Identification of Novel Protein Substrates and Chemical Inhibitors of the T3SA in *Shigella*

Navoun Silué

A thesis submitted to the University of Ottawa
In partial Fulfillment of the requirements for the
Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Sciences
University of Ottawa
April 2023

Abstract

Identification of Novel Protein Substrates and Chemical Inhibitors of the T3SA in *Shigella*

Enteropathogenic bacteria, such as *Shigella* and *Salmonella,* are associated with diarrheal diseases, which remain a significant cause of infant mortality worldwide. The secretion of protein effectors by the type III secretion apparatus (T3SA) is used by these pathogens to invade human cells and modulate host cell functions. First, we used RNA-Seq to analyze the differential transcriptome of *Shigella flexneri* when the T3SA is active or inactive. This allowed us to identify two uncharacterized genes that were temporarily named gem1 and gem3 and whose expression was regulated by MxiE and IpgC as other late substrates of the T3SA. Finally, we pursued the characterization of gem1 and gem3 at the protein level and renamed them *icaT* and *icaR*, respectively, when we found their protein products were secreted by the T3SA. Furthermore, we find homologs of *icaT* and *icaR* with a conserved MxiE box in several *E. coli* phylogroups. We also demonstrated that these homologous genes could be reactivated when both MxiE and IpgC were introduced in these strains. This discovery paved a new perspective on the evolution of pathogenesis into the *E. coli* lineage as both commensal and pathogenic strains harbored these genes.

Treating infections caused by *Enterobacteriaceae* is becoming more challenging due to growing antibiotic resistance and no vaccines are widely available. Accordingly, the World Health Organization (WHO) recognized that we entered the « post-antibiotic era, » where new antibiotics or antivirulence drugs are urgently needed, including for *Shigella*. The T3SA is an attractive target for antivirulence drugs, which may become alternative to classical antibiotics. Through screening 3.000 compounds, we found two novel inhibitors of the T3SA. Our data suggested that one of these candidate inhibitors, a dipyridyl-containing compound, reduces the virulence of *Shigella* at the transcriptional level. Indeed, the virulence inhibition occurs via the repression of the transcriptional activator VirB by the small chromosomal RNA RyhB, which is upregulated by this compound through an unknown mechanism involving the pyridyl groups. The repression of VirB induced by this molecule reduce the expression of several genes encoding parts of the T3SA. In comparison, the second compound is a quinone that seems to affect the assembly of the T3SA.

Résumé (French)

Les bactéries entéropathogènes, telles que *Shigella* et *Salmonella* sont associées aux maladies diarrhéiques, qui restent une cause importante de mortalité infantile dans le monde. La sécrétion d'effecteurs protéiques par l'appareil de sécrétion de type trois (T3SA) est utilisée par ces agents pathogènes pour envahir les cellules humaines et moduler leur fonction. Nous avons d'abord utilisé le RNA-Seq pour analyser le transcriptome différentiel de *Shigella flexneri* lorsque le T3SA est actif ou inactif. Cela nous a permis d'identifier deux gènes non caractérisés qui ont été temporairement nommés gem1 et gem3 et dont l'expression était régulée par MxiE et IpgC comme d'autres substrats tardifs du T3SA. Enfin, nous avons poursuivi la caractérisation de gem1 et gem3 au niveau des protéines et les avons renommées respectivement *icaT* et *icaR* lorsque nous avons constaté que leurs produits protéiques étaient sécrétés par le T3SA. De plus, nous avons trouvé des homologues de *icaT* et *icaR* avec une boîte MxiE conservée dans plusieurs phylogroupes d'*E. coli*. Nous avons également démontré que ces gènes homologues pouvaient être réactivés lorsque MxiE et IpgC étaient introduits dans ces souches. Cette découverte a ouvert une nouvelle perspective sur l'évolution de la pathogenèse dans la lignée *E. coli*, car les souches commensales et pathogènes hébergeaient aussi ces gènes.

Le traitement des infections causées par les entérobactéries devient de plus en plus difficile en raison de la résistance croissante aux antibiotiques et aucun vaccin n'est largement disponible. En conséquence, l'Organisation mondiale de la santé (OMS) a reconnu que nous entrions dans « l'ère post-antibiotique », où de nouveaux antibiotiques ou des drogues antivirulents sont nécessaires de toute urgence, y compris pour *Shigella*. Le T3SA est une cible intéressante pour les antivirulences, qui pourraient devenir une alternative aux antibiotiques classiques. Grâce au criblage de 3 000 composés, nous avons trouvé deux nouveaux inhibiteurs du T3SA. Nos données suggèrent que le premier de ces composés, qui comporte deux groupements pyridyl, réduit la virulence de *Shigella* au niveau transcriptionnel. En effet, l'inhibition de la virulence se produit via la répression de l'activateur transcriptionnel VirB par le petit ARN chromosomique RyhB dont l'expression est augmentée par ce composé par un mechanism inconnu impliquant les groupements pyridyls. La répression de VirB induite par cette molécule réduit l'expression de plusieurs gènes codant pour des composants du T3SA. En comparaison, le second composé est une quinone qui semble interférée avec l'assemblage du T3SA.

Dedication

I would like to dedicate this work to my fantastic wife Laurentine and my wonderful daughter Anaya who supported me throughout my entire graduate studies.

Acknowledgements

I would like to acknowledge all the Department of Chemistry and biomolecular sciences for their support throughout my graduate career and their innumerable assistance in this work. Firstly, we want to express our sincere gratitude to my advisor François-Xavier Campbell-Valois, who has been patient, encouraging, and continuously supporting and guiding me through this project. I would also like to thank my committee members, John Pezacki, Corrie Dacosta, and Daniel Figeys, for their insightful comments and encouragement. My thanks also go to all members of the Host-Microbe-Interaction Laboratory (HMI). Finally, I thank my brother Fanignoro Silué for his wise advice.

Table of contents

List of figures

List of tables

List of abbreviations

BMP: 1,5-Bis(3-Methoxyphenyl)-1,4-pentadien-3-one.
BPY: 2,2′-Bipyridine or 2,2′-Dipyridine or 2,2′-Dipyridyl.
CR: Congo red
DBA: DiBenzyldeneAcetone or 1,5-Diphenyl-1,4-pentadien-3-one or trans, trans-1,5-Diphenyl-1,4-pentadien-3-one).
ddPCR: droplet digital polymerase chain reaction.

DF: Deferoxamine mesylate salt.
DM or DMSO: Dimethyl sulfoxide.
DP: Deferiprone
DPK: Di(2-pyridyl) Ketone or 2,2'-Carbonyldipyridine or 2,2'-Dipyridyl Ketone
DPP: 1,5-Di(1h-Pyrrol-2-yl)-1,4-Pentadien-3-one
DS: Deferasirox
DTP: 1,5-Di-thiophen-2-yl-Penta-1,4-dien-3-one
f-TSAR: fluorescent transcription-based activation reporter of the T3SA

FC: on/off fold change.

FDA: Food Drug Administration
FPP: 3-(5-fluoropyridin-3-yl)-1-(pyridin-3-yl) prop-2-en-1-one.
GFP: green fluorescent protein.

icaR: Invasion chromosome Antigen with homology for a Transcription Regulator
icaT: Invasion chromosome Antigen with homology for a Transposase
ipaHs: *ipaH* genes.

IS: insertion sequence.

l-TSAR: luminescent transcription-based activation reporter of the T3SA

MOI: multiplicity of infection

NCI: National Cancer Institute
Notes orf (lowercase): Refers to some hypothetical open reading frames in pINV of M90T; usually identified by digits corresponding to the approximate predicted transcription start site.

ORF: open reading frame.

pINV: invasion plasmid
PMH: 3-[(E)-[(2E)-2-[(pyridin-3-yl) methylidene] hydrazin-1-ylidene] methyl] pyridine.
PPP: 3-(pyridin-3-yl)-N-[3-(pyridin-3-yl) prop-2-en-1-yl] prop-2-enamide.
pWR100: the large virulence plasmid of *Shigella flexneri* strain M90T.

qRT-ddPCR: Real-time quantitative Reverse transcriptase-droplet digital polymerase chain reaction

QUE: Quercetin.

RNA-Seq: ribonucleic acid sequencing.

rRNA: ribosomal RNA; RT-PCR: reverse transcriptase polymerase chain reaction.

T3SA: Type III Secretion Apparatus.

T3SS: Type III Secretion System.

TPY: Terpyridine (α,α′,α″-Tripyridyl, 2,6-Di(2-pyridyl)pyridine
TSA: Tryptic Soy Agar

TSB: Tryptic Soy Broth

WB: western blot.

Chapter 1 : The Regulation of the Expression of the T3SS in *Shigella*

1.1 Preface

Section 1.2 to 1.5 of this chapter contains edited excerpts from the review published in Microorganism (Bajunaid et al., 2020) . All authors contributed equally by writing one section of this review. Therefore, we adapted the introduction (written by François-Xavier Campbell-Valois) and used the entire section 3 of this paper, written by me and reviewed by François-Xavier Campbell-Valois. Finally, we completed this chapter with section 1.6. As authors in this journal, we have the copyright to use this article, which may be reused and quoted, provided that the original published version is cited.

