]. At the same time, the SsrA-SsrB TCS was discovered to be the predominant regulator for this island. Over the last decade, further work has identified numerous individual virulence factors that are translocated by the SPI-2 T3SS into host cells during the intracellular stage of *Salmonella* infection. Of twenty-eight known SPI-2-associated virulence effectors, only two, SseF and SseG, are encoded within the SPI-2 island [63, 126, 156]. The remaining twenty-six are distributed throughout the genome, in association with horizontally-acquired DNA. In many cases, SsrB was found to coordinate expression of these virulence factors with the expression of SPI-2, thereby ensuring synchronized activation of both the T3SS and its associated effectors [101, 105, 111, 157, 159, 162]. Despite identification of a repertoire of SsrB-dependent promoters, a distinct SsrB DNA recognition motif had not yet been discerned.

Our work sought to identify the regulatory logic controlling SsrB in the context of intracellular virulence. By combining whole genome arrays for transcriptional profiling with genome-scale DNA binding studies, the *cis*-regulatory sequences involved in SsrB-dependent gene regulation were identified (Figure 2.2, Table 2.2). Additionally, comparative genomics of *S*. Typhimurium SPI-2 with the orthologous *Sodalis*

glossinidius SSR-3 island enabled identification of a degenerate SsrB DNA binding motif that coincided with previously published DNA footprinting data (Figure 2.4) [101, 102, 152]. This motif, revealed to be an 18 bp degenerate palindromic sequence, was confirmed by functional validation using the *ssaG* promoter. Mutagenesis of the *ssaG* promoter palindrome revealed it to be flexible to change, allowing significant substitutions while maintaining transcriptional activity (Figure 2.5, Figure 2.S2). The 18 bp consensus motif from SPI-2 provided a training set to guide further bioinformatics analysis of the ChIP-on-chip data to find related palindromes within the promoters of SsrB-dependent genes that had undergone natural selection, thus further validating the motif (Figure 2.6).

In the course of this work, transcriptional profiling of the SsrB regulon identified six genes in SPI-12 that exhibited SsrB dependence for transcription but did not have a readily identifiable SsrB binding site in upstream regulatory DNA. These genes are bacteriophage in origin, and part of a horizontally-acquired remnant phage which encodes a known SsrB-dependent effector SspH2. This locus had previously been identified as contributing to intramacrophage survival and *S*. Typhimurium virulence, prompting us to evaluate the genes within SPI-12 as putative novel virulence factors [66, 168, 169]. Assessment of bacterial fitness contribution for each individual phage gene using *in vivo* and *in vitro* infection experiments demonstrated a role for four of six of these genes in *S.* Typhimurium fitness, while STM2239 was specifically found to influence macrophage intracellular survival (Figure 3.1). SsrB was found to contribute to regulation of SPI-12 indirectly by transcriptional read-through from the *sspH2* promoter (Figure 3.2, Figure

3.5). This read-through gives rise to expression of STM2239, a putative Q-like antiterminator, which was required for activation of the *STM2237-34* operon as evidenced by direct interaction with the *STM2237* promoter (Figure 3.3, Figure 3.4). These findings suggest that genes of bacteriophage origin can contribute to the fitness of bacterial pathogens within the host environment, and reveal a unique regulatory cascade incorporating a phage regulatory mechanism into a bacterial transcription factor regulon.

Remaining Questions: SsrB-Dependent Regulation

The acquisition of SPI-2 into the *Salmonella* genome provided a quantifiable fitness advantage for survival in the host environment. This was demonstrated through functional transfer of SPI-2 into *S. bongori*, where SPI-2 was effectively able to produce a T3SS and translocate effector proteins into the host during intracellular infection [271]. Additionally, it is well established that SPI-2 T3SS mutants are avirulent [62, 63, 80, 82]. Nonetheless, complete intracellular survival and systemic propagation of disease was impeded likely due to the absence in *S. bongori* of all non-SPI-2 encoded effectors involved in intracellular pathogenesis [90, 162, 271, 272]. Expression of the T3SS in this scenario would have been mediated by SsrA-SsrB, as would have been the case when SPI-2 was first introduced as a horizontally-acquired locus in what is now *S. enterica*. Many of the effectors that are coordinated by SsrB also exhibit evidence of horizontal acquisition. Interestingly, initial examination of SsrB DNA footprinted promoters implied that SsrB had an affinity for AT-rich sequence [101, 102]. This is indeed evident when evaluating the SsrB palindrome for tolerance to base substitution. All but two

positions within each half site show a preference for A or T, and the degeneracy of the overall palindrome across SsrB-dependent promoters implies great flexibility in the types of sequence that SsrB recognizes (Figure 2.S2, Figure 2.6). Additionally, SsrB has exhibited traits of promiscuity in that it was found to bind throughout the *Salmonella* chromosome, including within CDS, and showed lower sequence specificity *in vitro* (Figure 2.2 and Tomljenovic-Berube AM, unpublished observations). Given that most horizontally-acquired DNA exhibits lower GC content (and thus is more AT rich) [13, 273], and that many SsrB-dependent factors are products of HGT, including SPI-2 itself, affinity for AT-rich sequence may be one mechanism in which SsrB initially recognizes target DNA. This flexibility on the part of both the TF and *cis*-regulatory sequence composition may decrease the threshold necessary for initial interaction and would reduce the period of neutral evolution necessary for palindrome development [237].

AT-rich DNA is known to be bound by nucleoid-associated proteins that contribute to gene expression regulation. One important factor, H-NS, functions as a repressor of horizontally-acquired DNA, including SPI-2 [2, 3, 102]. Additional nucleoid-associated factors such as Hha and YdgT work in concert with H-NS to target horizontally-acquired DNA [30, 110, 274]. Studies by Walthers and colleagues found that SsrB functions to counter H-NS silencing at SsrB-dependent promoters by displacement of H-NS [102, 111]. As part of our ChIP-on-chip analysis we compared SsrB binding to H-NS binding at SPI-2 and found that their binding profiles were distinct (data not shown). Nonetheless, the evidence that SsrB may antagonize H-NS and directly

activate transcription at SPI-2 lends support to the hypothesis that it may also perform a similar role at other AT-rich loci.

The degenerate nature of SsrB DNA recognition sequences raises several questions. What is the minimum sequence change required for SsrB-dependent recognition? What evolutionary pressures are at work to develop such a regulatory site? How would one assess this? We established that maintenance of one heptamer was sufficient to ensure recognition of the *ssaG* promoter despite atypical base composition in the second half of the sequence (Figure 2.5). Only introduction of intolerant bases within one heptamer (scrambled) or inversion of the heptamer sequences (7"-4-7' and H2H reporter) was able to abrogate transcriptional activation. Our work did note that there were differences in the palindrome sequence fidelity between subsets of SsrB-dependent promoters in terms of how strongly they scored against the consensus position weight matrix (Figure 2.6A). In general, most showed some conservation in the right heptamer with variation in the left heptamer and spacer. Interestingly, an intricate study looking at the DNA recognition sequences for the ancestral virulence regulator PhoP identified select nucleotide differences between ancestral and horizontally-acquired binding motifs [36, 275]. This study also demonstrated that PhoP-dependent promoters can be divided into subclasses within PhoP box sequences which reflected higher order promoter architecture affecting the mechanism of regulation and conversely promoter strength [36, 276]. Similar bioinformatics studies might be applied to SsrB-dependent promoters to evaluate whether transcriptional activation correlates with patterns within the recognition motif.

The identification of the SsrB recognition motif lends itself to exploration of the questions stated above. One potential avenue of research could involve experiments to "evolve" an SsrB-dependent promoter in the laboratory. Numerous studies have attempted to do this using adaptive laboratory evolution, where bacteria of interest are grown in iterative rounds of culture under a specified selective pressure [277-279]. Following numerous replication cycles, evolved isolates are sequenced to identify adaptive changes that have been incurred in the genome through natural selection [277-279]. This type of work has been done to assess adaptation to new carbon sources, improve growth in minimal media conditions, engender antibiotic overproduction and most recently enhance pathogenic fitness and host range [280-283]. In a similar approach, clinical isolates of various pathogens have been compared to identify possible adaptations to a particular host niche, such as *Pseudomonas aeruginosa* isolates from cystic fibrosis patients [284]. Studies have identified gene sets within *S.* Typhimurium which are involved in intracellular replication in macrophages and long-term persistence in animals [64, 285], hence the promoter of an appropriate SsrB-dependent virulence effector could be used for targeted regulatory evolution in these environments [283]. One such candidate is SseI as it has been shown to be required for long-term systemic infection in mice [28, 286]. Evaluation of the minimal SsrB binding motif could be accomplished *in vitro* using the "scrambled" sequence construct as a template to generate combinatorially derived reporter constructs which contain base substitutions in the scrambled hexamer. Reporter experiments would allow for identification of key nucleotide changes needed to establish SsrB-dependent activation. Identification of

motifs that, upon mutagenesis of one to two bases, results in SsrB-dependent reporter activation could then be used in an *sseI* deletion strain to "re-introduce" a new copy of *sseI* under the control of a mutated SsrB motif. This strain could then be used in an adaptive laboratory evolution experiment to promote regulatory evolution of the deficient *sseI* promoter using the host environment as the selective pressure, such as iterative macrophage infection and/or persistent mouse infection. The outcome of such experiments could provide further insights into how SsrB recognition motifs evolve for pathoadaptation in the context of the host.

With respect to SsrB-dependent pathoadaptation, there remain several avenues of investigation to pursue. To date, the SsrB regulon has been assessed solely in *S.* Typhimurium, one of over 2,500 different *Salmonella enterica* serovars. Genome sequencing has revealed many differences between serovars, particularly with respect to horizontally-acquired DNA. One notable example is the serovar *S.* Typhi, which appears to have undergone reductive evolution due to its adaptation to the human host, having lost 5% of its functional genome as compared to *S.* Typhimurium [161, 287, 288]. *S.* Typhi lacks many SPI-2 associated virulence genes, including the *spv* locus, *sspH1*, *slrP, sseI, sseK1, sseK2, sseK3, sopD2,* and *sseJ* [71]. As SsrB regulates many of these genes in *S.* Typhimurium, it would be interesting to evaluate the SsrB regulon in *S.* Typhi due to its presumed contribution to host survival. For instance, one study has suggested that SPI-2 is dispensable for intracellular replication in *S.* Typhi [289]. In contrast, a transposon mutagenesis screen for avirulent *S.* Typhi mutants in a humanized mouse infection model showed that SPI-2 was necessary for infection [290]. As *S.* Typhi is a host-restricted

pathogen, studies relating to its virulence strategy have been limited to *in vitro* work and inferred by comparison to murine *S.* Typhimurium pathogenesis. The advent of humanized mouse models may permit further assessment of SPI-2- and conversely SsrB-dependent virulence contributions in this pathogen [290-292]. Additionally, investigation of the SsrB regulon in other host-specific serovars may provide further insights into how these pathogens modulate virulence gene expression in specific niches.