1.2 Introduction

Shigella spp. are Gram-negative bacterial pathogens infecting the large intestine of humans and responsible for hundreds of millions of infections every year (Lampel et al., 2018; Kotloff et al., 2018). *Shigella* spp. are transmitted through a fecal-oral route that causes acute bacillary dysentery; they are highly infectious with a low infective dose of 10-500 bacteria sufficient to cause disease in healthy human volunteers (Kang et al., 2018). Shigellosis is associated with the consumption of contaminated food and water, or increasingly with sexual transmission among men who have sex with men (Richardson et al., 2021; Charles et al., 2022). The symptoms range from mild, self-limiting diarrhea and abdominal cramps to severe colitis, bacteremia, intestinal perforation, and peritonitis but the infection can also be asymptomatic (Charles et al., 2022). *Shigella* spp. consist in four subgroups, such as *Shigella dysenteriae* (group A), *S. flexneri* (group B), *S. boydii* (group C) and *S. sonnei* (group D) (Barry et al., 2013). Shigellosis mainly affects

children under five years old and the elderly by provoking bloody diarrhea, which can be lethal. Indeed, the global burden of *Shigella* infection was estimated at around 269 million cases and over 212,000 deaths in 2016 (Khalil et al., 2018). This study also reported that children under five years are the most frequently afflicted with 74 million cases and 63,000 deaths. In Canada, around 880 cases of shigellosis are recorded annually (Canada Public Health Agency, 2020; NCCID, 2022). Recently, a *Shigella* outbreak was reported in the Edmonton area with 211 cases, of which 144 required hospitalizations, but no deaths have been reported connected to this outbreak (Alberta Health Services, 2023). The Center for Diseases Control and Prevention in the USA (CDC) observed an increase in extensively antimicrobial resistant (XDR) strains of *Shigella* among adults in the United States from 0% in 2015 to 5% in 2022 (CDC, 2023). The predominant strain was *Shigella sonnei* (66%), previously reported in Europe as extensively resistant to antibiotics (Trivett, 2022), and followed by *Shigella flexneri* (34%) (CDC, 2023). Antibiotics remain the standard treatment for shigellosis, but the widespread prevalence of extensively resistant strains coupled with the dearth of new antibiotics restricts these therapeutic options (von Seidlein et al., 2006; Kotloff et al., 1999; Raso et al., 2023). Generally, antibiotic resistance occurs by the use of efflux pumps to export antibiotics in the extracellular environment, the decrease of the membrane permeability to prevent their penetration, the mutation of their biomolecular target to diminish their binding affinity, or by expressing enzymes able to alter and inactivate the antibiotics (Ranjbar and Farahani, 2019; Raso et al., 2023). As the discovery of novel antibiotics is scarce, the development of vaccine or antivirulence drugs against *Shigella* and other enterobacteria becomes more crucial (Barry et al., 2013; Charles et al., 2022; Lefèvre et al., 2023; Raso et al., 2023). As of early 2023, no vaccines against *Shigella* have been approved by the World Health Organization (WHO, 2022). Protecting against the multiple serogroups of *Shigella* and the absence of an

appropriate animal model for shigellosis makes it challenging to develop cross-protective candidates (Sansonetti, 2006; Barry et al., 2013; Kotloff et al., 2017; Raso et al., 2023). The alternatives to antibiotics and vaccines are phage therapy, natural products derived from bacteria or plant (Ranjbar and Farahani, 2019), or antivirulent compounds to disarm the pathogenicity of bacteria.

Shigella spp. mainly differ from commensal *Escherichia coli* by the presence of a large virulence plasmid (VP) that encodes the various components of the *Shigella* type III secretion system (T3SS). This comprises the various proteins required to assemble the syringe-like type III secretion apparatus (T3SA), also known as the injectisome, as well as those acting as transcription regulators, substrates, and their cognate chaperones. Some of the T3SA substrates are effectors that rewire human cell biochemical processes through their enzymatic or binding activities. The effectors allow the invasion of the cytosol of epithelial cells by *Shigella*. This entry process is akin to phagocytosis. *Shigella* then uses actin-based motility and effectors to invade the cytosol of neighboring cells in a process named cell-to-cell spread that shares some similarities with the entry (Campbell-Valois and Pontier, 2016). Among these two components, the T3SA is the most important for virulence and is used by many other gram-negative bacteria. Thus, it is an attractive drug target, which has been tested in several small molecule screens (Muschiol et al., 2006; Dey et al., 2017).

Iron is an essential nutrient for most microbes, but free iron is scarce in many environments, particularly in mammalian hosts which restrict iron to prevent infection (Ratledge and Dover, 2000; Mey et al., 2021). Under these circumstances, bacteria exploit strategies to access extracellular and intracellular iron stores while avoiding excess exposure. Indeed, disproportionate concentrations of free intracellular iron can generate hydroxyl radical ($OH^{\cdot}$) products via the

Fenton reaction in the presence of oxygen. Therefore, the tight control of this free iron pool by bacteria is critical to avoid reaching toxic iron concentrations. Indeed, the iron acquisition is the primary determinant of the free iron pool and directly impacts the capacity of a microorganism to maintain iron homeostasis in diverse settings, including during the infection of its host. To circumvent iron scarcity, several pathogens synthesize small iron chelators called siderophores, which extract the iron from the host sources (Ratledge and Dover, 2000; Massé et al., 2003; Payne et al., 2006). In this chapter, we will also discuss the regulation of the expression of T3SS genes by the iron homeostasis pathway in *Shigella*.

1.3 Overview of the regulatory cascade of the T3SS

The regulation of the production of proteins encoded by T3SS genes intervenes mostly at the transcriptional level. The T3SS regulatory cascade is composed of a repressor named the histone-like nucleoid structure protein (H-NS), and of three transcription activators named VirF (AraC-family), VirB (ParB family), and MxiE (AraC family) (Figure1.1A) (Dorman and Dorman, 2018). The first three players coordinate the transcriptional response to the rise of temperature occurring upon *Shigella's* transition from an external environment to its host. The interaction between these factors and DNA culminates with the production of VirB, which activates the expression of most T3SS genes, hence ensuring that the T3SA is assembled in the *Shigella* membrane only when it is inside its host. The last transcription activator MxiE senses the presence of host cells using T3SA activity as a proxy. Indeed, MxiE is inhibited at the posttranslational level when T3SAs are inactive. This inhibition is released when T3SAs are active, thereby allowing MxiE to upregulate the expression of genes encoding late substrates B. Finally, a handful of genes called late substrates A/B are dually controlled by VirB- and MxiE-dependent promoters (Parsot,

2009; Silué et al., 2020) (Figure1.1B). Below, we detail the action of these different transcription factors in the regulation of the main virulence genes.

1.4 Master regulators of the T3SS in *Shigella*: H-NS, VirF, VirB

The repressor of the T3SS regulatory cascade in *Shigella* is the histone-like nucleoid structure protein (H-NS). It forms homomers in isolation but may also heteromerize with two related proteins named StpA (prevalent in *Shigella* spp.) and Sfh (*S. flexneri* str. 2457T) (Deighan et al., 2003; Picker and Wing, 2016). Interestingly, StpA and Sfh can complement the loss of H-NS by, for example, ensuring the repression of T3SS genes, thus suggesting that these three paralogs share this function (Beloin et al., 2003; Picker and Wing, 2016).

H-NS binds AT-rich DNA through the insertion of a conserved three-residue motif into the minor groove according to molecular dynamics simulations (Riccardi et al., 2019). In response to changing physicochemical conditions, H-NS can toggle between a stiffening or a bridging DNA binding mode (Liu et al., 2010). In the filament mode, several H-NS protomers are associated to a single DNA segment, while in the stiffening mode, they bundle two or more DNA segments together (Ali et al., 2012). Both binding modes inhibit transcription initiation by impeding the access of transcriptional regulators and the RNA polymerase to promoters. The binding of H-NS to DNA can also hinder transcription elongation, which can favor Rho-dependent termination (Kotlajich et al., 2015). To our knowledge, this phenomenon, however, has not been assessed for *Shigella* T3SS genes.

In *Shigella*, T3SS genes have 60% to 70% AT content (Buchrieser et al., 2000). H-NS thus has a countless number of potential binding sites spread over the VP. Those located in the entry region have been particularly scrutinized due to their key role in the regulation of virulence. Below

32 °C, H-NS binding is favored, thus imposing three locks on the regulatory cascade of *Shigella*. The two most upstream result in the silencing of *virF* and *virB* that encode two transcription activators; the third is downstream and results in the silencing of T3SA parts and several substrates (Tobe et al., 1993; Falconi et al., 1998; Beloin and Dorman, 2003; Martino et al., 2016; Picker and Wing, 2016). The main consequence of the switch to human body temperature is to open these three locks[1]. For example, the conformation of the promoter of *virF* favors H-NS binding and silencing of *virF* below 32 °C (Falconi et al., 1998; Martino et al., 2016).By contrast, at the human body temperature (or >32 °C), the conformation of the promoter precludes H-NS binding, hence favoring transcription initiation and, ultimately, the production of the VirF protein (Falconi et al., 1998; Martino et al., 2016). *virF* integrates other environmental cues to ensure that the virulence genes are expressed to optimal level when the right conditions are met. For example, *virF* is silenced at acidic pH due to the inactivation of the CpxAR two-component system. In neutral conditions, CpxA-activated CpxR upregulates the expression of *virF*.