In complement to evaluating SsrB in other serovars, assessment of its orthologous role in *S. glossinidius* for symbiosis of the tsetse fly may also prove useful. While the SSR-3 locus has been implicated in symbiosis, its role has not be been experimentally validated in contrast to SSR-1 and SSR-2 [230, 293, 294]. SSR-3 and SPI-2 are fairly well conserved for most components of the T3SS, and SsrB shares 69% identity and 81% similarity with SG1279, the orthologous gene in *Sodalis* (Figure 2.S3) [230]. We demonstrated that the regulatory logic controlling SG1279 was sufficiently conserved for SsrB-dependent activation, and subsequent experiments in our laboratory have shown that SG1279 can complement an *ssrB* deletion in *S.* Typhimurium (Figure 2.5D, D.T. Mulder – personal communication). Given that *S. glossinidius* has adapted to a mutualistic lifestyle in an insect, it is likely that it senses different signals in this host. Hence, examination of SSR-3 symbiosis contribution as well as functional assessment of the regulators of this system may provide further insights into its role in mutualism.

The question of how SsrB interacts with other regulatory proteins is an additional potential avenue of research that has thus far remained unexplored. One important aspect to SsrB regulation is its activation by its cognate HK SsrA. The intravacuolar signal that

stimulates the SsrA-SsrB regulatory cascade remains unknown as well as how it interfaces with SsrB. Characterization of SsrA functional domains with particular emphasis on its periplasmic sensing domain would provide great insights into its putative regulatory signal and is currently being investigated in our laboratory. RNAP is another potential interaction partner to SsrB when bound to DNA. Most activating TFs interface with the RNAP holoenzyme as either class I or class II activators [295]. SsrB binding sites vary with respect to their proximity to the -10 and -35 sites, and preliminary analysis of SPI-2 promoters indicated no consensus on either class [102]. As well, direct interaction of SsrB with RNAP holoenzyme has never been assessed, thus meriting further study. Finally, questions remain regarding SsrB self-interaction with respect to its DNA interaction. Most bacterial TFs function as dimers, as evidenced by crystal structure determination as well as binding site composition [296]. As the SsrB recognition sequence is a palindrome composed of two heptamers , this implies that it too functions as a dimer. Structural characterization of the SsrB C-terminal DNA binding domain indicated dimerization is important for function, however it was implied that it may be possible that SsrB first binds DNA as a monomer prior to dimerization [152]. Thus further experiments evaluating SsrB dimerization in the context of DNA interaction may provide insights into its mode of recognition and may contribute to our understanding of TF binding site evolution.

Remaining Questions: SPI-12

Bacteriophages are common agents of HGT. This is particularly evident in *S.* Typhimurium, as its genome encodes several intact prophage in addition to remnant phage islands that bear lysogenic conversion genes and other fitness factors [297, 298]. Many of these loci encode virulence factors that play a role in *S.* Typhimurium pathogenesis. These include *sopE* encoded within SopEΦ, *gogB* in Gifsy-1, *sseI* in Gifsy-2, as well as *sopE2* and *sspH2* encoded within remnant prophage [27, 76, 162, 208, 243, 299, 300]. While much research has been invested into the virulence contribution of these lysogenic conversion genes, little attention has been directed towards genes involved in phage biology despite there being numerous phage ORFs in the *Salmonella* genome with uncharacterized functions. Lysogeny had been shown to contribute to bacterial fitness, but only with respect to ensuring immunity from superinfection by other phage as well as carriage of lysogenic conversion genes [24]. The findings presented here regarding the fitness conferred by remnant phage genes in SPI-12 provide rationale for additional study to uncover their function(s) during host-pathogen interactions.

Many questions remain regarding SPI-12 and the function of its genes. Comparative genomics reveals that the SPI-12 locus as presented here is largely *S.* Typhimurium specific, with evidence of phage island decay in other serovars. Some strains such as *S.* Heidelburg and *S.* Paratyphi B carry mostly the same genes while others like *S.* Typhi exhibit gene loss with exception to *sspH2* and some upstream phage genes. Still others encode pseudogenes at this site including inactivated forms of *sspH2* (Figure 4.1). These observations beg the question of why the island composition and architecture

has been retained and maintained in *S.* Typhimurium. A recent study by Koskiniemi et al. demonstrated that environmental selection can be a driver of gene loss [301]. Further evaluation of SPI-12 phage genes in serovars such as *S.* Heidelburg and *S.* Paratyphi B as well as other remnant phage genes throughout the *Salmonella* genome would be necessary to differentiate the selective pressures on this island.

We found that SPI-12 contributes to *S.* Typhimurium fitness within the host niche. This was demonstrated by several studies including our own [66, 168, 169], however ours specifically evaluated the contribution of each individual gene. Of all six genes tested, STM2239 demonstrated a prominent fitness contribution for both intracellular replication and systemic survival (Figure 3.1). Identification of this protein as a putative Q-like antiterminator lead to evaluation of its role as a regulator at the *STM2237* promoter and resulted in the discovery of the regulatory cascade in effect at SPI-12 (Figure 4.2). Transcriptional profiling of a strain over-expressing STM2239 did not reveal any additional STM2239-dependent genes outside of SPI-12 as verified by reporter experiments, however its function at the *STM2237* promoter alone does not directly correlate with fitness contribution, as deletion of STM2239 results in a more prominent survival defect than the deletion of all six SPI-12 phage genes (Figure 3.6, Figure 3.1). Interestingly, a recent study showed that the CII phage regulator from Stx2 phage inhibits expression of the LEE-encoded T3SS, thus showing that bacteriophage regulators can influence other genetic loci involved in bacterial pathogenesis [302]. Similarly, Q from Stx2 phage is also directly involved in transcriptional activation of Shiga toxin, another virulence determinant, although this process ultimately results in release of the toxin by

bacterial cell lysis as a bi-product of phage particle production [196, 268]. STM2239 has not been directly assessed for regulation of other SPIs, however our microarray data would suggest that it does not influence T3SS expression at either SPI-1 or SPI-2 (data not shown). Thus further functional characterization of this Q protein as an antiterminator may reveal its full contribution to gene regulation and *S.* Typhimurium pathogenesis. Future work could include identifying its putative *qut* site within $P_{STM2237}$ and confirming its interaction with RNAP, as well as structural characterization of the protein to identify important residues for function.

The functions of the other phage genes in the context of *S.* Typhimurium virulence remain a mystery – for now. STM2234 and STM2235 show robust expression under SPI-2 inducing conditions as shown by this work and others (Figure 3.2D)[105]. Additionally, expression of STM2234 within intracellular *Salmonella* has been confirmed experimentally (data not shown). BLAST results indicate that these genes resemble phage tail fiber genes. Phage tail fiber proteins have been implicated in a few studies involving bacterial pathogenesis and inter-species competition. The *Streptococcus mitis* PblA and PblB proteins were found to be secreted and function as adhesins during infective endocarditis [209-211]. Additionally, tail fiber proteins dictate target recognition specificity for a subset of bacteriocins. Examples include the carotovoricins of the plant pathogen *Erwinia carotovora* and the R- and F-type pyocins of *Pseudomonas aeruginosa* [263-265]. We found no evidence that STM2234 and STM2235 are secreted from *Salmonella* (data not shown), thus further work is necessary to evaluate their role within the bacterium. The putative functions of the remaining genes, STM2236,

STM2237, and STM2240, are less intuitive with respect to roles in the bacterial cell. STM2236 resembles a phage head maturation protease, which may have an applicable function in bacteria. STM2237 appears to be a hybrid gene composed of a holin gene and part of a portal protein gene and does not appear to be functionally expressed nor contribute to virulence. Nonetheless, one study of *S*. Typhi found a functional role for a remnant holin in cell envelope stability [303]. Finally, STM2240 belongs to a family of phage genes of unknown function (DUF122). This protein also does not appear to be expressed under any culture conditions tested (data not shown). Potential avenues of research to investigate expressed proteins could include co-immunoprecipitation studies to find interacting partners within the bacterial cell as well as cellular localization studies using cell fractionation and fluorescence microscopy. The results of these experiments may provide further insights as to adapted functions for these phage genes within *S.* Typhimurium.

Concluding Remarks

The studies presented here validate the hypothesis that SsrB functions to integrate *Salmonella* virulence. SsrB was found to play a vital role in regulating many horizontally-acquired genes necessary for *S.* Typhimurium fitness within the host environment, including SPI-2, numerous virulence genes, as well as the phage genes of SPI-12. Combined genomic, computational, and functional approaches ultimately lead to the identification of the SsrB regulon and discovery of its DNA recognition sequence. Additionally, evaluation of its regulatory function at SPI-12 led to the characterization of

the regulatory landscape governing phage gene expression at this site and exhibited how bacteriophage genes contribute to pathogen fitness within the host environment. These findings demonstrate how pathogens integrate horizontally-acquired DNA into existing regulatory networks through *cis*-regulatory evolution to ensure pathoadaptation, and provide further insights into the complex process of coordinating intra-host survival through transcriptional regulation.

Figure 4.1 SPI-12 variants from a selection of *Salmonella* serovars. A graphical representation of SPI-12 loci from seven *Salmonella enterica* serovars. The *pro2* tRNA and *narP* gene were used to define island boundaries. Arrows indicate gene direction while their colour indicates origin and/or status as follows: genes of bacterial origin (black), genes of phage origin (blue), lysogenic conversion gene *sspH2* (red), and pseudogenes (white stripe).