[1] Updated in section 1.6 of this chapter

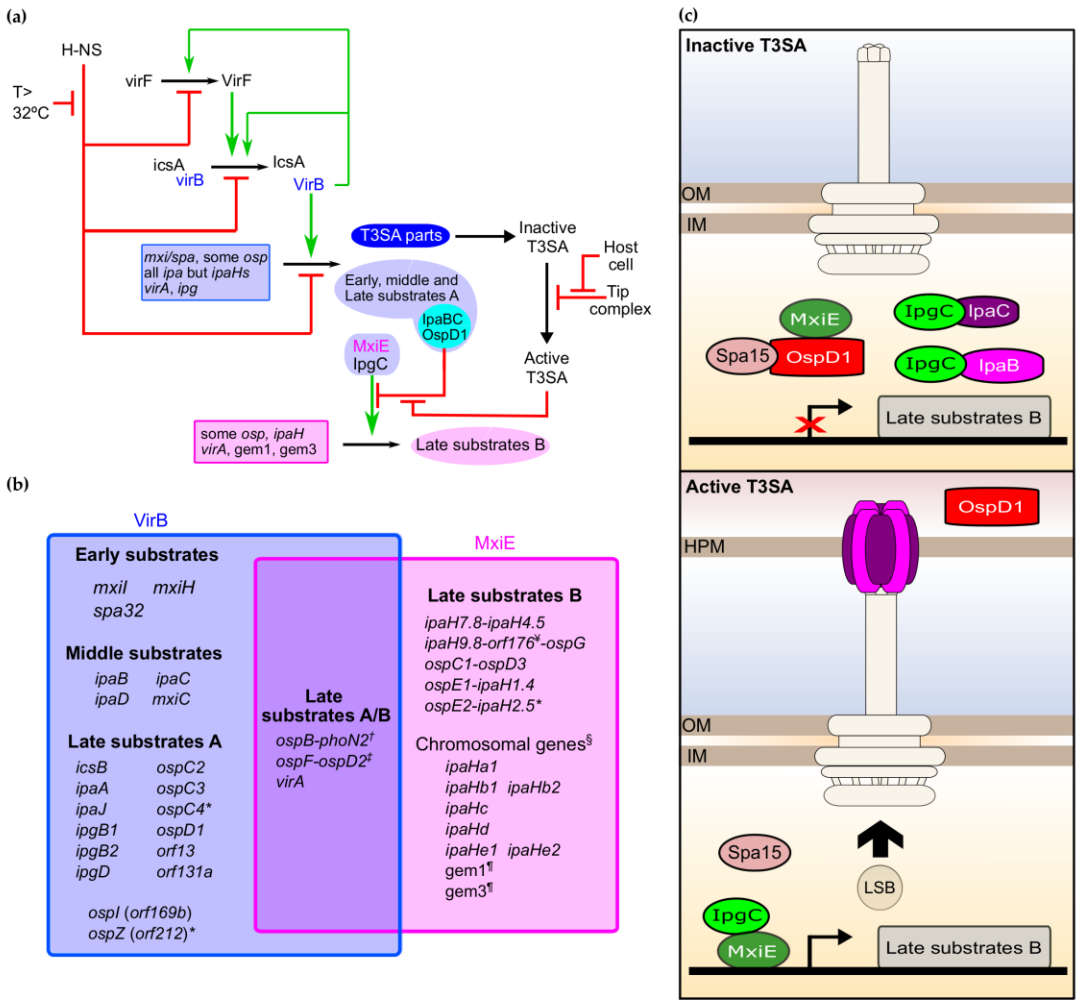


Figure 1. 1. The transcription regulatory cascade of the T3SA. (**A**) Key transcriptional regulatory cascade of the type III secretion system (T3SS) in *Shigella*. (**B**) The VirB and MxiE regulons; this figure is inspired from a prior version (Parsot, 2009), and updated to take into account the most recent findings **(Pinaud et al., 2017a; Silué et al., 2020)**. * These are likely pseudogenes in strain M90T; † this gene belongs to the VirB and MxiE regulon but its product is not a T3SA substrate; ‡ this gene is weakly expressed and a new addition to the MxiE regulon; § the chromosomal *ipaH*s are annotated according to (Bongrand et al., 2012); ¶ whether the product of these genes are T3SA substrates is currently unknown[2]; ¥this gene encodes an antitoxin from a TA system that is weakly secreted by the T3SA. Note that other antitoxin coding genes *gmvA* (orf48), orf86, and *mvpA*, which are not part of these two regulons, were also suggested to be T3SA substrates (Pinaud et al., 2017a). (**C**) Heteromers formed by MxiE and IpgC in the inactive and active state of the T3SA and their effect on the transcription of late substrates B.

[2] Renamed *icaT* and *icaR* and confirmed as substrates of T3SA (Chapter 3)

The translation factor EF-P and its regulator PoxA are required for the production of CpxA, and thus play an indirect role in the expression of *virF* (Marman et al., 2014). The antisilencing of *virF* is also favored by nucleoid-associated proteins such as IHF and FIS whose functions are modulated by the physiological state of the cell, most notably in function of the growth phase (Dorman, 2004; Marman et al., 2014; Martino et al., 2016) Furthermore, the production and the function of VirF are respectively regulated at the translational level through posttranscriptional modifications of several tRNAs (Durand et al., 1997, 2000; Martino et al., 2016), and by the production of an N-terminally truncated isoform of VirF, which, in contrast to its full-length counterpart, represses target genes (Di Martino et al., 2016). As the most upstream transcription activator in the virulence cascade, VirF activates the transcription of *icsA*, which encodes a virulence factor essential for cell-to-cell spread, as well as of *virB*, which encodes the second transcription activator (Martino et al., 2016). VirF is a promiscuous DNA binder that tolerates degeneracy within its AT-rich consensus binding site. This property, combined with the high AT-content of T3SS genes, allows the formation of large VirF oligomers bridging discrete binding sites within the same gene (Tran et al., 2011). Interestingly, VirF may be acting through the displacement of H-NS around the +1 of transcription and in the coding sequence of *icsA* (Tran et al., 2011), while in the case of *virB*, it more classically binds to the promoter (Tobe et al., 1993; Martino et al., 2016). On top of this antisilencing effect on transcription, VirF was recently suggested to regulate the expression of *icsA* by directly binding its mRNA and its noncoding antisense regulator RnaG (Giangrossi et al., 2017).

VirB is the key to activating the transcription of T3SS-related genes for which the repression of H-NS function by temperature (>32 °C) is not sufficient. The DNA binding site of VirB is disputed, as independent studies indicated that an eight-nucleotide motif or two inverted repeats thereof separated by a single nucleotide were required for VirB activity (Taniya et al., 2003; Kane

and Dorman, 2012; Karney et al., 2019). As suggested in the most recent study, the number of sites matching to the larger recognition site is commensurate with the number of genes controlled by VirB, thus arguing in favor of its relevance. Alternately, noncontiguous eight-nucleotide motifs, provided they can be brought together in space, might act as nucleation sites for the formation of VirB-DNA complexes. VirB is deemed an antisilencing factor because its capacity to control the expression of its target genes depends on the silencing activity of H-NS on them (Basta et al., 2013; Picker and Wing, 2016; Weatherspoon-Griffin et al., 2018). Furthermore, VirB binding motifs both upstream and downstream, and sometimes at a considerable distance from the +1 of transcription of several T3SS genes, are documented, suggesting that the formation of large VirB complexes nucleating from these sites may help displace H-NS complexes (Beloin and Dorman, 2003; Wing et al., 2004; Turner and Dorman, 2007; Basta et al., 2013; Weatherspoon-Griffin et al., 2018; Karney et al., 2019; McKenna and Wing, 2020). The existence of two positive feedback loops of VirB on *virB* and *virF* was proposed to explain the great change of expression of T3SS genes observed upon shifting to the permissive temperature (Kane and Dorman, 2012) (Figure1.1A). The VirB regulon was determined with DNA microarrays by comparing mRNA abundance in Δ*virB* or WT grown at 30 °C versus WT grown at 37 °C. This study showed that T3SS genes are regulated by VirB, thus supporting its role in mediating the effect of temperature on virulence (Le Gall et al., 2005). Although this study was not pangenomic, it probed almost all T3SS genes on the VP. Based on these findings and the mode of action described above, VirB is hypothesized to directly activate the transcription of T3SA parts, early substrates, middle substrates, and late substrates A (Figure1.1A-B). Besides transcription regulation, a few cases of posttranscriptional regulation of the virulence regulatory cascade have been reported. First, the mRNA-binding protein and posttranscriptional regulator CsrA acts as a positive regulator of *virF*

and *virB* through an unknown mechanism involving the glycolytic pathway (Gore and Payne, 2010). Interestingly, CsrA controls the expression of T3SS genes in other species (Vakulskas et al., 2015). In enteropathogenic *E. coli*, the function of CsrA is even coupled to the activity of the T3SAs (Katsowich et al., 2017). Second, VirB was identified in the phosphotyrosine proteome of *Shigella*; a phosphomimetic mutation at one of the two phosphorylated sites phenocopied *ΔvirB* (Standish et al., 2016). Further work is needed to identify the putative tyrosine kinase responsible for this posttranslational modification and determine its relevance for pathogenesis.

The fumarate and nitrate reduction factor (FNR), which is a transcriptional activator important for bacterial adaptation to anaerobic conditions (Mettert and Kiley, 2018), is also known to act as a transcriptional repressor through hindering of RNA polymerase recruitment. It usually binds to DNA as a homodimer, which is destabilized strongly by dioxygen (Mettert and Kiley, 2018). Indeed, in hypoxia, FNR directly represses the expression of *spa32* and *spa33* by binding to their promoter (Marteyn et al., 2010). This led to the formation of partly functional T3SAs characterized by longer needles and poor secretion of middle and late substrates. An RNA-Seq study further described the role of FNR for the adaptation of *Shigella* to anaerobic conditions; the most salient finding was that the absence of oxygen induced an indirect FNR-dependent increase in H-NS repression that contributed to silencing *virF*, *virB*, and their target genes (Figure1.1A) (Vergara-Irigaray et al., 2014), which reinforces the notion that *Shigella's* virulence is stunted in anaerobic conditions. In brief, the control of the key transcription activator VirB by VirF, the presence of positive feedback loops involving VirB, and the three locks imposed by H-NS on the two antisilencers and T3SS genes ensures a 5-10-fold change in the expression level of mxi/spa and *ipaBCD* operons at 37 °C compared to 30 °C (Le Gall et al., 2005). These mechanisms, by ensuring

that the assembly of the T3SAs takes place only in the host, allow the optimal use of cellular resources.

1.5 The MxiE regulon

VirB also directly controls the transcription of *mxiE*, which encodes a transcription activator of the AraC family (Le Gall et al., 2005) (Figure1.1A). The full-length transcript of *mxiE* is the result of transcriptional slippage (Penno et al., 2005); this phenomenon might be used in so far unknown environmental conditions to reduce the production of MxiE. Besides, VirB controls the expression of genes *ipgC*, *ipaB*, *ipaC*, *ospD1*, and *spa15* (Le Gall et al., 2005), which encode MxiE posttranslational regulators. Hence, VirB indirectly controls the expression of late substrates B through the regulation of these genes and others encoding T3SA parts. When T3SAs are inactive, MxiE and IpgC are segregated by the antiactivator OspD1, and its cognate cargos IpaB and IpaC, respectively. The chaperone of OspD1 named Spa15 also acts as co-antiactivator probably through the stabilization of OspD1 in a conformation that favors binding to MxiE (Figure 1. 1C) (Mavris et al., 2002a; Parsot et al., 2005; Demers et al., 1998; Kane et al., 2002). By contrast, when T3SAs are active, IpaB, IpaC, and OspD1 are secreted, allowing the cytosol-resident MxiE and IpgC to oligomerize and activate the transcription of late substrates B (Figure 1.1C). Such direct coupling of transcription to the secretion activity is as elegant as it is rare. The combination of activator and antiactivator factors affords a switch-like behavior displaying 5 to 50 fold upregulation in expression of the MxiE-regulated genes in the active state compared to the inactive state (Bongrand et al., 2012; Campbell-Valois et al., 2014; Silué et al., 2020).

The coelution of MxiE and IpgC from affinity columns supports the existence of an MxiE–IpgC complex (Pilonieta and Munson, 2008), but its stoichiometry and relevance in vivo are not known. Furthermore, neither MxiE nor IpgC alone can activate the transcription of target genes,

supporting the requirement of a MxiE-IpgC heteromer for transcription activation (Mavris et al., 2002a). All monocistronic genes or short operons regulated by the MxiE-IpgC harbor a 17-nucleotide consensus MxiE box in their promoter (Parsot et al., 2005; Bongrand et al., 2012; Silué et al., 2020). The conservation of key nucleotides of this box is critical for transcriptional activity (Bongrand et al., 2012), most likely because of their role in stabilizing a tripartite promoter-MxiE-IpgC complex. The MxiE box being positioned at -35, MxiE-IpgC is thought to act as an RNA polymerase recruiting factor as expected from classical AraC-like transcriptional activators (Mavris et al., 2002b).

Flow cytometry, a DNA microarray study previously mentioned, and a more recent RNA-Seq analysis defined the MxiE regulon, which is composed of eight operons encoding fifteen proteins on the VP and ten[3] chromosomal genes (Kane et al., 2002; Le Gall et al., 2005; Silué et al., 2020; Hall et al., 2022) (Figure 1.1B). Seven chromosomal genes of this regulon belonged to the *ipaH* family and likely originated from the duplication of one or more ancestral *ipaH* genes from the VP (Le Gall et al., 2005; Silué et al., 2020). The pangenomic RNA-Seq study added two previously unknown chromosomal genes temporarily named gem1[4] and gem3[5] that have no homology with T3SS genes to the MxiE regulon (Silué et al., 2020). Further studies are required to determine whether they encode proteins that are T3SA substrates.

[3] Updated (Hall et al. 2022)
[4] Renamed *icaT* (Chapter 3)
[5] Renamed *icaR* (Chapter 3)

1.6 Iron regulates the T3SS in *Shigella*[6]

Iron plays a key role in the virulence of *Shigella*. For example, the subgroup *Shigella flexneri* produces aerobactin, a hydroxamate siderophore synthesized by the *iucABCD* operon (Payne et al., 2006). Consequently, a live attenuated vaccine (*S. flexneri* 2a SC602 vaccine) based on a strain carrying a mutation in *icsA* and *iuc* was developed (Sansonetti and Arondel, 1989; Rahman et al., 2011). The fact that the vaccine strain had a defect in iron acquisition argues for its importance during infection. This vaccine was immunogenic and protective in North Americans, whereas it was weakly immunogenic and protective in healthy Bangladeshis. The discrepancy in efficacy may be due to the difference in diet between the two populations, as Bangladeshi diets tend to be low in iron compared to North American diets. In addition, since the vaccine has a mutation that prevents the production of the siderophore aerobactin, essential for iron uptake in bacteria, its potency may be reduced in populations with low iron intake (Rahman et al., 2011; Barry et al., 2013; Raso et al., 2023).

The Ferric uptake regulator (Fur) plays a crucial role in iron homeostasis in bacterial cells. In iron-rich conditions, Fur binds to free cytosolic Fe^{2+} ions, forming holo-Fur to repress a series of genes, which reduces iron uptake to maintain a safe concentration of iron. This prevents the harmful iron-dependent Fenton reaction (Lloyd et al., 1997; Massé et al., 2003), thus avoiding oxidative stress damage. Specifically, holo-Fur represses genes involved in producing siderophores (aerobactin and others) and iron channel transporters, which are essential for iron uptake in bacteria (Payne et al., 2006; Porcheron and Dozois, 2015). Conversely, if the concentration of cytosolic Fe^{2+} is low, these genes are expressed to increase iron uptake.

[6] Section 1.6 was not published.

In addition, *ryhB* is another gene repressed by holo-Fur, and upregulated in low iron conditions in both *E.coli* and *Shigella* (Massé and Gottesman, 2002; Murphy and Payne, 2007). RyhB is a small RNA that plays an important gene repression role in the latter condition. It was also noted that the RNA chaperone Hfq is required for the stability and activity of RyhB (Massé et al., 2003; Lalaouna et al., 2021). The upregulation of RyhB in iron-starved cells leads to the repression of many genes, including genes for tricarboxylic acid cycle enzymes *sdh*CDAB, and the superoxide dismutase gene *sodB* (Massé et al., 2003; Seo et al., 2014). On these genes, RyhB acts post-transcriptional by destabilizing the mRNA or interfering with translation (Massé and Gottesman, 2002; Massé et al., 2003). Interestingly, RyhB was also found to downregulate the T3SS in *Shigella* through repressing the expression of *virB*. This likely occurred at the transcription level because RyhB is complementary to the template strand of *virB*. Thus, *ryhB* indirectly downregulates the VirB-regulon independently of *virF* in *Shigella* (Massé et al., 2003; Murphy and Payne, 2007; Africa et al., 2011; Broach et al., 2012). It was also reported that iron deprivation upregulated the *stx* gene that encodes the Shiga toxin in *Shigella dysenteriae* (Calderwood and Mekalanos, 1987). Thus, iron balances the relative expression of the T3SS and Shiga toxin, which may have an important impact on the pathogenesis of *S. dysenteriae*.

Besides, since our review was published, two studies have provided additional insights into regulating the T3SS in *Shigella*. The first study suggests that the complex MxiE-IpgC represses the promoter of *virB*, and this inhibition is dependent on *virF* activation. Still, this repression did not impact the expression of *icsA* (McKenna et al., 2022), whereas *virF* activation positively controls *virB* and *icsA* genes (Figure 1.1A). The crucial binding sites for this negative regulation of the promoter *virB* are coextensive with the binding sites needed for *virF*-dependent activation of the promoter of *virB* (McKenna et al., 2022). Indeed, MxiE-IpgC might prevent VirF from

binding to the promoter of *virB* to activate it, which is the most plausible hypothesis to describe this mechanism of repression (McKenna et al., 2022). However, the detailed mechanism used by MxiE to bind to *virB* is not yet clear because the putative MxiE boxes found upstream of *virB* are degenerated compared to canonical MxiE boxes (McKenna et al., 2022). In addition, the model of the cooperation between MxiE-IpgC and VirF is vague because no physical interactions were demonstrated between them. Also, the second group revealed another role of the MxiE-IpgC complex. This MxiE-IpgC complex activates transcription through the release of H-NS-mediated repression. Unassisted by the VirB repression of H-NS occurring at other T3SS-related genes (described in section 1.4), this recent study supports that MxiE and IpgC are sufficient to counterbalance H-NS silencing at the promoters of both canonical and novel late substrate B. Indeed, overexpression of MxiE-IpgC was sufficient for the expression of *ipaH*7.8 and *yccE* in the *virB* mutant in *Shigella flexneri* (Hall et al., 2022). Both studies refined the role and the mechanism through which the MxiE-IpgC complex regulates the T3SA.

The studies conducted during my thesis have exploited prior knowledge about the transcriptional regulation of the T3SS to identify two novel genes in the MxiE regulon and to identify small molecules that repress the expression of the T3SS. First, the proteins encoded by two chromosomal genes harboring a MxiE box and upregulated by the MxiE-IpgC complex, have been characterized as late substrates B of the T3SA. Initially quipped as gem1 and gem3 (Chapter 2), we renamed them *icaT* and *icaR* after we discovered they encoded T3SA substrates (Chapter 3). Next, we sought to identify inhibitors of the T3SA in *Shigella* by screening approximately 3,000 compounds. We focused on the characterization of the mechanism of action of a novel inhibitor that repressed expression of the T3SA through the small RNA RyhB (Chapter 4).

Chapter 2: RNA-Seq analysis of the T3SA regulon in *Shigella flexneri* reveals two new chromosomal genes upregulated in the on-state

Navoun Silué[a,b], Endrei Marcantonio[a,b,1], F-X Campbell-Valois[a,b,*]

[a]Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

[b]Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Ontario, K1N 6N5, Canada

[1]Current address: Department of Microbiology, University of Hawaii at Mānoa, Honolulu, Hawaii, United States of America

*Corresponding author.
E-mail address: fcampbel@uottawa.ca (F-X Campbell-Valois)

2.1 Preface

This chapter includes the full text edited from the paper published in the journal Methods (Silué et al., 2020). I did all the experiments in this paper, prepared all the figures, and contributed to the redaction of the text. Endrei Marcantonio contributed to the analyses of the RNA-Seq data that led to the identification of two genes that were named gem1 and gem3. This author also highlighted the putative MxiE boxes of these genes by manual inspection of their promoter. As authors in this journal, we have the copyright to use this article in a thesis or dissertation, which may be reused and quoted, provided that the original published version is cited. To harmonize the presentation in each chapter, we moved the section Materials and Methods of this article after the discussion. The complete Supporting Material is available on the journal's website. Most supplemental figures and tables are available in the Appendix 1.

2.2 Abstract (250 words)

Shigella spp. are enterobacteria that invade human colonic mucosal cells using their Type III Secretion Apparatus (T3SA). *Shigella spp.* possess a large plasmid that encodes most of its virulence factors and has been the focus of seminal work that defined the T3SA regulon. Thus, a global assessment of the transcriptional response regulated by the T3SA has been lacking. Herein we used RNA-Seq to identify genes that are differentially expressed when the T3SA is active (on-state) versus inactive (off-state). The quality of the RNA-Seq dataset was validated by its correlation with a prior microarray study. Using novel insights about the expression of non-coding regions, bioinformatic tools and experimentations, we demonstrated the existence of six operons and evidence that *ipaH2.5* is a pseudogene. In addition, 86 chromosomal genes were downregulated in the on-state including several non-coding transcripts corresponding to short antisense RNA embedded in the 16S and 23S RNA genes, and 40 coding transcripts, whose cognate proteins were highly connected at the genetic and biochemical levels. Finally, we identified two novel chromosomal genes dubbed gem1 and gem3, which were upregulated in the on-state similarly to genes belonging to the T3SA regulon. The latter findings were validated on biological triplicates by droplet digital PCR. To our knowledge gem1 and gem3 are the first chromosomal members of the T3SA regulon that have no homologs on the plasmid. Our approach provides a path to optimizing RNA-Seq studies in case of bacterial models that had previously been the subject of medium to large scale studies.