Figure 4.1

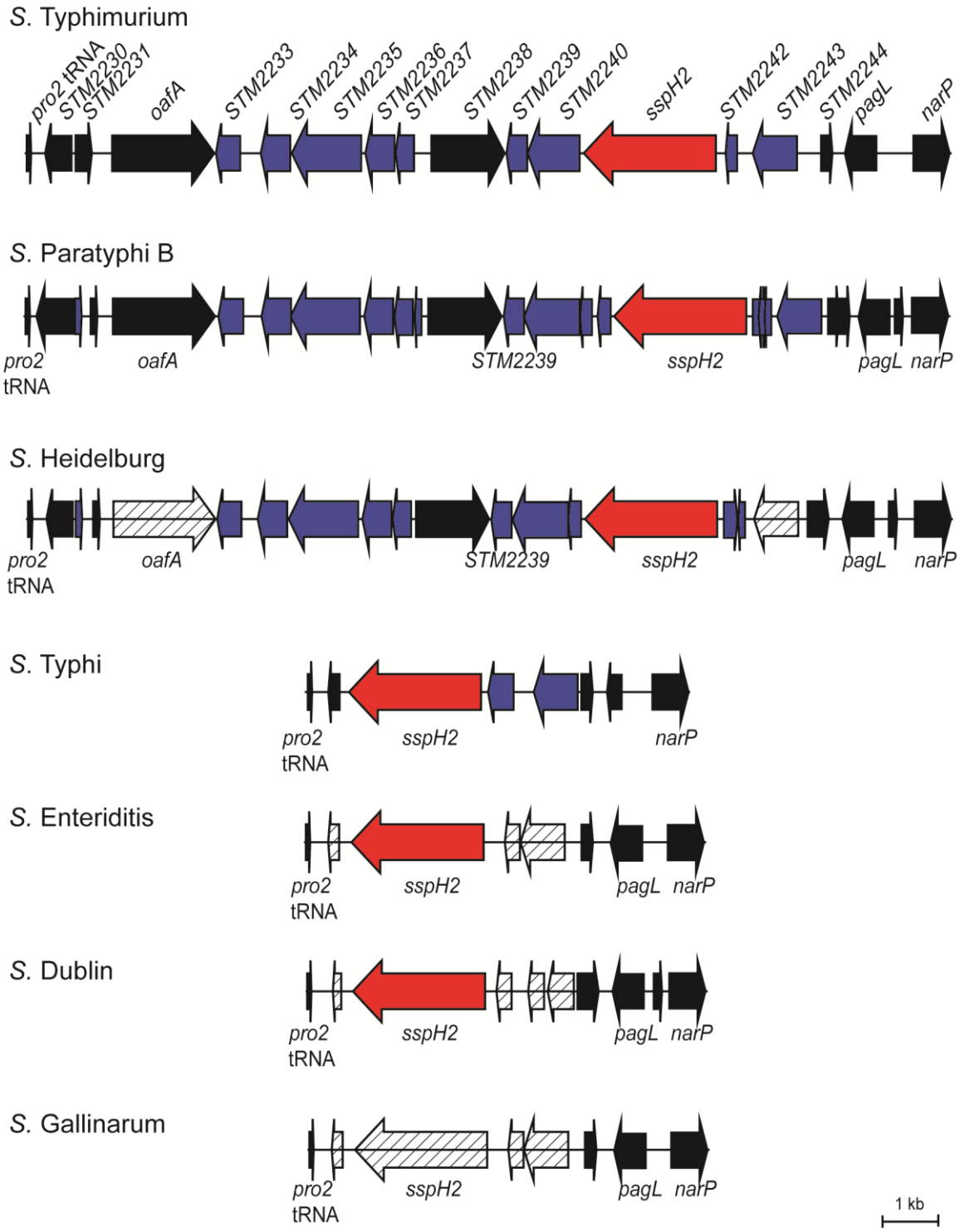
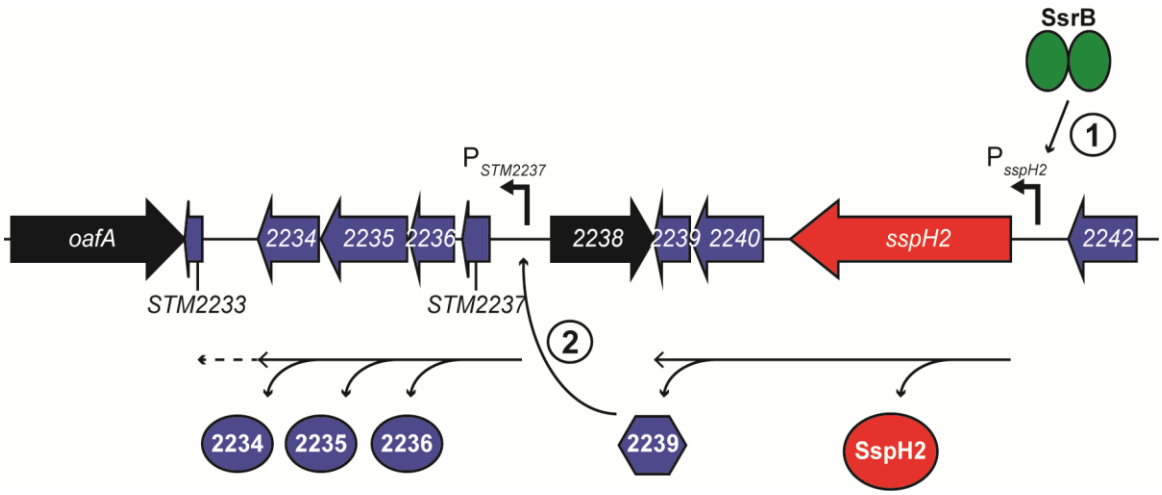

S. Typhimurium
pro2 tRNA
STM2230
STM2231
oafA
STM2233
STM2234
STM2235
STM2236
STM2237
STM2238
STM2239
STM2240
sspH2
STM2242
STM2243
STM2244
pagL
narP
S. Paratyphi B
pro2 tRNA
oafA
STM2239
sspH2
pagL
narP
S. Heidelburg
pro2 tRNA
oafA
STM2239
sspH2
pagL
narP
S. Typhi
pro2 tRNA
sspH2
narP
S. Enteriditis
pro2 tRNA
sspH2
pagL
narP
S. Dublin
pro2 tRNA
sspH2
pagL
narP
S. Gallinarum
pro2 tRNA
sspH2
pagL
narP
1 kb

Figure 4.2 A model for the SsrB-STM2239 regulatory cascade at SPI-12. Expression of SPI-12 genes is proposed to occur in two steps: (1) In the host environment, the transcription factor SsrB activates the expression of *sspH2* at the *sspH2* promoter (P_{sspH2}). This results in expression of SspH2 and STM2239 by way of transcriptional read-through from P_{sspH2}. (2) STM2239 acts directly on the *STM2237* promoter ($P_{STM2237}$) to activate expression of the STM2237 operon.

Figure 4.2


SsrB
1
P STM2237
P sspH2
oafA
2234
2235
2236
2238
2239
2240
sspH2
2242
STM2233
STM2237
2
2234
2235
2236
2239
SspH2

REFERENCES

1. Ochman H, Lawrence JG, Groisman EA. (2000) Lateral gene transfer and the nature of bacterial innovation. Nature 405(6784): 299-304.

2. Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M, Hinton JC. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. PLoS Pathog 2(8): e81.

3. Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H, Libby SJ, Fang FC. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science 313(5784): 236-238.

4. Doyle M, Fookes M, Ivens A, Mangan MW, Wain J, Dorman CJ. (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. Science 315(5809): 251-252.

5. Wray GA. (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8(3): 206-216.

6. Wittkopp PJ, Kalay G. (2011) Cis-regulatory elements: Molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13(1): 59-69.

7. Osborne SE, Walthers D, Tomljenovic AM, Mulder DT, Silphaduang U, Duong N, Lowden MJ, Wickham ME, Waller RF, Kenney LJ, Coombes BK. (2009) Pathogenic adaptation of intracellular bacteria by rewiring a *cis*-regulatory input function. Proc Natl Acad Sci U S A 106(10): 3982-3987.

8. Osborne SE, Tuinema BR, Mok MC, Lau PS, Bui NK, Tomljenovic-Berube AM, Vollmer W, Zhang K, Junop M, Coombes BK. (2012) Characterization of DalS, an ATP-binding cassette transporter for D-alanine, and its role in pathogenesis in *Salmonella enterica*. J Biol Chem 287(19): 15242-15250.

9. Chen I, Dubnau D. (2004) DNA uptake during bacterial transformation. Nat Rev Microbiol 2(3): 241-249.

10. Thomas CM, Nielsen KM. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol 3(9): 711-721.

11. Burrus V, Waldor MK. (2004) Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155(5): 376-386.

12. Frost LS, Leplae R, Summers AO, Toussaint A. (2005) Mobile genetic elements: The agents of open source evolution. Nat Rev Microbiol 3(9): 722-732.

13. Lawrence JG, Ochman H. (1998) Molecular archaeology of the *Escherichia coli* genome. Proc Natl Acad Sci U S A 95(16): 9413-9417.

14. Langille MG, Hsiao WW, Brinkman FS. (2008) Evaluation of genomic island predictors using a comparative genomics approach. BMC Bioinformatics 9: 329.

15. Gal-Mor O, Finlay BB. (2006) Pathogenicity islands: A molecular toolbox for bacterial virulence. Cell Microbiol 8(11): 1707-1719.

16. Langille MG, Hsiao WW, Brinkman FS. (2010) Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol 8(5): 373-382.

17. Hughes VM, Datta N. (1983) Conjugative plasmids in bacteria of the 'pre-antibiotic' era. Nature 302(5910): 725-726.

18. Norman A, Hansen LH, Sorensen SJ. (2009) Conjugative plasmids: Vessels of the communal gene pool. Philos Trans R Soc Lond B Biol Sci 364(1527): 2275-2289.

19. Enright MC, Robinson DA, Randle G, Feil EJ, Grundmann H, Spratt BG. (2002) The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A 99(11): 7687-7692.

20. Threlfall EJ. (2000) Epidemic *Salmonella typhimurium* DT 104--a truly international multiresistant clone. J Antimicrob Chemother 46(1): 7-10.

21. Hall RM. (2010) *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica*. Future Microbiol 5(10): 1525-1538.

22. Fournier PE, Vallenet D, Barbe V, Audic S, Ogata H, Poirel L, Richet H, Robert C, Mangenot S, Abergel C, Nordmann P, Weissenbach J, Raoult D, Claverie JM. (2006) Comparative genomics of multidrug resistance in *Acinetobacter baumannii*. PLoS Genet 2(1): e7.

23. Davies J, Davies D. (2010) Origins and evolution of antibiotic resistance. Microbiol Mol Biol Rev 74(3): 417-433.

24. Brussow H, Canchaya C, Hardt WD. (2004) Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68(3): 560-602.

25. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, Hendrix RW. (2000) Genomic sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299(1): 27-51.

26. Hardt WD, Urlaub H, Galan JE. (1998) A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. Proc Natl Acad Sci U S A 95(5): 2574-2579.

27. Coombes BK, Wickham ME, Brown NF, Lemire S, Bossi L, Hsiao WW, Brinkman FS, Finlay BB. (2005) Genetic and molecular analysis of GogB, a phage-encoded type III-secreted substrate in *Salmonella enterica* serovar Typhimurium with autonomous expression from its associated phage. J Mol Biol 348(4): 817-830.

28. Worley MJ, Nieman GS, Geddes K, Heffron F. (2006) *Salmonella typhimurium* disseminates within its host by manipulating the motility of infected cells. Proc Natl Acad Sci U S A 103(47): 17915-17920.

29. Miao EA, Miller SI. (2000) A conserved amino acid sequence directing intracellular type III secretion by *Salmonella typhimurium*. Proc Natl Acad Sci U S A 97(13): 7539-7544.

30. Coombes BK, Wickham ME, Lowden MJ, Brown NF, Finlay BB. (2005) Negative regulation of *Salmonella* pathogenicity island 2 is required for contextual control of virulence during typhoid. Proc Natl Acad Sci U S A 102(48): 17460-17465.

31. Silphaduang U, Mascarenhas M, Karmali M, Coombes BK. (2007) Repression of intracellular virulence factors in *Salmonella* by the Hha and YdgT nucleoid-associated proteins. J Bacteriol 189(9): 3669-3673.

32. Bina J, Zhu J, Dziejman M, Faruque S, Calderwood S, Mekalanos J. (2003) ToxR regulon of *Vibrio cholerae* and its expression in vibrios shed by cholera patients. Proc Natl Acad Sci U S A 100(5): 2801-2806.

33. Krukonis ES, DiRita VJ. (2003) DNA binding and ToxR responsiveness by the wing domain of TcpP, an activator of virulence gene expression in *Vibrio cholerae*. Mol Cell 12(1): 157-165.

34. Cathelyn JS, Ellison DW, Hinchliffe SJ, Wren BW, Miller VL. (2007) The RovA regulons of *Yersinia enterocolitica* and *Yersinia pestis* are distinct: Evidence that many RovA-regulated genes were acquired more recently than the core genome. Mol Microbiol 66(1): 189-205.

35. Perez JC, Groisman EA. (2009) Transcription factor function and promoter architecture govern the evolution of bacterial regulons. Proc Natl Acad Sci U S A 106(11): 4319-4324.