Keywords

Shigella, transcriptomics, RNA-Seq, ddPCR, type three secretion system, operon

2.3 Introduction

Shigella spp. are enterobacteria infecting the large intestine of humans and causing a diarrheal disease with significant morbidity. They are intracellular pathogens that employ a Type III Secretion System (T3SS) to invade non-phagocytic epithelial cells and kill macrophages (Kang et al., 2018). *Shigella spp.* differ from closely related commensal *Escherichia coli* mainly by the presence of a large virulence plasmid of approximately 200 kbp that encodes most components of the T3SS (Buchrieser et al., 2000; Venkatesan et al., 2001). The virulence plasmid has been constructed through the acquisition of genetic elements from at least three different evolutionary origins and a third of its sequence is composed of highly repetitive insertion sequences (IS). Central to the T3SS is the injectisome also known as the Type III Secretion Apparatus (T3SA), a syringe-like nanomachine weighing in the range of megadaltons and composed of tens of proteins that form a cytosolic complex involved in cargo selection, rings that span the bacterial membranes, and a needle that protrudes in the extracellular environment (Hu et al., 2015). It is hypothesized that the T3SA is activated when the tip complex installed at the distal end of the needle is brought into physical contact with the plasma membrane of host cells (reviewed in (Campbell-Valois and Pontier, 2016)). This event leads to the secretion of approximately 35 substrates by the T3SA of *Shigella flexneri* (Pinaud et al., 2017a).

Shigella spp. are cytosol-dwelling pathogens. Following the entry into a host cell in a process akin to phagocytosis, a *Shigella* cell is captured into a vacuole, which it rapidly exits. Once inside the cytosol, it replicates at a high rate (i.e. close to its rate in rich broth), and moves using hijacked F-actin as a propeller (Bernardini et al., 1989). A motile bacterium eventually forms a protrusion from within the initially infected cell into a receiving neighboring cell; this receiving cell then phagocytoses the resulting dissemination vacuole containing the incoming bacterium. This

bacterium rapidly exits this vacuole as well to access the cytosol of the receiving cell. The iterations of this cell-to-cell spreading cycle allow invasion of the mucosae from within. Entry, cell-to-cell spread and vacuole lysis depend on a functional T3SA (reviewed in (Campbell-Valois and Pontier, 2016)). Using a GFP-based transcription secretion activity reporter (TSAR), we have shown that the T3SA activity was high in plasma membrane derived compartments, while it was low in the cytosol (Campbell-Valois et al., 2014, 2015).

The TSAR is based on decades of *Shigella* genetics that led to the identification of promoters that are regulated by the T3SA activity. A key finding was that strains that do not express the tip complex proteins IpaB or IpaD are constitutively secreting (Ménard et al., 1994). This facilitated the description of the T3SA regulon of *Shigella*, which is composed of 16 promoters that are upregulated when the bacteria are secreting (Le Gall et al., 2005; Bongrand et al., 2012). These genes encode substrates of the T3SA that act as effectors within host cells (Parsot, 2009). Hence, the secretion hierarchy of the T3SA in *Shigella* is slightly more complex than the canonical description (Deng et al., 2017), hence displaying two classes of late substrates: a) pre-stored effectors, and b) members of the T3SA regulon. It is noteworthy though that microscopic observation suggests that productive secretion events (e.g. entry or cell-to-cell spread) lead to activation of the T3SA regulon (Campbell-Valois et al., 2014, 2015), suggesting that all classes of substrates are generally secreted within a short period of time (< 60 minutes). The expression of genes of the T3SA regulon is upregulated by the transcription activator MxiE and the co-activator IpgC. However, when the T3SA is inactive, these positive regulators are inhibited by the anti-activactor OspD1 and the anti-coactivators IpaB and IpaC. The activation of the T3SA leads to the secretion of these inhibitory proteins, hence allowing the binding of a MxiE-IpgC complex on MxiE-boxes found in genes constituting the T3SA regulon, and ultimately the upregulation of their

transcription (Mavris et al., 2002a; Le Gall et al., 2005; Parsot et al., 2005; Pilonieta and Munson, 2008; Bongrand et al., 2012) (reviewed in (Campbell-Valois and Pontier, 2016)). Some of these genes are hypothesized to be expressed on polycistronic mRNA (Le Gall et al., 2005; Bongrand et al., 2012), although the existence of these operons has not been directly tested in most cases.

The T3SA regulon was initially defined through a microarray study that was restrained to genes encoded on the virulence plasmid pWR100 (Le Gall et al., 2005; Parsot, 2009). The chromosomal *ipaH* genes (*ipaH*s) that are derived from the plasmid *ipaHs* through gene duplication events had been the only exception to this rule (Penno and Parsot, 2006; Ashida et al., 2007; Bongrand et al., 2012). A genomic scale assessment of the impact of the T3SA activity on the transcriptional program of *Shigella* is warranted to probe the existence of other chromosomal genes in the T3SA regulon. RNA-Seq is an ideal approach to test this hypothesis. Prior to this work, only a few RNA-Seq studies had been performed in *Shigella* (Ni et al., 2015), including one that used comparison between the wild-type and a *fnr* mutant to study the dioxygen regulon in culture medium (Vergara-Irigaray et al., 2014). In the future, one expects that RNA-Seq studies will often be performed on microbes for which medium to large scale studies are already available as is the case with the T3SA regulon in *S. flexneri*. Approaches are needed to optimize labor and cost of RNA-Seq studies in such cases.

Herein, we have compared using RNA-Seq the transcriptional program of *Shigella* cells with T3SA either in their off-, or on-state. We validated this new transcriptomic dataset through its comparison with a previous microarray study. Furthermore, we used bioinformatic analyses and various reverse transcriptase PCR (RT-PCR) techniques to validate the most interesting findings. Specifically, we have demonstrated the existence of six operons on the plasmid and identified 86 transcripts that are downregulated between two- to three-fold in the on-state. More

importantly, we have identified two chromosomal genes that are strongly upregulated in the on-state similarly to members of the T3SA regulon. Their lack of homology with plasmid genes supports the notion that they are novel members of the T3SA regulon.

2.4 Results

2.4.1 Validation of the RNA-Seq analysis

RNA-Seq was used to identify variations in the transcriptional program of *Shigella flexneri* M90T when the T3SA is active (on-state, *ipaD* strain) versus inactive (off-state, WT). To do this, an overnight culture grown at 30°C was subcultured 1/200 into fresh LB broth and incubated under constant shaking at 37°C for 4 hours. Then, the total RNA from the resulting cellular suspension was isolated with Qiagen RNAprotect Bacterial Reagent and RNeasy mini kit. rRNA was removed before performing the library synthesis and the RNA-Seq was performed on the Hi-seq Illumina platform. This dataset is retrievable on the GEO repository using accession number GSE123435. Approximately 80 million sequences were obtained for both tested samples. Sequence reads obtained were analyzed with the Rockhopper analysis software to yield the reads per kilobases per millions of reads (RPKM) for each gene in both conditions (McClure et al., 2013). Approximately 90% of the sequencing reads were mapped on the virulence plasmid pWR100 or the chromosome by Rockhopper. Previous work had reported the expression level in the off-state and on-state of approximately 70% of the genes encoded on the virulence plasmid using microarrays (Le Gall et al., 2005). The on/off fold change (FC) of the 72 genes of this study were well correlated with those obtained by RNA-Seq (R = 0.89) (Figure 2.1).

This similarity is remarkable since the locus mutated to model the on-state differed between the two studies (i.e. RNA-Seq: *ipaD*; Microarray: *ipaB*4). In addition, the slope of the liner regression suggested that FC observed for the RNA-Seq were approximately 5 times higher than their microarray counterparts, hence highlighting the higher sensitivity of the former. Taken together, this analysis validated the RNA-Seq dataset.

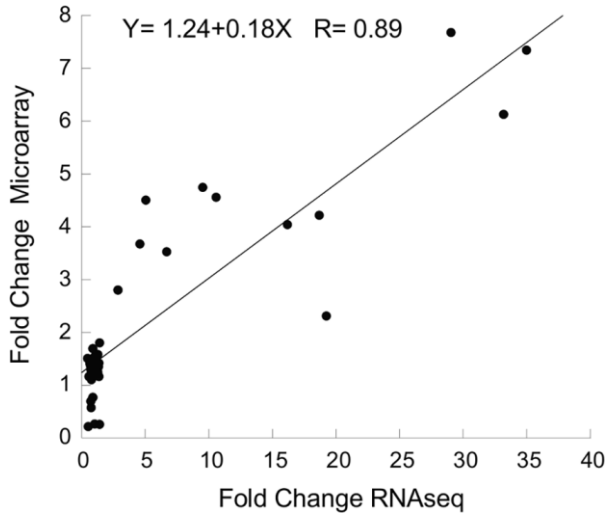


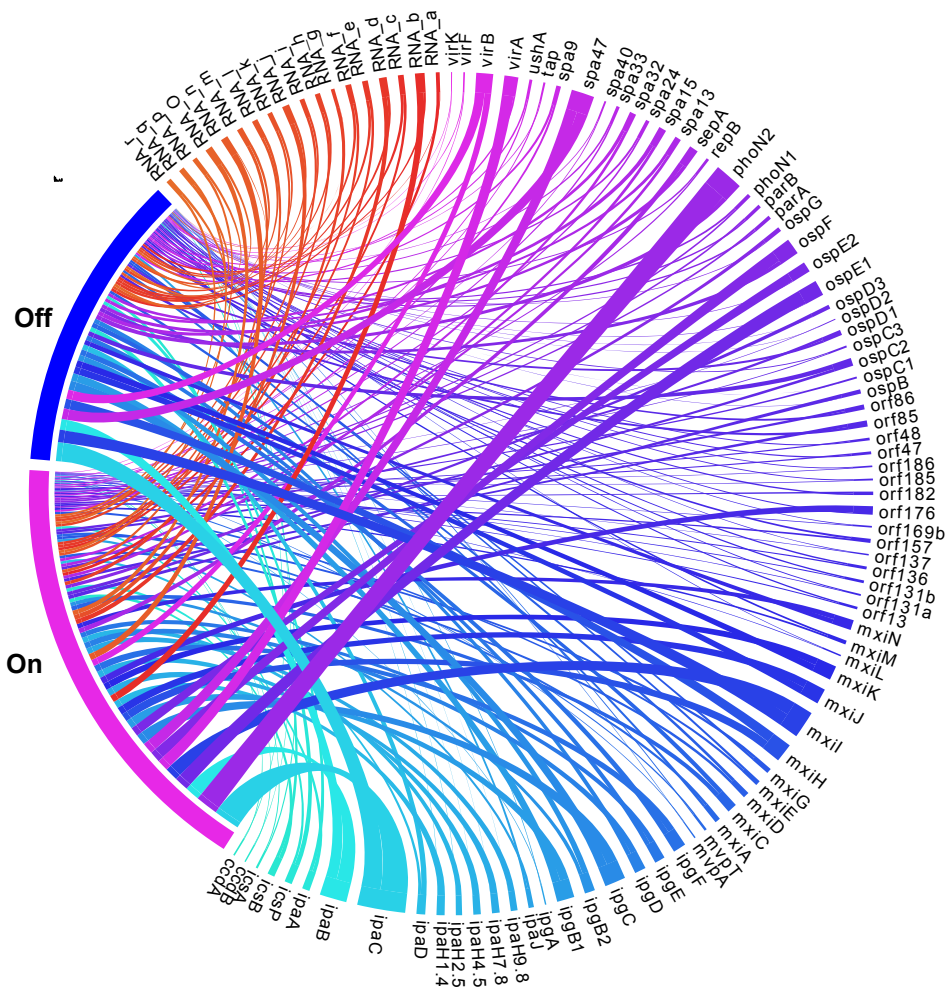
Figure 2. 1. Validation of the RNA-Seq dataset. Correlation plot of the on/off fold change (FC) for virulence plasmid genes according to the RNA-Seq data reported in this study versus a previously reported microarray study (Le Gall et al., 2005; Parsot et al., 2005; Pilonieta and Munson, 2008; Bongrand et al., 2012). The two datasets were linearly fitted using Kaleidagraph 3.5. The quality of this fitting was tested with a Chi-square goodness of fit (R). Mean values of the raw data of the microarray study were calculated and used directly in this plot without further modifications. These microarray data are protected under a CC-BY license. The use of these data does not constitute an endorsement on the part of the licensor.