36. Zwir I, Latifi T, Perez JC, Huang H, Groisman EA. (2012) The promoter architectural landscape of the *Salmonella* PhoP regulon. Mol Microbiol 84(3): 463-485.

37. Stoebel DM, Free A, Dorman CJ. (2008) Anti-silencing: Overcoming H-NS-mediated repression of transcription in gram-negative enteric bacteria. Microbiology 154(Pt 9): 2533-2545.

38. Coburn B, Grassl GA, Finlay BB. (2007) *Salmonella*, the host and disease: A brief review. Immunol Cell Biol 85(2): 112-118.

39. Ochman H, Groisman EA. (1994) The origin and evolution of species differences in *Escherichia coli* and *Salmonella typhimurium*. EXS 69: 479-493.

40. Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Baumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR. (2011) *Salmonella bongori* provides insights into the evolution of the salmonellae. PLoS Pathog 7(8): e1002191.

41. Tindall BJ, Grimont PA, Garrity GM, Euzeby JP. (2005) Nomenclature and taxonomy of the genus *Salmonella*. Int J Syst Evol Microbiol 55(Pt 1): 521-524.

42. Fierer J, Guiney DG. (2001) Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. J Clin Invest 107(7): 775-780.

43. Hansen-Wester I, Hensel M. (2001) *Salmonella* pathogenicity islands encoding type III secretion systems. Microbes Infect 3(7): 549-559.

44. Ohl ME, Miller SI. (2001) *Salmonella*: A model for bacterial pathogenesis. Annu Rev Med 52: 259-274.

45. World Health Organization. (2005) Drug-resistant *Salmonella*. 2007(02/22): 4.

46. Crump JA, Luby SP, Mintz ED. (2004) The global burden of typhoid fever. Bull World Health Organ 82(5): 346-353.

47. Crump JA, Mintz ED. (2010) Global trends in typhoid and paratyphoid fever. Clin Infect Dis 50(2): 241-246.

48. Feasey NA, Dougan G, Kingsley RA, Heyderman RS, Gordon MA. (2012) Invasive non-typhoidal salmonella disease: An emerging and neglected tropical disease in africa. Lancet 379(9835): 2489-2499.

49. Scharff RL. (2012) Economic burden from health losses due to foodborne illness in the United States. J Food Prot 75(1): 123-131.

50. Thomas MK, Majowicz SE, Sockett PN, Fazil A, Pollari F, Dore K, Flint JA, Edge VL. (2006) Estimated numbers of community cases of illness due to *Salmonella*, *Campylobacter* and verotoxigenic *Escherichia coli*: Pathogen-specific community rates. Can J Infect Dis Med Microbiol 17(4): 229-234.

51. Jones BD, Ghori N, Falkow S. (1994) *Salmonella typhimurium* initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches. J Exp Med 180(1): 15-23.

52. Santos RL, Zhang S, Tsolis RM, Baumler AJ, Adams LG. (2002) Morphologic and molecular characterization of *Salmonella typhimurium* infection in neonatal calves. Vet Pathol 39(2): 200-215.

53. Galan JE. (1996) Molecular genetic bases of *Salmonella* entry into host cells. Mol Microbiol 20(2): 263-271.

54. Schlumberger MC, Hardt WD. (2006) *Salmonella* type III secretion effectors: Pulling the host cell's strings. Curr Opin Microbiol 9(1): 46-54.

55. Jepson MA, Collares-Buzato CB, Clark MA, Hirst BH, Simmons NL. (1995) Rapid disruption of epithelial barrier function by *Salmonella typhimurium* is associated with structural modification of intercellular junctions. Infect Immun 63(1): 356-359.

56. McCormick BA, Hofman PM, Kim J, Carnes DK, Miller SI, Madara JL. (1995) Surface attachment of *Salmonella typhimurium* to intestinal epithelia imprints the subepithelial matrix with gradients chemotactic for neutrophils. J Cell Biol 131(6 Pt 1): 1599-1608.

57. McCormick BA, Miller SI, Carnes D, Madara JL. (1995) Transepithelial signaling to neutrophils by salmonellae: A novel virulence mechanism for gastroenteritis. Infect Immun 63(6): 2302-2309.

58. Weinstein DL, O'Neill BL, Hone DM, Metcalf ES. (1998) Differential early interactions between *Salmonella enterica* serovar Typhi and two other pathogenic *Salmonella* serovars with intestinal epithelial cells. Infect Immun 66(5): 2310-2318.

59. House D, Bishop A, Parry C, Dougan G, Wain J. (2001) Typhoid fever: Pathogenesis and disease. Curr Opin Infect Dis 14(5): 573-578.

60. Hersh D, Monack DM, Smith MR, Ghori N, Falkow S, Zychlinsky A. (1999) The *Salmonella* invasin SipB induces macrophage apoptosis by binding to caspase-1. Proc Natl Acad Sci U S A 96(5): 2396-2401.

61. Puri AW, Broz P, Shen A, Monack DM, Bogyo M. (2012) Caspase-1 activity is required to bypass macrophage apoptosis upon *Salmonella* infection. Nat Chem Biol .

62. Cirillo DM, Valdivia RH, Monack DM, Falkow S. (1998) Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. Mol Microbiol 30(1): 175-188.

63. Hensel M, Shea JE, Waterman SR, Mundy R, Nikolaus T, Banks G, Vazquez-Torres A, Gleeson C, Fang FC, Holden DW. (1998) Genes encoding putative effector proteins of the type III secretion system of *Salmonella* pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. Mol Microbiol 30(1): 163-174.

64. Lawley TD, Chan K, Thompson LJ, Kim CC, Govoni GR, Monack DM. (2006) Genome-wide screen for *Salmonella* genes required for long-term systemic infection of the mouse. PLoS Pathog 2(2): e11.

65. Richter-Dahlfors A, Buchan AM, Finlay BB. (1997) Murine salmonellosis studied by confocal microscopy: *Salmonella typhimurium* resides intracellularly inside macrophages and exerts a cytotoxic effect on phagocytes in vivo. J Exp Med 186(4): 569-580.

66. Fields PI, Swanson RV, Haidaris CG, Heffron F. (1986) Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. Proc Natl Acad Sci U S A 83(14): 5189-5193.

67. Monack DM, Bouley DM, Falkow S. (2004) *Salmonella typhimurium* persists within macrophages in the mesenteric lymph nodes of chronically infected *Nramp1*+/+ mice and can be reactivated by IFNgamma neutralization. J Exp Med 199(2): 231-241.

68. Sheppard M, Webb C, Heath F, Mallows V, Emilianus R, Maskell D, Mastroeni P. (2003) Dynamics of bacterial growth and distribution within the liver during *Salmonella* infection. Cell Microbiol 5(9): 593-600.

69. Bhan MK, Bahl R, Bhatnagar S. (2005) Typhoid and paratyphoid fever. Lancet 366(9487): 749-762.

70. Hensel M. (2004) Evolution of pathogenicity islands of *Salmonella enterica*. Int J Med Microbiol 294(2-3): 95-102.

71. Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F. (2010) So similar, yet so different: Uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett 305(1): 1-13.

72. Cornelis GR. (2006) The type III secretion injectisome. Nat Rev Microbiol 4(11): 811-825.

73. Ochman H, Groisman EA. (1996) Distribution of pathogenicity islands in *Salmonella* spp. Infect Immun 64(12): 5410-5412.

74. Ellermeier JR, Slauch JM. (2007) Adaptation to the host environment: Regulation of the SPI1 type III secretion system in *Salmonella enterica* serovar Typhimurium. Curr Opin Microbiol 10(1): 24-29.

75. Bakowski MA, Braun V, Lam GY, Yeung T, Heo WD, Meyer T, Finlay BB, Grinstein S, Brumell JH. (2010) The phosphoinositide phosphatase SopB manipulates membrane surface charge and trafficking of the *Salmonella*-containing vacuole. Cell Host Microbe 7(6): 453-462.

76. Mirold S, Rabsch W, Rohde M, Stender S, Tschape H, Russmann H, Igwe E, Hardt WD. (1999) Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. Proc Natl Acad Sci U S A 96(17): 9845-9850.

77. Bajaj V, Lucas RL, Hwang C, Lee CA. (1996) Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of *hilA* expression. Mol Microbiol 22(4): 703-714.

78. Francis CL, Starnbach MN, Falkow S. (1992) Morphological and cytoskeletal changes in epithelial cells occur immediately upon interaction with *Salmonella typhimurium* grown under low-oxygen conditions. Mol Microbiol 6(21): 3077-3087.

79. Coburn B, Sekirov I, Finlay BB. (2007) Type III secretion systems and disease. Clin Microbiol Rev 20(4): 535-549.

80. Shea JE, Hensel M, Gleeson C, Holden DW. (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. Proc Natl Acad Sci U S A 93(6): 2593-2597.

81. Hensel M, Shea JE, Baumler AJ, Gleeson C, Blattner F, Holden DW. (1997) Analysis of the boundaries of *Salmonella* pathogenicity island 2 and the corresponding chromosomal region of *Escherichia coli* K-12. J Bacteriol 179(4): 1105-1111.

82. Ochman H, Soncini FC, Solomon F, Groisman EA. (1996) Identification of a pathogenicity island required for *Salmonella* survival in host cells. Proc Natl Acad Sci U S A 93(15): 7800-7804.

83. Brumell JH, Rosenberger CM, Gotto GT, Marcus SL, Finlay BB. (2001) SifA permits survival and replication of *Salmonella typhimurium* in murine macrophages. Cell Microbiol 3(2): 75-84.

84. Brumell JH, Grinstein S. (2004) *Salmonella* redirects phagosomal maturation. Curr Opin Microbiol 7(1): 78-84.

85. Vazquez-Torres A, Xu Y, Jones-Carson J, Holden DW, Lucia SM, Dinauer MC, Mastroeni P, Fang FC. (2000) *Salmonella* pathogenicity island 2-dependent evasion of the phagocyte NADPH oxidase. Science 287(5458): 1655-1658.

86. Chakravortty D, Hansen-Wester I, Hensel M. (2002) *Salmonella* pathogenicity island 2 mediates protection of intracellular *Salmonella* from reactive nitrogen intermediates. J Exp Med 195(9): 1155-1166.

87. Haraga A, Ohlson MB, Miller SI. (2008) Salmonellae interplay with host cells. Nat Rev Microbiol 6(1): 53-66.

88. Blanc-Potard AB, Solomon F, Kayser J, Groisman EA. (1999) The SPI-3 pathogenicity island of *Salmonella enterica*. J Bacteriol 181(3): 998-1004.

89. Gerlach RG, Jackel D, Stecher B, Wagner C, Lupas A, Hardt WD, Hensel M. (2007) *Salmonella* pathogenicity island 4 encodes a giant non-fimbrial adhesin and the cognate type 1 secretion system. Cell Microbiol 9(7): 1834-1850.

90. Knodler LA, Celli J, Hardt WD, Vallance BA, Yip C, Finlay BB. (2002) *Salmonella* effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. Mol Microbiol 43(5): 1089-1103.

91. Folkesson A, Lofdahl S, Normark S. (2002) The *Salmonella enterica* subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. Res Microbiol 153(8): 537-545.