2.4.2 Virulence plasmid gene expression profiles

The analysis performed by the software Rockhopper indicated that 8% and 11% of sequencing reads were mapped to pWR100 in the off- and on-state, respectively. Of the 104 predicted open reading frames originally annotated (Buchrieser et al., 2000; Venkatesan et al., 2001), we were able to confirm the expression of 100 (Table 2.S.2 and Figure 2.S.1). The four that were not detected belong to the same operon (*orf160-163*). Six more orfs were expressed to very low level. Thus, of the 27 hypothetical protein-coding genes (orfs) (Buchrieser et al., 2000; Venkatesan et al., 2001), 17 were well expressed (Table 2.S.2 and Figure 2.S.1), suggesting they may encode functional proteins. In addition, 18 transcripts mapped to non-coding regions were annotated RNA_a-r (Figure 2.S.1) and belong essentially to two groups. The first group included

transcripts present in multiple copies mapping to a small group of IS sequences (IS629.03, 05, 08, 12 and 15, IS600.07 and .11 and IS1600.06b), and whose expression was not at all or moderately upregulated in the on-state (1< FC< 2) (Figure 2.S.1 and Table 2.S.3). These data suggested that only a small group of related IS are transcribed. The second group included unique transcripts (a, b, g and n) mapping to loci that were rich in coding genes, and whose expression was upregulated in the on-state (FC> 3) (Figure 2.S.1 and Table 2.S.3). We will come back to them specifically in the next section. Next, using CIRCOS, we represented the off- and on-state RPKM values of the 77 most highly expressed transcripts, including the 18 that mapped to non-coding regions (Figure 2.2A). Of the 59 coding transcripts, none were downregulated, 44 did not significantly vary and 15 were upregulated in the on-state (Figure 2.2B). The FC of the latter was on average much greater than two (15.5 ± 10.8) and most displayed q-value with a false discovery rate smaller than 0.1%, suggesting that their upregulation was statistically significant (Table 2.S.3). Fourteen of these genes are members of the T3SA regulon whose expression is controlled by the transcription activator MxiE (Le Gall et al., 2005; Parsot, 2009). A notable addition is *orf176*, which is found between *ipaH9.8* and *ospG* and has similarity with the *yacA* antitoxin (McVicker and Tang, 2016; Pinaud et al., 2017a). However, only 9 of the 15 genes harbored an active MxiE-box immediately upstream of their coding sequence (*ospB*, *ospF, ospE2*, *ipaH7.8*, *ipaH4.*5, ospC1, *virA*, *ipaH9.*8 and *ospE1*) (Le Gall et al., 2005; Bongrand et al., 2012). The transcriptional control of the remaining genes (*phoN2*, *ipaH2.5*, ospD3*, orf176, ospG and ipaH1.4*) is thus likely through a MxiE box located further upstream.

A


Off
On
RNA_a
RNA_b
RNA_c
RNA_d
RNA_e
RNA_f
RNA_g
RNA_h
RNA_i
RNA_j
RNA_k
RNA_l
RNA_m
RNA_n
RNA_o
virB
virA
ushA
tap
spa9
spa47
spa40
spa33
spa32
spa24
spa15
spa13
sepA
repB
phoN2
phoN1
parB
parA
ospG
ospF
ospE2
ospE1
ospD3
ospD2
ospD1
ospC3
ospC2
ospC1
ospB
orf86
orf85
orf48
orf47
orf186
orf185
orf182
orf176
orf169b
orf157
orf137
orf136
orf131b
orf131a
orf13
mxiN
mxiM
mxiL
mxiK
mxiJ
mxiI
mxiH
mxiG
mxiE
mxiD
mxiC
mxiA
mvpT
mvpA
ipgF
ipgE
ipgD
ipgC
ipgB2
ipgB1
ipgA
ipaJ
ipaH9.8
ipaH7.8
ipaH4.5
ipaH2.5
ipaH1.4
ipaD
ipaC
ipaB
ipaA
icsP
icsB
icsA
cpsB
cpsA

B

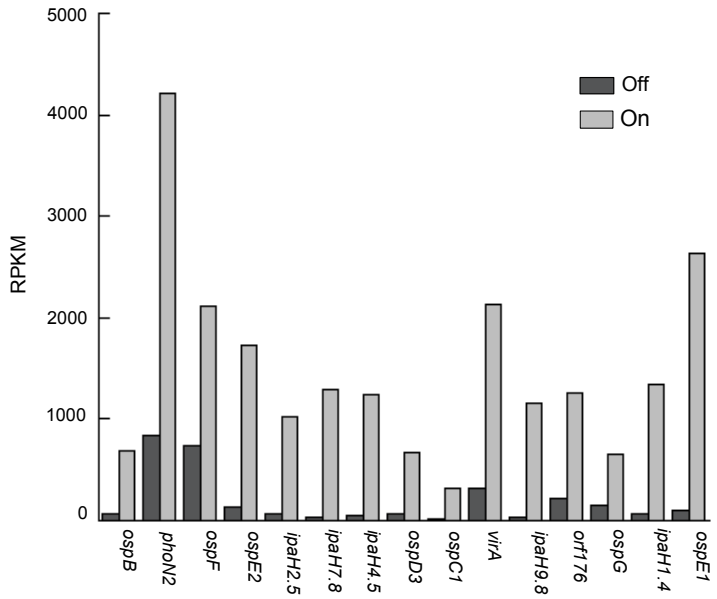

RPKM
5000
4000
3000
2000
1000
0
Off
On
ospB
phoN2
ospF
ospE2
ipaH2.5
ipaH7.8
ipaH4.5
ospD3
ospC1
virA
ipaH9.8
orf176
ospG
ipaH1.4
ospE1

Figure 2. 2. Expression of virulence plasmid transcripts by RNA-Seq. (**A**) Wheel representation of the relative abundance in the off- and on-state of virulence plasmid transcripts with cumulative Reads per Kilobase of transcripts per Million bases read (RPKM) > 177 (Table 2.S.3). Coding transcripts are sorted in alphabetical order. Transcripts corresponding to non-coding regions (RNA) are clustered on the graph and were given a unique letter based on their position on the virulence plasmid. The blue and purple arcs are the anchoring point for data (ribbon) pertaining to the off- and on-state, respectively. The width of the two ribbons attached to each transcript name and their respective anchoring point is proportional to the RPKM values measured in both conditions. Ribbons color are intended solely to facilitate visualization of the graph and do not carry any quantitative information. (**B**) Histogram plot representing the RPKM values in the off-state and on-state of virulence plasmid genes with on/off fold change (FC)> 2. The genes are ordered according to their position on the virulence plasmid.

2.4.3 Operon characterizations

Previously six MxiE-regulated operons had been hypothesized based on the absence of *mxiE* boxes immediately upstream of genes that were, however, upregulated when the T3SA was active; in all cases a functional MxiE box was identified in an upstream gene (Le Gall et al., 2005; Parsot, 2009; Bongrand et al., 2012). To our knowledge, the operon composed of *ospB* and *phoN2* is the only one that had been experimentally validated (Santapaola et al., 2002). This study also showed that both *ospB-phoN2* and *phoN2* transcripts existed, the latter being regulated by a MxiE-independent promoter located between the two genes. Rockhopper uses correlation between expression levels and physical proximity between genes to predict operons. It identified the ORFs-dense mxi/spa and ipa operons, but not the low-density MxiE-regulated operons (data not shown). We therefore used the software WoPPER to identify the MxiE-regulated operons (Puccio et al., 2017). WoPPER was able to predict five (O3-O7) out of the seven operons (Figure 2.S.1 and Table 2.S.1). We also noted that both O5 and O6 had one upregulated transcript that mapped to their intercistronic region (e.g. RNA_g and RNA_n, respectively), hence supporting further the existence of these operons. Similarly, non-coding transcripts RNA_a and RNA_b, which unambiguously matched respectively to the intercistronic regions of O1 (4663-4679 bp) and O2

(11,476-11,539), were also upregulated when the T3SA was in the on-state (Table 2.S.3). Hence, we also hypothesized the existence of operons O1 and O2. O1 corresponds to the *ospB-phoN2* operon mentioned above. The higher expression level of *phoN2* in both states and its lower FC compared to *ospB* is in agreement with previous studies (Santapaola et al., 2002; Le Gall et al., 2005), hence confirming the existence of *ospB-phoN2* and *phoN2* transcripts. O2 is composed of *ospF* and *ospD2*. The FC of 2.84 for *ospF* is lower than the average value for MxiE-regulated genes (15.5 ± 10.8). Furthermore, the FC of 1.47 for *ospD2* is even closer to those of non-MxiE regulated genes (1.0 ± 0.2). A MxiE box is present 53 bp upstream of *ospF* coding sequence (Le Gall et al., 2005; Bongrand et al., 2012). In addition, the existence of a *mxiE*-independent promoter was postulated to explain the lower FC of both *ospF* and *ospD2* (Le Gall et al., 2005; Parsot, 2009). The RNA-Seq FC of these genes are highly similar to those reported in the microarray study (Santapaola et al., 2002; Le Gall et al., 2005). What remained unclear though was whether *ospD2* was co-transcribed with *ospF* and if its expression was also upregulated by MxiE.