92. Lambert MA, Smith SG. (2008) The PagN protein of *Salmonella enterica* serovar Typhimurium is an adhesin and invasin. BMC Microbiol 8: 142.

93. Mulder DT, Cooper CA, Coombes BK. (2012) Type VI secretion system-associated gene clusters contribute to pathogenesis of *Salmonella enterica* serovar Typhimurium. Infect Immun 80(6): 1996-2007.

94. Hansen-Wester I, Hensel M. (2002) Genome-based identification of chromosomal regions specific for *Salmonella* spp. Infect Immun 70(5): 2351-2360.

95. Wilson RP, Raffatellu M, Chessa D, Winter SE, Tukel C, Baumler AJ. (2008) The Vi-capsule prevents Toll-like receptor 4 recognition of *Salmonella*. Cell Microbiol 10(4): 876-890.

96. Wilson RP, Winter SE, Spees AM, Winter MG, Nishimori JH, Sanchez JF, Nuccio SP, Crawford RW, Tukel C, Baumler AJ. (2011) The Vi capsular polysaccharide prevents complement receptor 3-mediated clearance of *Salmonella enterica* serotype Typhi. Infect Immun 79(2): 830-837.

97. Pickard D, Wain J, Baker S, Line A, Chohan S, Fookes M, Barron A, Gaora PO, Chabalgoity JA, Thanky N, Scholes C, Thomson N, Quail M, Parkhill J, Dougan G. (2003) Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. J Bacteriol 185(17): 5055-5065.

98. Husain M, Jones-Carson J, Song M, McCollister BD, Bourret TJ, Vazquez-Torres A. (2010) Redox sensor SsrB Cys203 enhances *Salmonella* fitness against nitric oxide generated in the host immune response to oral infection. Proc Natl Acad Sci U S A 107(32): 14396-14401.

99. Hensel M, Hinsley AP, Nikolaus T, Sawers G, Berks BC. (1999) The genetic basis of tetrathionate respiration in *Salmonella typhimurium*. Mol Microbiol 32(2): 275-287.

100. Hensel M, Shea JE, Raupach B, Monack D, Falkow S, Gleeson C, Kubo T, Holden DW. (1997) Functional analysis of *ssaJ* and the *ssaK/U* operon, 13 genes encoding components of the type III secretion apparatus of *Salmonella* pathogenicity island 2. Mol Microbiol 24(1): 155-167.

101. Feng X, Walthers D, Oropeza R, Kenney LJ. (2004) The response regulator SsrB activates transcription and binds to a region overlapping OmpR binding sites at *Salmonella* pathogenicity island 2. Mol Microbiol 54(3): 823-835.

102. Walthers D, Carroll RK, Navarre WW, Libby SJ, Fang FC, Kenney LJ. (2007) The response regulator SsrB activates expression of diverse *Salmonella* pathogenicity island 2 promoters and counters silencing by the nucleoid-associated protein H-NS. Mol Microbiol 65(2): 477-493.

103. Feng X, Oropeza R, Kenney LJ. (2003) Dual regulation by phospho-OmpR of *ssrA/B* gene expression in *Salmonella* pathogenicity island 2. Mol Microbiol 48(4): 1131-1143.

104. Linehan SA, Rytkonen A, Yu XJ, Liu M, Holden DW. (2005) SlyA regulates function of *Salmonella* pathogenicity island 2 (SPI-2) and expression of SPI-2-associated genes. Infect Immun 73(7): 4354-4362.

105. Yoon H, McDermott JE, Porwollik S, McClelland M, Heffron F. (2009) Coordinated regulation of virulence during systemic infection of *Salmonella enterica* serovar Typhimurium. PLoS Pathog 5(2): e1000306.

106. Lim S, Kim B, Choi HS, Lee Y, Ryu S. (2006) Fis is required for proper regulation of *ssaG* expression in *Salmonella enterica* serovar Typhimurium. Microb Pathog 41(1): 33-42.

107. Bijlsma JJ, Groisman EA. (2005) The PhoP/PhoQ system controls the intramacrophage type three secretion system of *Salmonella enterica*. Mol Microbiol 57(1): 85-96.

108. Lee AK, Detweiler CS, Falkow S. (2000) OmpR regulates the two-component system SsrA-SsrB in *Salmonella* pathogenicity island 2. J Bacteriol 182(3): 771-781.

109. Osborne SE, Coombes BK. (2011) Transcriptional priming of *Salmonella* pathogenicity island-2 precedes cellular invasion. PLoS One 6(6): e21648.

110. Vivero A, Banos RC, Mariscotti JF, Oliveros JC, Garcia-del Portillo F, Juarez A, Madrid C. (2008) Modulation of horizontally acquired genes by the Hha-YdgT proteins in *Salmonella enterica* serovar Typhimurium. J Bacteriol 190(3): 1152-1156.

111. Walthers D, Li Y, Liu Y, Anand G, Yan J, Kenney LJ. (2011) *Salmonella enterica* response regulator SsrB relieves H-NS silencing by displacing H-NS bound in polymerization mode and directly activates transcription. J Biol Chem 286(3): 1895-1902.

112. Nikolaus T, Deiwick J, Rappl C, Freeman JA, Schroder W, Miller SI, Hensel M. (2001) SseBCD proteins are secreted by the type III secretion system of

Salmonella pathogenicity island 2 and function as a translocon. J Bacteriol 183(20): 6036-6045.

113. Klein JR, Jones BD. (2001) *Salmonella* pathogenicity island 2-encoded proteins SseC and SseD are essential for virulence and are substrates of the type III secretion system. Infect Immun 69(2): 737-743.

114. Zurawski DV, Stein MA. (2003) SseA acts as the chaperone for the SseB component of the *Salmonella* pathogenicity island 2 translocon. Mol Microbiol 47(5): 1341-1351.

115. Coombes BK, Brown NF, Kujat-Choy S, Vallance BA, Finlay BB. (2003) SseA is required for translocation of *Salmonella* pathogenicity island-2 effectors into host cells. Microbes Infect 5(7): 561-570.

116. Ruiz-Albert J, Mundy R, Yu XJ, Beuzon CR, Holden DW. (2003) SseA is a chaperone for the SseB and SseD translocon components of the *Salmonella* pathogenicity-island-2-encoded type III secretion system. Microbiology 149(Pt 5): 1103-1111.

117. Zurawski DV, Stein MA. (2004) The SPI2-encoded SseA chaperone has discrete domains required for SseB stabilization and export, and binds within the C-terminus of SseB and SseD. Microbiology 150(Pt 7): 2055-2068.

118. Miki T, Shibagaki Y, Danbara H, Okada N. (2009) Functional characterization of SsaE, a novel chaperone protein of the type III secretion system encoded by *Salmonella* pathogenicity island 2. J Bacteriol 191(22): 6843-6854.

119. Yu XJ, Liu M, Holden DW. (2004) SsaM and SpiC interact and regulate secretion of *Salmonella* pathogenicity island 2 type III secretion system effectors and translocators. Mol Microbiol 54(3): 604-619.

120. Yu XJ, McGourty K, Liu M, Unsworth KE, Holden DW. (2010) pH sensing by intracellular *Salmonella* induces effector translocation. Science 328(5981): 1040-1043.

121. Shotland Y, Kramer H, Groisman EA. (2003) The *Salmonella* SpiC protein targets the mammalian Hook3 protein function to alter cellular trafficking. Mol Microbiol 49(6): 1565-1576.

122. Lee AH, Zareei MP, Daefler S. (2002) Identification of a NIPSNAP homologue as host cell target for *Salmonella* virulence protein SpiC. Cell Microbiol 4(11): 739-750.

123. Yu XJ, Liu M, Matthews S, Holden DW. (2011) Tandem translation generates a chaperone for the *Salmonella* type III secretion system protein SsaQ. J Biol Chem 286(41): 36098-36107.

124. Pfeifer CG, Marcus SL, Steele-Mortimer O, Knodler LA, Finlay BB. (1999) *Salmonella typhimurium* virulence genes are induced upon bacterial invasion into phagocytic and nonphagocytic cells. Infect Immun 67(11): 5690-5698.

125. Hindle Z, Chatfield SN, Phillimore J, Bentley M, Johnson J, Cosgrove CA, Ghaem-Maghami M, Sexton A, Khan M, Brennan FR, Everest P, Wu T, Pickard D, Holden DW, Dougan G, Griffin GE, House D, Santangelo JD, Khan SA, Shea JE, Feldman RG, Lewis DJ. (2002) Characterization of *Salmonella enterica* derivatives harboring defined *aroC* and *Salmonella* pathogenicity island 2 type III secretion system (*ssaV*) mutations by immunization of healthy volunteers. Infect Immun 70(7): 3457-3467.

126. Hansen-Wester I, Stecher B, Hensel M. (2002) Type III secretion of *Salmonella enterica* serovar Typhimurium translocated effectors and SseFG. Infect Immun 70(3): 1403-1409.

127. Chakravortty D, Rohde M, Jager L, Deiwick J, Hensel M. (2005) Formation of a novel surface structure encoded by *Salmonella* pathogenicity island 2. EMBO J 24(11): 2043-2052.

128. Rappl C, Deiwick J, Hensel M. (2003) Acidic pH is required for the functional assembly of the type III secretion system encoded by *Salmonella* pathogenicity island 2. FEMS Microbiol Lett 226(2): 363-372.

129. Geddes K, Worley M, Niemann G, Heffron F. (2005) Identification of new secreted effectors in *Salmonella enterica* serovar Typhimurium. Infect Immun 73(10): 6260-6271.

130. Akeda Y, Galan JE. (2005) Chaperone release and unfolding of substrates in type III secretion. Nature 437(7060): 911-915.

131. Cooper CA, Zhang K, Andres SN, Fang Y, Kaniuk NA, Hannemann M, Brumell JH, Foster LJ, Junop MS, Coombes BK. (2010) Structural and biochemical characterization of SrcA, a multi-cargo type III secretion chaperone in *Salmonella* required for pathogenic association with a host. PLoS Pathog 6(2): e1000751.

132. Dai S, Zhou D. (2004) Secretion and function of *Salmonella* SPI-2 effector SseF require its chaperone, SscB. J Bacteriol 186(15): 5078-5086.

133. Cooper CA. (2011) The virulence chaperone network associated with the SPI-2 encoded type-three secretion system of *Salmonella enterica*. Ph.D. dissertation. Hamilton: McMaster University. 206p.

134. Guy RL, Gonias LA, Stein MA. (2000) Aggregation of host endosomes by *Salmonella* requires SPI2 translocation of SseFG and involves SpvR and the *fms-aroE* intragenic region. Mol Microbiol 37(6): 1417-1435.

135. Ramsden AE, Mota LJ, Munter S, Shorte SL, Holden DW. (2007) The SPI-2 type III secretion system restricts motility of *Salmonella*-containing vacuoles. Cell Microbiol 9(10): 2517-2529.

136. Kuhle V, Jackel D, Hensel M. (2004) Effector proteins encoded by *Salmonella* pathogenicity island 2 interfere with the microtubule cytoskeleton after translocation into host cells. Traffic 5(5): 356-370.

137. Robinson VL, Buckler DR, Stock AM. (2000) A tale of two components: A novel kinase and a regulatory switch. Nat Struct Biol 7(8): 626-633.