To test the existence of these operons (O1-O7), we performed a RT-PCR on the cDNA of the *ipaD* strain using primers located at the 3' end of the first cistron and at the 5' end of the last cistron (Figure 2.3A and Table 2.S.2). We included a negative control with no reverse transcriptase and a positive control containing genomic DNA. O1, O2, O4, O5, O6 and O7 showed no band in the negative control, while a band appeared in both the RT and in the positive control. These data confirmed the existence of all operons but O3 (Figure 2.3B). Thus, an *ospF-ospD2* transcript exists (O2). To verify whether the expression of *ospD2* was dependent on MxiE, we measured by ddPCR the relative expression of *ospD2* and *ospF* in the BS176 (virulence plasmid-cured), *ipaB4* (on-state) and *ipaB4 mxiE* (on-state Δ*mxiE*) strains (Figure 2.S.2). Both genes, albeit to a lesser extent in the case of *ospD2*, showed reduced expression in the absence of MxiE. The expression in

absence of MxiE was, however, much higher than in the negative control plasmid cured strain, supporting the existence of a MxiE independent promoter. O3 is composed of *ospE2* and *ipaH2.5*. It is virtually identical to O7, which is composed of *ospE1* and *ipaH1.4*. The expression of *ipaH2.5* may be reduced by the presence of two large IS between it and *ospE2* (Figure 2.S.1). The RNA-Seq data indicated that *ospE1* and *ipaH1.4* were transcribed to slightly higher level than *ospE2* and *ipaH2.5*. However, the RPKM values are not reliable for almost identical genes such as those. Interestingly, no transcript corresponding to the region between *ospE2* and *ipaH2.5* ORFs were identified contrary to other MxiE-regulated operons. It was therefore possible that *ipaH2.5* was expressed to lower levels than indicated by the RNA-Seq, as the absence of a product in the RT-PCR experiment seemed to suggest. To test this, we cloned the DNA fragments corresponding to operons O3 and O7 into a plasmid that allowed expression of IpaH1.4-2xMyc and IpaH2.5-2xMyc from their native promoter and measured their relative expression level in the off- and on-state by western blot (WB). These data demonstrated that the expression of both IpaH1.4 and IpaH2.5 is upregulated in the on-state, although the latter is expressed at much lower level (Figure 2.3C). The relative expression of IpaH1.4 and IpaH2.5 in the on-state was evaluated by measuring the density of the bands. It is noteworthy that it was necessary to dilute 10-fold the cell lysate of the *ipaD* strain (on-state) expressing IpaH1.4 to obtain unsaturated bands, as required for density measurements (Figure 2.S.3A). These measurements indicated that IpaH1.4 was expressed 16-fold more than IpaH2.5 in the on-state (Figure 2.3D). Previous work had shown that the activity of the MxiE-box of *ospE1* was moderately superior to that of *ospE2* (Bongrand et al., 2012), indicating that promoter strength is insufficient to explain the difference of expression between *ipaH1.4* and *ipaH2.5*. However, to verify that the low expression of IpaH2.5 was not due to an artefact that would have reduced the activity of the upstream promoter in the reporter plasmids, we introduced

a FLAG-tag at the C-terminus of *ospE1* and *ospE2* and probed their expression by the same method. These data demonstrated that the relative expression levels of OspE1 and OspE2 were indistinguishable (Figure 2.S.3B). Taken together, the RT-PCR and WB data suggest that the transcript *ospE2-ipaH2.5* is present in much lower copies than *ospE1-ipaH1.4* and *ospE2* transcripts.

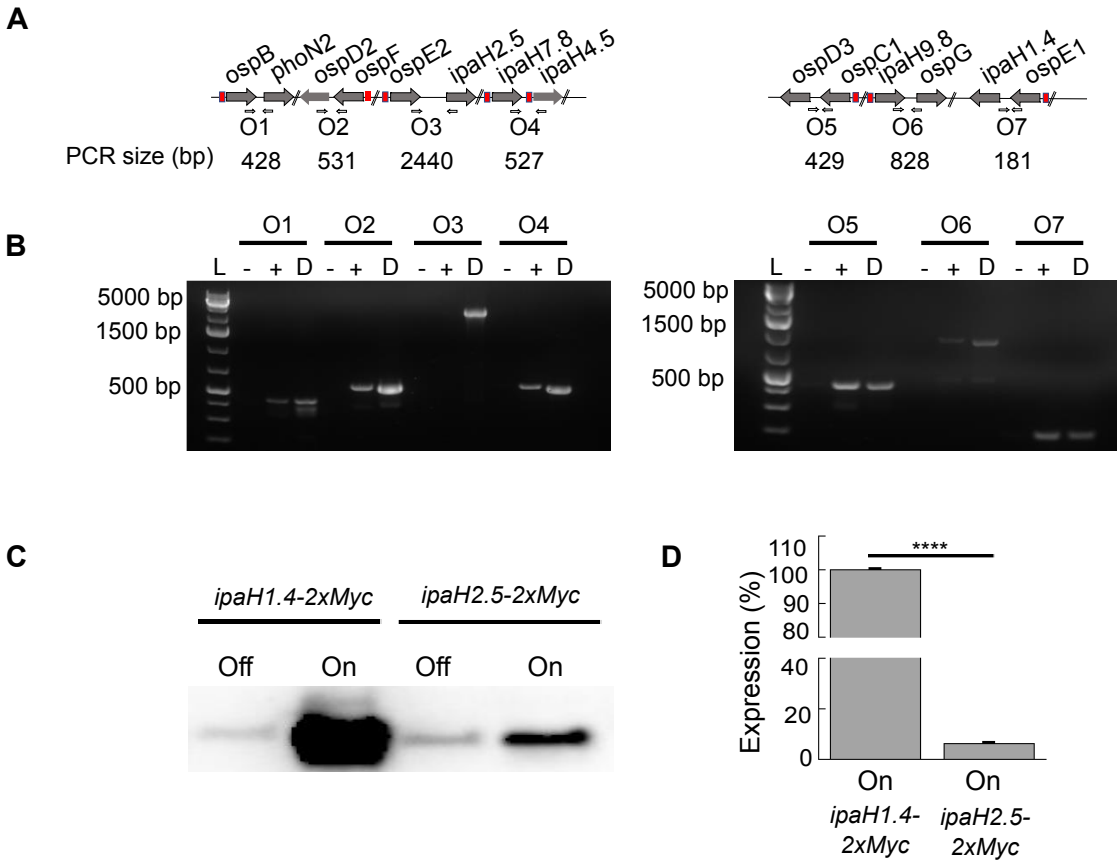


Figure 2. 3. Characterization of the virulence plasmid operons. (**A**) Schematic representation of the seven predicted operons (O1-O7) with their different ORFs (large arrows), their cognate MxiE boxes (red rectangles), and the position of primers used for the RT-PCR (small arrows). (**B**) RT-PCR analysis to validate the existence of O1-O7. Total RNA extracted from an *ipaD strain* culture was isolated and used to prepare the no reverse transcriptase negative control (-), or to synthesize cDNA with the reverse transcriptase (+). Chromosomal DNA extracted from an *ipaD* strain colony (D) was used as a positive control. Each was used as a template for PCR reactions preformed with each of the primer pairs testing the existence of O1-O7. The lane labeled with a L contains DNA ladder. Note that the cell lysate of the *ipaD* (on-state) pUC18Δ *ospE1-ipaH1.4-2xMyc* strain was diluted 10-fold prior to electrophoresis. (**C**). Detection of the expression of IpaH1.4-2xMyc and IpaH2.5-Myc by western blot (WB) in the off-state and on-state using cell lysate of WT and *ipaD* cells harboring plasmids pUC18Δ *ospE1-ipaH1.4-2xMyc*, or pUC18Δ *ospE2-ipaH2.5-2xMyc.* (**D**) A histogram plot of the relative expression of IpaH1.4-2xMyc and IpaH2.5-2xMyc in the off-state and in the on-state estimated by dosimetry of WB bands corresponding to biological triplicates (Figure 2.S.2A). The mean value and standard deviation, as well as the result of Student's t-tests for unpaired data are shown; *p< 0.05, ** p< 0.01, *** p< 0.001, ****p<0.0001, NS: non-statistically significant (p> 0.05).

2.4.4 Chromosomal gene expression profiles

The analysis performed by the software Rockhopper indicated that 83% and 79% of reads mapped to the chromosome in the off- and on-state, respectively. Of the 3311 transcripts with RPKM greater than 4, there were 86 downregulated transcripts and 10 upregulated transcripts in the on-state (q-value< 0.1 and FC< 0.6 or > 2, respectively; Table A.5). 46 of 86 downregulated transcripts were non-coding; of these, 45 corresponded to antisense embedded within the 16S and 23S rRNA genes (Figure 2.4A and Table 2.S.4). Transcripts of 16, 82, 149-153, 181, 190 and 303 nucleotides (nts) originated from the 16S RNA loci. The transcripts of 82 and 190 nts completely overlapped with the transcript of 303 nucleotides, but they did not co-occur on the same gene except on one occasion. The overlapping transcripts appeared due to ambiguities in the sequence of the corresponding gene in the M90T chromosomal sequence recovered from the database (Onodera et al., 2012), which may have introduced mapping errors. Transcripts of 82, 132, 291, 774 and 868 nts originated from the 23S RNA loci. The transcript of 774 nts completely overlapped with the transcript of 868 nts, but it did not co-occur on the same gene. It is noteworthy that the RPKM value from the matching antisense from each rRNA gene converged. Since there are seven copies each of 16S and 23S genes present in *S. flexneri* serotype 5a (Jin et al., 2002; Onodera et al., 2012), the RPKM values of these transcripts represent the summation of the expression level stemming from each rRNA gene. Therefore, the individual RPKM of each transcript may be as much as sevenfold lower. The remaining 40 downregulated transcripts originated from protein-coding genes (Figure 2.4B and Table 2.S.5). 39 of the corresponding proteins were present in the STRING database (Figure 2.4C). The average node degree was 4.31 and the protein-protein interaction p-value smaller than $1x10^{-16}$, suggesting that these proteins were biologically connected to some degree. Specifically, 31 of the 39 proteins formed two distinct networks that contained