138. Deiwick J, Nikolaus T, Erdogan S, Hensel M. (1999) Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. Mol Microbiol 31(6): 1759-1773.

139. Coombes BK, Brown NF, Valdez Y, Brumell JH, Finlay BB. (2004) Expression and secretion of *Salmonella* pathogenicity island-2 virulence genes in response to acidification exhibit differential requirements of a functional type III secretion apparatus and SsaL. J Biol Chem 279(48): 49804-49815.

140. Miao EA, Freeman JA, Miller SI. (2002) Transcription of the SsrAB regulon is repressed by alkaline pH and is independent of PhoPQ and magnesium concentration. J Bacteriol 184(5): 1493-1497.

141. Garmendia J, Beuzon CR, Ruiz-Albert J, Holden DW. (2003) The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the *Salmonella typhimurium* SPI-2 type III secretion system. Microbiology 149(Pt 9): 2385-2396.

142. Lober S, Jackel D, Kaiser N, Hensel M. (2006) Regulation of *Salmonella* pathogenicity island 2 genes by independent environmental signals. Int J Med Microbiol 296(7): 435-447.

143. Duong N, Osborne S, Bustamante VH, Tomljenovic AM, Puente JL, Coombes BK. (2007) Thermosensing coordinates a cis-regulatory module for transcriptional activation of the intracellular virulence system in *Salmonella enterica* serovar Typhimurium. J Biol Chem 282(47): 34077-34084.

144. Wolanin PM, Thomason PA, Stock JB. (2002) Histidine protein kinases: Key signal transducers outside the animal kingdom. Genome Biol 3(10): REVIEWS3013.

145. Bijlsma JJ, Groisman EA. (2003) Making informed decisions: Regulatory interactions between two-component systems. Trends Microbiol 11(8): 359-366.

146. Cheung J, Hendrickson WA. (2010) Sensor domains of two-component regulatory systems. Curr Opin Microbiol 13(2): 116-123.

147. Zhang Z, Hendrickson WA. (2010) Structural characterization of the predominant family of histidine kinase sensor domains. J Mol Biol 400(3): 335-353.

148. Kenney LJ. (2010) How important is the phosphatase activity of sensor kinases? Curr Opin Microbiol 13(2): 168-176.

149. Volz K. (1993) Structural conservation in the CheY superfamily. Biochemistry 32(44): 11741-11753.

150. Baikalov I, Schroder I, Kaczor-Grzeskowiak M, Grzeskowiak K, Gunsalus RP, Dickerson RE. (1996) Structure of the *Escherichia coli* response regulator NarL. Biochemistry 35(34): 11053-11061.

151. Zhang JH, Xiao G, Gunsalus RP, Hubbell WL. (2003) Phosphorylation triggers domain separation in the DNA binding response regulator NarL. Biochemistry 42(9): 2552-2559.

152. Carroll RK, Liao X, Morgan LK, Cicirelli EM, Li Y, Sheng W, Feng X, Kenney LJ. (2009) Structural and functional analysis of the C-terminal DNA binding domain of the *Salmonella typhimurium* SPI-2 response regulator SsrB. J Biol Chem 284(18): 12008-12019.

153. Siebold C, Flukiger K, Beutler R, Erni B. (2001) Carbohydrate transporters of the bacterial phosphoenolpyruvate: Sugar phosphotransferase system (PTS). FEBS Lett 504(3): 104-111.

154. Choi J, Shin D, Yoon H, Kim J, Lee CR, Kim M, Seok YJ, Ryu S. (2010) *Salmonella* pathogenicity island 2 expression negatively controlled by EIIANtr-SsrB interaction is required for *Salmonella* virulence. Proc Natl Acad Sci U S A 107(47): 20506-20511.

155. Xu X, Hensel M. (2010) Systematic analysis of the SsrAB virulon of *Salmonella enterica*. Infect Immun 78(1): 49-58.

156. Figueira R, Holden DW. (2012) Functions of the *Salmonella* pathogenicity island 2 (SPI-2) type III secretion system effectors. Microbiology 158(Pt 5): 1147-1161.

157. Worley MJ, Ching KH, Heffron F. (2000) *Salmonella* SsrB activates a global regulon of horizontally acquired genes. Mol Microbiol 36(3): 749-761.

158. Rytkonen A, Poh J, Garmendia J, Boyle C, Thompson A, Liu M, Freemont JC, Hinton JCD, Holden DW. (2007) SseL, a *Salmonella* deubiquitinase required for macrophage killing and virulence. Proc Natl Acad Sci U S A 104(9): 3502-3507.

159. Coombes BK, Lowden MJ, Bishop JL, Wickham ME, Brown NF, Duong N, Osborne S, Gal-Mor O, Finlay BB. (2007) SseL is a *Salmonella*-specific translocated effector integrated into the SsrB-controlled *Salmonella* pathogenicity island 2 type III secretion system. Infect Immun 75(2): 574-580.

160. Chiu CH, Tang P, Chu C, Hu S, Bao Q, Yu J, Chou YY, Wang HS, Lee YS. (2005) The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. Nucleic Acids Res 33(5): 1690-1698.

161. McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P, Courtney L, Porwollik S, Ali J, Dante M, Du F, Hou S, Layman D, Leonard S, Nguyen C, Scott K, Holmes A, Grewal N, Mulvaney E, Ryan E, Sun H, Florea L, Miller W, Stoneking T, Nhan M, Waterston R, Wilson RK. (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature 413(6858): 852-856.

162. Miao EA, Scherer CA, Tsolis RM, Kingsley RA, Adams LG, Baumler AJ, Miller SI. (1999) *Salmonella typhimurium* leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. Mol Microbiol 34(4): 850-864.

163. Quezada CM, Hicks SW, Galan JE, Stebbins CE. (2009) A family of *Salmonella* virulence factors functions as a distinct class of autoregulated E3 ubiquitin ligases. Proc Natl Acad Sci U S A 106(12): 4864-4869.

164. Levin I, Eakin C, Blanc MP, Klevit RE, Miller SI, Brzovic PS. (2010) Identification of an unconventional E3 binding surface on the UbcH5 ~ Ub conjugate recognized by a pathogenic bacterial E3 ligase. Proc Natl Acad Sci U S A 107(7): 2848-2853.

165. Miao EA, Brittnacher M, Haraga A, Jeng RL, Welch MD, Miller SI. (2003) *Salmonella* effectors translocated across the vacuolar membrane interact with the actin cytoskeleton. Mol Microbiol 48(2): 401-415.

166. Hicks SW, Charron G, Hang HC, Galan JE. (2011) Subcellular targeting of *Salmonella* virulence proteins by host-mediated S-palmitoylation. Cell Host Microbe 10(1): 9-20.

167. Slauch JM, Lee AA, Mahan MJ, Mekalanos JJ. (1996) Molecular characterization of the *oafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: OafA is a member of a family of integral membrane trans-acylases. J Bacteriol 178(20): 5904-5909.

168. Baumler AJ, Kusters JG, Stojiljkovic I, Heffron F. (1994) *Salmonella typhimurium* loci involved in survival within macrophages. Infect Immun 62(5): 1623-1630.

169. Haneda T, Ishii Y, Danbara H, Okada N. (2009) Genome-wide identification of novel genomic islands that contribute to *Salmonella* virulence in mouse systemic infection. FEMS Microbiol Lett 297(2): 241-249.

170. Lwoff A. (1953) Lysogeny. Bacteriol Rev 17(4): 269-337.

171. Ptashne M. (2004) A genetic switch: Phage lambda revisited. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. pp.52-57.

172. Dodd IB, Shearwin KE, Egan JB. (2005) Revisited gene regulation in bacteriophage lambda. Curr Opin Genet Dev 15(2): 145-152.

173. Maurer R, Meyer B, Ptashne M. (1980) Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor. J Mol Biol 139(2): 147-161.

174. Meyer BJ, Maurer R, Ptashne M. (1980) Gene regulation at the right operator (O_R) of bacteriophage lambda. II. O_{R1}, O_{R2}, and O_{R3}: Their roles in mediating the effects of repressor and cro. J Mol Biol 139(2): 163-194.

175. Meyer BJ, Ptashne M. (1980) Gene regulation at the right operator (O_R) of bacteriophage lambda. III. lambda repressor directly activates gene transcription. J Mol Biol 139(2): 195-205.

176. Court DL, Oppenheim AB, Adhya SL. (2007) A new look at bacteriophage lambda genetic networks. J Bacteriol 189(2): 298-304.

177. Ptashne M, Jeffrey A, Johnson AD, Maurer R, Meyer BJ, Pabo CO, Roberts TM, Sauer RT. (1980) How the lambda repressor and cro work. Cell 19(1): 1-11.

178. Couturier M, Dambly C. (1970) Sequential activation of late functions in temperate bacteriophages. C R Acad Sci Hebd Seances Acad Sci D 270(2): 428-430.

179. Friedman DI, Court DL. (1995) Transcription antitermination: The lambda paradigm updated. Mol Microbiol 18(2): 191-200.

180. Roberts JW, Yarnell W, Bartlett E, Guo J, Marr M, Ko DC, Sun H, Roberts CW. (1998) Antitermination by bacteriophage lambda Q protein. Cold Spring Harb Symp Quant Biol 63: 319-325.

181. Kobiler O, Koby S, Teff D, Court D, Oppenheim AB. (2002) The phage lambda CII transcriptional activator carries a C-terminal domain signaling for rapid proteolysis. Proc Natl Acad Sci U S A 99(23): 14964-14969.

182. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. (2005) Switches in bacteriophage lambda development. Annu Rev Genet 39: 409-429.

183. Kobiler O, Rokney A, Oppenheim AB. (2007) Phage lambda CIII: A protease inhibitor regulating the lysis-lysogeny decision. PLoS One 2(4): e363.

184. Santangelo TJ, Artsimovitch I. (2011) Termination and antitermination: RNA polymerase runs a stop sign. Nat Rev Microbiol 9(5): 319-329.

185. Nodwell JR, Greenblatt J. (1991) The nut site of bacteriophage lambda is made of RNA and is bound by transcription antitermination factors on the surface of RNA polymerase. Genes Dev 5(11): 2141-2151.

186. Mogridge J, Mah TF, Greenblatt J. (1995) A protein-RNA interaction network facilitates the template-independent cooperative assembly on RNA polymerase of a stable antitermination complex containing the lambda N protein. Genes Dev 9(22): 2831-2845.

187. Adhya S, Gottesman M, De Crombrugghe B. (1974) Release of polarity in *Escherichia coli* by gene *N* of phage lambda: Termination and antitermination of transcription. Proc Natl Acad Sci U S A 71(6): 2534-2538.

188. Franklin NC. (1974) Altered reading of genetic signals fused to the *N* operon of bacteriophage lambda: Genetic evidence for modification of polymerase by the protein product of the *N* gene. J Mol Biol 89(1): 33-48.

189. Somasekhar G, Szybalski W. (1987) The functional boundaries of the Q-utilization site required for antitermination of late transcription in bacteriophage lambda. Virology 158(2): 414-426.

190. Yarnell WS, Roberts JW. (1992) The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell 69(7): 1181-1189.

191. Yang XJ, Hart CM, Grayhack EJ, Roberts JW. (1987) Transcription antitermination by phage lambda gene Q protein requires a DNA segment spanning the RNA start site. Genes Dev 1(3): 217-226.

192. Schleif R. (1972) The specificity of lamboid phage late gene induction (lamboid phage late gene specificity). Virology 50(2): 610-612.

193. Schechtman MG, Alegre JN, Roberts JW. (1980) Assay and characterization of late gene regulators of bacteriophages lambda82 and phi. J Mol Biol 142(3): 269-288.

194. Deighan P, Hochschild A. (2007) The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription elongation complex. Mol Microbiol 63(3): 911-920.

195. Deighan P, Diez CM, Leibman M, Hochschild A, Nickels BE. (2008) The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A 105(40): 15305-15310.

196. Neely MN, Friedman DI. (1998) Functional and genetic analysis of regulatory regions of coliphage H-19B: Location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. Mol Microbiol 28(6): 1255-1267.

197. Wagner PL, Neely MN, Zhang X, Acheson DW, Waldor MK, Friedman DI. (2001) Role for a phage promoter in Shiga toxin 2 expression from a pathogenic *Escherichia coli* strain. J Bacteriol 183(6): 2081-2085.

198. Plunkett G,3rd, Rose DJ, Durfee TJ, Blattner FR. (1999) Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. J Bacteriol 181(6): 1767-1778.

199. Zinder ND, Lederburg J. (1952) Genetic exchange in *Salmonella*. J Bacteriol 64(5): 679-699.

200. Schicklmaier P, Schmieger H. (1995) Frequency of generalized transducing phages in natural isolates of the *Salmonella typhimurium* complex. Appl Environ Microbiol 61(4): 1637-1640.

201. Hodgson DA. (2000) Generalized transduction of serotype 1/2 and serotype 4b strains of *Listeria monocytogenes*. Mol Microbiol 35(2): 312-323.

202. Canchaya C, Fournous G, Chibani-Chennoufi S, Dillmann ML, Brussow H. (2003) Phage as agents of lateral gene transfer. Curr Opin Microbiol 6(4): 417-424.

203. Waldor MK, Mekalanos JJ. (1996) Lysogenic conversion by a filamentous phage encoding cholera toxin. Science 272(5270): 1910-1914.

204. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S. (2000) The origins and ongoing evolution of viruses. Trends Microbiol 8(11): 504-508.

205. Boyd EF, Brussow H. (2002) Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol 10(11): 521-529.

206. Yokoyama K, Makino K, Kubota Y, Watanabe M, Kimura S, Yutsudo CH, Kurokawa K, Ishii K, Hattori M, Tatsuno I, Abe H, Yoh M, Iida T, Ohnishi M, Hayashi T, Yasunaga T, Honda T, Sasakawa C, Shinagawa H. (2000) Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli* O157:H7 strain derived from the Sakai outbreak. Gene 258(1-2): 127-139.

207. Figueroa-Bossi N, Uzzau S, Maloriol D, Bossi L. (2001) Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol Microbiol 39(2): 260-271.

208. Ho TD, Figueroa-Bossi N, Wang M, Uzzau S, Bossi L, Slauch JM. (2002) Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella enterica* serovar Typhimurium. J Bacteriol 184(19): 5234-5239.

209. Bensing BA, Siboo IR, Sullam PM. (2001) Proteins PblA and PblB of *Streptococcus mitis*, which promote binding to human platelets, are encoded within a lysogenic bacteriophage. Infect Immun 69(10): 6186-6192.

210. Bensing BA, Rubens CE, Sullam PM. (2001) Genetic loci of *Streptococcus mitis* that mediate binding to human platelets. Infect Immun 69(3): 1373-1380.

211. Mitchell J, Siboo IR, Takamatsu D, Chambers HF, Sullam PM. (2007) Mechanism of cell surface expression of the *Streptococcus mitis* platelet binding proteins PblA and PblB. Mol Microbiol 64(3): 844-857.

212. Kuhle V, Hensel M. (2004) Cellular microbiology of intracellular *Salmonella enterica*: Functions of the type III secretion system encoded by *Salmonella* pathogenicity island 2. Cell Mol Life Sci 61(22): 2812-2826.

213. Friedman DI, Court DL. (2001) Bacteriophage lambda: Alive and well and still doing its thing. Curr Opin Microbiol 4(2): 201-207.

214. Waldor MK, Friedman DI. (2005) Phage regulatory circuits and virulence gene expression. Curr Opin Microbiol 8(4): 459-465.

215. Prud'homme B, Gompel N, Carroll SB. (2007) Emerging principles of regulatory evolution. Proc Natl Acad Sci U S A 104 Suppl 1: 8605-8612.

216. Stone JR, Wray GA. (2001) Rapid evolution of *cis*-regulatory sequences via local point mutations. Mol Biol Evol 18(9): 1764-1770.

217. Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U. (2006) Plasticity of the *cis*-regulatory input function of a gene. PLoS Biol 4(4): e45.

218. Isalan M, Lemerle C, Michalodimitrakis K, Horn C, Beltrao P, Raineri E, Garriga-Canut M, Serrano L. (2008) Evolvability and hierarchy in rewired bacterial gene networks. Nature 452(7189): 840-845.

219. Perez JC, Groisman EA. (2009) Evolution of transcriptional regulatory circuits in bacteria. Cell 138(2): 233-244.

220. Galan JE, Wolf-Watz H. (2006) Protein delivery into eukaryotic cells by type III secretion machines. Nature 444(7119): 567-573.

221. Fass E, Groisman EA. (2009) Control of *Salmonella* pathogenicity island-2 gene expression. Curr Opin Microbiol 12(2): 199-204.

222. Brown NF, Vallance BA, Coombes BK, Valdez Y, Coburn BA, Finlay BB. (2005) *Salmonella* pathogenicity island 2 is expressed prior to penetrating the intestine. PLoS Pathog 1(3): e32.

223. Hokamp K, Roche FM, Acab M, Rousseau ME, Kuo B, Goode D, Aeschliman D, Bryan J, Babiuk LA, Hancock RE, Brinkman FS. (2004) ArrayPipe: A flexible processing pipeline for microarray data. Nucleic Acids Res 32(Web Server issue): W457-9.

224. Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. (2001) The COG database: New developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res 29(1): 22-28.

225. Dobrindt U, Hochhut B, Hentschel U, Hacker J. (2004) Genomic islands in pathogenic and environmental microorganisms. Nat Rev Microbiol 2(5): 414-424.

226. Meng X, Brodsky MH, Wolfe SA. (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. Nat Biotechnol 23(8): 988-994.

227. Zwir I, Shin D, Kato A, Nishino K, Latifi T, Solomon F, Hare JM, Huang H, Groisman EA. (2005) Dissecting the PhoP regulatory network of *Escherichia coli* and *Salmonella enterica*. Proc Natl Acad Sci U S A 102(8): 2862-2867.

228. Bailey TL, Williams N, Misleh C, Li WW. (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34(Web Server issue): W369-73.

229. Liu XS, Brutlag DL, Liu JS. (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol 20(8): 835-839.

230. Toh H, Weiss BL, Perkin SA, Yamashita A, Oshima K, Hattori M, Aksoy S. (2006) Massive genome erosion and functional adaptations provide insights into the symbiotic lifestyle of *Sodalis glossinidius* in the tsetse host. Genome Res 16(2): 149-156.

231. Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P, Moreau Y. (2001) A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. Bioinformatics 17(12): 1113-1122.

232. Thijs G, Marchal K, Lescot M, Rombauts S, De Moor B, Rouze P, Moreau Y. (2002) A Gibbs sampling method to detect overrepresented motifs in the upstream regions of coexpressed genes. J Comput Biol 9(2): 447-464.

233. Coombes BK. (2009) Type III secretion systems in symbiotic adaptation of pathogenic and non-pathogenic bacteria. Trends Microbiol 17(3): 89-94.

234. Dale C, Plague GR, Wang B, Ochman H, Moran NA. (2002) Type III secretion systems and the evolution of mutualistic endosymbiosis. Proc Natl Acad Sci U S A 99(19): 12397-12402.

235. Shimada T, Ishihama A, Busby SJ, Grainger DC. (2008) The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. Nucleic Acids Res 36(12): 3950-3955.

236. Cho BK, Knight EM, Barrett CL, Palsson BO. (2008) Genome-wide analysis of fis binding in *Escherichia coli* indicates a causative role for A-/AT-tracts. Genome Res 18(6): 900-910.

237. MacArthur S, Brookfield JF. (2004) Expected rates and modes of evolution of enhancer sequences. Mol Biol Evol 21(6): 1064-1073.

238. Salcedo SP, Holden DW. (2003) SseG, a virulence protein that targets *Salmonella* to the golgi network. EMBO J 22(19): 5003-5014.

239. Deiwick J, Salcedo SP, Boucrot E, Gilliland SM, Henry T, Petermann N, Waterman SR, Gorvel JP, Holden DW, Meresse S. (2006) The translocated *Salmonella* effector proteins SseF and SseG interact and are required to establish an intracellular replication niche. Infect Immun 74(12): 6965-6972.

240. Shin D, Lee EJ, Huang H, Groisman EA. (2006) A positive feedback loop promotes transcription surge that jump-starts *Salmonella* virulence circuit. Science 314(5805): 1607-1609.

241. Langille MG, Brinkman FS. (2009) IslandViewer: An integrated interface for computational identification and visualization of genomic islands. Bioinformatics 25(5): 664-665.

242. Crooks GE, Hon G, Chandonia JM, Brenner SE. (2004) WebLogo: A sequence logo generator. Genome Res 14(6): 1188-1190.

243. Uzzau S, Figueroa-Bossi N, Rubino S, Bossi L. (2001) Epitope tagging of chromosomal genes in *Salmonella*. Proc Natl Acad Sci U S A 98(26): 15264-15269.

244. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. (2009) Circos: An information aesthetic for comparative genomics. Genome Res 19(9): 1639-1645.

245. Wray C, Sojka WJ. (1978) Experimental *Salmonella typhimurium* infection in calves. Res Vet Sci 25(2): 139-143.

246. Bjarnason J, Southward CM, Surette MG. (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar Typhimurium by high-throughput screening of a random promoter library. J Bacteriol 185(16): 4973-4982.

247. Mills DM, Bajaj V, Lee CA. (1995) A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. Mol Microbiol 15(4): 749-759.

248. Li J, Ochman H, Groisman EA, Boyd EF, Solomon F, Nelson K, Selander RK. (1995) Relationship between evolutionary rate and cellular location among the

Inv/Spa invasion proteins of *Salmonella enterica*. Proc Natl Acad Sci U S A 92(16): 7252-7256.

249. Canchaya C, Fournous G, Brussow H. (2004) The impact of prophages on bacterial chromosomes. Mol Microbiol 53(1): 9-18.

250. Fang FC, DeGroote MA, Foster JW, Baumler AJ, Ochsner U, Testerman T, Bearson S, Giard JC, Xu Y, Campbell G, Laessig T. (1999) Virulent *Salmonella typhimurium* has two periplasmic cu, zn-superoxide dismutases. Proc Natl Acad Sci U S A 96(13): 7502-7507.

251. Casjens S. (2003) Prophages and bacterial genomics: What have we learned so far? Mol Microbiol 49(2): 277-300.

252. Tomljenovic-Berube AM, Mulder DT, Whiteside MD, Brinkman FS, Coombes BK. (2010) Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB two-component system. PLoS Genet 6(3): e1000875.

253. 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(12): 6640-6645.

254. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. (1989) Engineering hybrid genes without the use of restriction enzymes: Gene splicing by overlap extension. Gene 77(1): 61-68.

255. Xia XQ, McClelland M, Porwollik S, Song W, Cong X, Wang Y. (2009) WebArrayDB: Cross-platform microarray data analysis and public data repository. Bioinformatics 25(18): 2425-2429.

256. Smyth GK. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3: Article3.

257. Kundinger MM, Zabala-Díaz IB, Chalova IV, Kim W-, Moore RW, Ricke SC. (2007) Characterization of *rsmC* as a potential reference gene for *Salmonella* Typhimurium gene expression during growth in spent media. Sens & Instrumen Food Qual 1(3): 99-103.

258. Grayhack EJ, Yang XJ, Lau LF, Roberts JW. (1985) Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. Cell 42(1): 259-269.

259. Hawley DK, McClure WR. (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11(8): 2237-2255.

260. Nickels BE, Roberts CW, Roberts JW, Hochschild A. (2006) RNA-mediated destabilization of the sigma(70) region 4/beta flap interaction facilitates engagement of RNA polymerase by the Q antiterminator. Mol Cell 24(3): 457-468.

261. Tominaga A, Ikemizu S, Enomoto M. (1991) Site-specific recombinase genes in three *Shigella* subgroups and nucleotide sequences of a *pinB* gene and an invertible B segment from *Shigella boydii*. J Bacteriol 173(13): 4079-4087.

262. van de Putte P, Goosen N. (1992) DNA inversions in phages and bacteria. Trends Genet 8(12): 457-462.

263. Nguyen HA, Tomita T, Hirota M, Kaneko J, Hayashi T, Kamio Y. (2001) DNA inversion in the tail fiber gene alters the host range specificity of carotovoricin Er, a phage-tail-like bacteriocin of phytopathogenic *Erwinia carotovora* subsp. *carotovora* Er. J Bacteriol 183(21): 6274-6281.

264. Veesler D, Cambillau C. (2011) A common evolutionary origin for tailed-bacteriophage functional modules and bacterial machineries. Microbiol Mol Biol Rev 75(3): 423-33, first page of table of contents.

265. Nakayama K, Takashima K, Ishihara H, Shinomiya T, Kageyama M, Kanaya S, Ohnishi M, Murata T, Mori H, Hayashi T. (2000) The R-type pyocin of *Pseudomonas aeruginosa* is related to P2 phage, and the F-type is related to lambda phage. Mol Microbiol 38(2): 213-231.

266. Garaizar J, Porwollik S, Echeita A, Rementeria A, Herrera S, Wong RM, Frye J, Usera MA, McClelland M. (2002) DNA microarray-based typing of an atypical monophasic *Salmonella enterica* serovar. J Clin Microbiol 40(6): 2074-2078.

267. Lawrence JG, Hendrix RW, Casjens S. (2001) Where are the pseudogenes in bacterial genomes? Trends Microbiol 9(11): 535-540.

268. Tyler JS, Mills MJ, Friedman DI. (2004) The operator and early promoter region of the Shiga toxin type 2-encoding bacteriophage 933W and control of toxin expression. J Bacteriol 186(22): 7670-7679.

269. Steele-Mortimer O, Meresse S, Gorvel JP, Toh BH, Finlay BB. (1999) Biogenesis of *Salmonella typhimurium*-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. Cell Microbiol 1(1): 33-49.

270. Wang RF, Kushner SR. (1991) Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene 100: 195-199.

271. Hansen-Wester I, Chakravortty D, Hensel M. (2004) Functional transfer of *Salmonella* pathogenicity island 2 to *Salmonella bongori* and *Escherichia coli*. Infect Immun 72(5): 2879-2888.

272. Hansen-Wester I, Stecher B, Hensel M. (2002) Analyses of the evolutionary distribution of *Salmonella* translocated effectors. Infect Immun 70(3): 1619-1622.

273. Daubin V, Lerat E, Perriere G. (2003) The source of laterally transferred genes in bacterial genomes. Genome Biol 4(9): R57.

274. Banos RC, Vivero A, Aznar S, Garcia J, Pons M, Madrid C, Juarez A. (2009) Differential regulation of horizontally acquired and core genome genes by the bacterial modulator H-NS. PLoS Genet 5(6): e1000513.

275. Groisman EA. (2001) The pleiotropic two-component regulatory system PhoP-PhoQ. J Bacteriol 183(6): 1835-1842.

276. Harari O, Park SY, Huang H, Groisman EA, Zwir I. (2010) Defining the plasticity of transcription factor binding sites by deconstructing DNA consensus sequences: The PhoP-binding sites among gamma/enterobacteria. PLoS Comput Biol 6(7): e1000862.

277. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, van den Boom D, Cantor CR, Palsson BO. (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38(12): 1406-1412.

278. Conrad TM, Lewis NE, Palsson BO. (2011) Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol 7: 509.

279. Portnoy VA, Bezdan D, Zengler K. (2011) Adaptive laboratory evolution--harnessing the power of biology for metabolic engineering. Curr Opin Biotechnol 22(4): 590-594.

280. Lee DH, Palsson BO. (2010) Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. Appl Environ Microbiol 76(13): 4158-4168.

281. Conrad TM, Frazier M, Joyce AR, Cho BK, Knight EM, Lewis NE, Landick R, Palsson BO. (2010) RNA polymerase mutants found through adaptive evolution

reprogram *Escherichia coli* for optimal growth in minimal media. Proc Natl Acad Sci U S A 107(47): 20500-20505.

282. Charusanti P, Fong NL, Nagarajan H, Pereira AR, Li HJ, Abate EA, Su Y, Gerwick WH, Palsson BO. (2012) Exploiting adaptive laboratory evolution of *Streptomyces clavuligerus* for antibiotic discovery and overproduction. PLoS One 7(3): e33727.

283. Ensminger AW, Yassin Y, Miron A, Isberg RR. (2012) Experimental evolution of *Legionella pneumophila* in mouse macrophages leads to strains with altered determinants of environmental survival. PLoS Pathog 8(5): e1002731.

284. Yang L, Jelsbak L, Marvig RL, Damkiaer S, Workman CT, Rau MH, Hansen SK, Folkesson A, Johansen HK, Ciofu O, Hoiby N, Sommer MO, Molin S. (2011) Evolutionary dynamics of bacteria in a human host environment. Proc Natl Acad Sci U S A 108(18): 7481-7486.

285. Eriksson S, Lucchini S, Thompson A, Rhen M, Hinton JC. (2003) Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. Mol Microbiol 47(1): 103-118.

286. McLaughlin LM, Govoni GR, Gerke C, Gopinath S, Peng K, Laidlaw G, Chien YH, Jeong HW, Li Z, Brown MD, Sacks DB, Monack D. (2009) The *Salmonella* SPI2 effector SseI mediates long-term systemic infection by modulating host cell migration. PLoS Pathog 5(11): e1000671.

287. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, Wain J, Churcher C, Mungall KL, Bentley SD, Holden MT, Sebaihia M, Baker S, Basham D, Brooks K, Chillingworth T, Connerton P, Cronin A, Davis P, Davies RM, Dowd L, White N, Farrar J, Feltwell T, Hamlin N, Haque A, Hien TT, Holroyd S, Jagels K, Krogh A, Larsen TS, Leather S, Moule S, O'Gaora P, Parry C, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature 413(6858): 848-852.

288. Dagan T, Blekhman R, Graur D. (2006) The "domino theory" of gene death: Gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. Mol Biol Evol 23(2): 310-316.

289. Forest CG, Ferraro E, Sabbagh SC, Daigle F. (2010) Intracellular survival of *Salmonella enterica* serovar Typhi in human macrophages is independent of *Salmonella* pathogenicity island (SPI)-2. Microbiology 156(Pt 12): 3689-3698.

290. Libby SJ, Brehm MA, Greiner DL, Shultz LD, McClelland M, Smith KD, Cookson BT, Karlinsey JE, Kinkel TL, Porwollik S, Canals R, Cummings LA, Fang FC. (2010) Humanized nonobese diabetic-scid IL2rgammanull mice are susceptible to lethal *Salmonella* Typhi infection. Proc Natl Acad Sci U S A 107(35): 15589-15594.

291. Song J, Willinger T, Rongvaux A, Eynon EE, Stevens S, Manz MG, Flavell RA, Galan JE. (2010) A mouse model for the human pathogen *Salmonella* Typhi. Cell Host Microbe 8(4): 369-376.

292. Mian MF, Pek EA, Chenoweth MJ, Coombes BK, Ashkar AA. (2011) Humanized mice for *Salmonella typhi* infection: New tools for an old problem. Virulence 2(3): 248-252.

293. Dale C, Young SA, Haydon DT, Welburn SC. (2001) The insect endosymbiont *Sodalis glossinidius* utilizes a type III secretion system for cell invasion. Proc Natl Acad Sci U S A 98(4): 1883-1888.

294. Dale C, Jones T, Pontes M. (2005) Degenerative evolution and functional diversification of type-III secretion systems in the insect endosymbiont *Sodalis glossinidius*. Mol Biol Evol 22(3): 758-766.

295. Barnard A, Wolfe A, Busby S. (2004) Regulation at complex bacterial promoters: How bacteria use different promoter organizations to produce different regulatory outcomes. Curr Opin Microbiol 7(2): 102-108.

296. van Helden J, Rios AF, Collado-Vides J. (2000) Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. Nucleic Acids Res 28(8): 1808-1818.

297. Ehrbar K, Hardt WD. (2005) Bacteriophage-encoded type III effectors in *Salmonella enterica* subspecies 1 serovar Typhimurium. Infect Genet Evol 5(1): 1-9.

298. Thomson N, Baker S, Pickard D, Fookes M, Anjum M, Hamlin N, Wain J, House D, Bhutta Z, Chan K, Falkow S, Parkhill J, Woodward M, Ivens A, Dougan G. (2004) The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. J Mol Biol 339(2): 279-300.

299. Figueroa-Bossi N, Bossi L. (1999) Inducible prophages contribute to *Salmonella* virulence in mice. Mol Microbiol 33(1): 167-176.

300. Stender S, Friebel A, Linder S, Rohde M, Mirold S, Hardt WD. (2000) Identification of SopE2 from *Salmonella typhimurium*, a conserved guanine

nucleotide exchange factor for Cdc42 of the host cell. Mol Microbiol 36(6): 1206-1221.

301. Koskiniemi S, Sun S, Berg OG, Andersson DI. (2012) Selection-driven gene loss in bacteria. PLoS Genet 8(6): e1002787.

302. Xu X, McAteer SP, Tree JJ, Shaw DJ, Wolfson EB, Beatson SA, Roe AJ, Allison LJ, Chase-Topping ME, Mahajan A, Tozzoli R, Woolhouse ME, Morabito S, Gally DL. (2012) Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli*. PLoS Pathog 8(5): e1002672.

303. Rodas PI, Trombert AN, Mora GC. (2011) A holin remnant protein encoded by STY1365 is involved in envelope stability of *Salmonella enterica* serovar Typhi. FEMS Microbiol Lett 321(1): 58-66.