four main clusters: 1) galactose transport and metabolism; 2) respiration, tricarboxylic acid cycle (TCA) and associated transporters and enzymes; 3) UMP *de novo* biosynthesis and purine metabolism; 4) translation regulation and polyamines synthesis or transport. Many of the proteins within these clusters are encoded from the same operons, co-expressed, or even physically interact. However, since the T3SA regulon is controlled by the transcription activator MxiE, it was not clear how these genes might be downregulated in the on-state. Hence, due to the lack of a solid working hypothesis, we decided not to pursue further the validation of the downregulated genes. The 10 upregulated transcripts displayed FC values between 2.9 and 47 (Table 2.1 and Figure 2.4B and 2.4D). Seven out of ten belong to the *ipaH* family. The chromosomal *ipaHs*, as their pWR100 homologs, are known members of the T3SA regulon (Penno and Parsot, 2006; Bongrand et al., 2012). *ipaHb1* and *ipaHb2* and *ipaHe1* and *ipaHe2* are virtually identical gene pairs. The RPKM values for these are therefore not accurate, similarly to what was described above for *ipaH1.4* and *ipaH2.5*. Interestingly, three hypothetical protein-coding genes (SFM590T_2352, SF5M90T_2760 et SF5M90T_4150), which we dubbed gem1, gem2 and gem3, were also upregulated in the on-state. The FC of gem1 and gem3, respectively of 47 and 38.5, was similar to that of chromosomal *ipaH*s (32.4 ± 9.0), while gem2 displayed a lower FC of 2.9. We also noted that the guanine-cytosine content of gem1 and gem3, but not gem2, is similar to that of T3SS-associated genes (e.g. the *mxi*/*spa* operon). In addition, their respective locus is not found in the vicinity of chromosomal *ipaHs* or of the *Shigella* pathogenicity Island 2 (SHI-2), the only pathogenicity island described in the M90T strain.

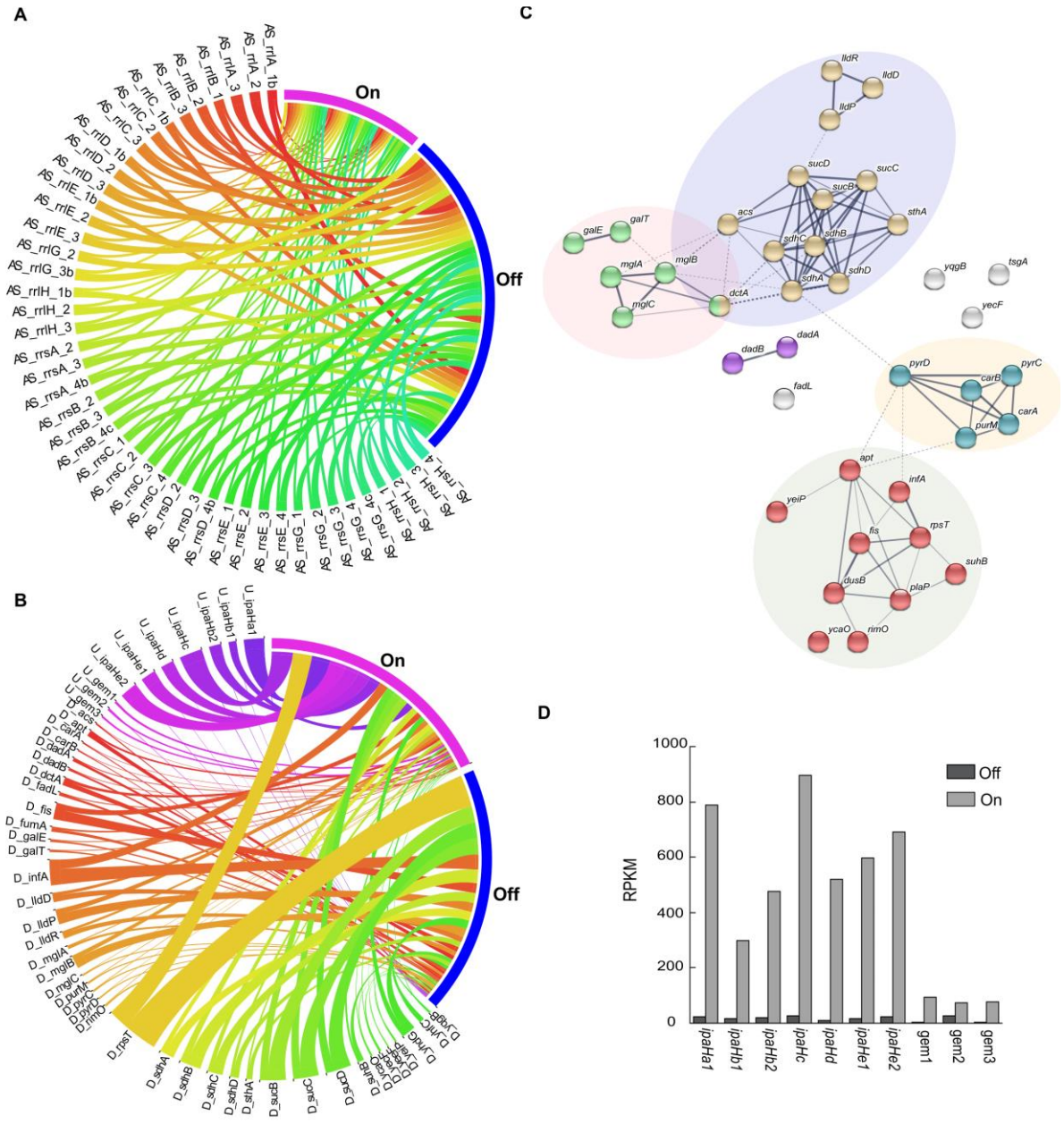


Figure 2. 4. Expression of chromosomal transcripts by RNA-Seq. (**A**) Wheel representation of the relative abundance in the off-state and on-state of non-coding chromosomal transcripts mapped to the antisense strand of 16S and 23S RNA genes. Transcripts with fold change smaller than 0.6 and q-value smaller than 0.1 are shown; none with on/off FC> 2 were detected (Table 2.S.4). The blue and purple arcs are the anchoring point for data (ribbon) pertaining to the off- and on-state, respectively. The width of the two ribbons attached to each transcript name and their respective anchoring point is proportional to the RPKM values measured in both conditions. Ribbons color are intended solely to facilitate visualization of the graph and do not carry any quantitative information. (**B**)Wheel representation of the relative abundance

in the off-state (blue arc) and on-state (purple arc) of coding chromosomal transcripts with on/off FC< 0.6 or > 1.7 and q-value smaller than 0.1 (Table 2.S.5 and Table 2.1). Transcripts are clustered in two groups in function of their on/off FC (e.g. U: upregulated, D: downregulated) and sorted in alphabetical order. The width of the two ribbons attached to each transcript name is proportional to the RPKM values measured in both conditions. See Panel A for details about the interpretation of the wheel. (**C**) STRING protein-protein interaction formed by proteins, which are products of the downregulated coding genes. The proteins formed two distinct networks that contained four main clusters: 1) galactose transport and metabolism (green nodes and pink area); 2) respiration, tricarboxylic acid cycle (TCA) and associated transporters and enzymes (yellow nodes and purple area); 3) UMP de novo biosynthesis and purine metabolism (blue nodes and yellow area); 4) translation regulation and polyamines synthesis or transport (red nodes and green area). (**D**) Histogram plot representing the RPKM values in the off-state and on-state of chromosomal genes with on/off FC> 2 (Table 2.1). Panel A. and B. were generated with the software CIRCOS. Panel D was generated with STRING.

Table 2. 1. Upregulated chromosomal genes of *Shigella flexneri* in the on-state of the T3SA.

Transcription start (nt. #)	Transcription end (nt. #)	Strand	Name	Off-state (RPKM)	On-state (RPKM)	FC (on/off)	q-value
2810440	2808630	-	*ipaHa1*	22	789	35.9	4.32×10^{-32}
777504	775385	-	*ipaHb1*	16	299	18.7	9.14×10^{-4}
2685971	2688115	+	*ipaHb2*	19	475	25.0	5.18×10^{-10}
1437267	1435560	-	*ipaHc*	27	896	33.2	5.16×10^{-41}
1921575	1923530	+	*ipaHd*	11	521	47.4	6.27×10^{-50}
2058935	2060660	+	*ipaHe1*	17	597	35.1	1.22×10^{-15}
2223374	2225241	+	*ipaHe2*	22	691	31.4	1.03×10^{-19}
2469433	2470496	+	gem1	2	94	47.0	2.3×10^{-186}
2901691	2902267	+	gem2	26	75	2.88	9.37×10^{-5}
4386701	4385594	-	gem3	2	77	38.5	1.97×10^{-41}

2.4.5 Confirmation of the upregulation of gem1 and gem3 by ddPCR

The identification by RNA-Seq of three novel genes in the T3SA regulon warranted further validation. To do this, we prepared three biological replicates of the cDNA of the WT and *ipaD* strains to model the off- and on-state of the T3SA and submitted them to ddPCR with primers for gem1, 2 or 3. In addition, we used as a positive control, *ipaH7.8*, which is upregulated in the on-state, and as a negative control, the house-keeping gene *recA*, whose expression is invariable. Droplet numbers versus fluorescent intensities graphs obtained from a typical ddPCR run are shown for each of these genes (Figure 2.5A). The number of positive droplets for each gene was converted into the number of copies of gene specific cDNA molecules detected per microliter of total cDNA (Figure 2.5B). The average value and standard deviation for the three biological replicates confirmed that gem1 and gem3 were significantly upregulated in the on-state with fold changes similar to *ipaH7.8*, although their respective expression level in the on-state were approximately 4-fold (gem1) and 18-fold (gem3) lower than that of *ipaH7.8* (Figure 2.5B). By contrast, gem2 upregulation was insignificant. ddPCR data for these five genes were well correlated by linear fitting with the RNAseq data (R= 0.90) (Figure 2.5C).

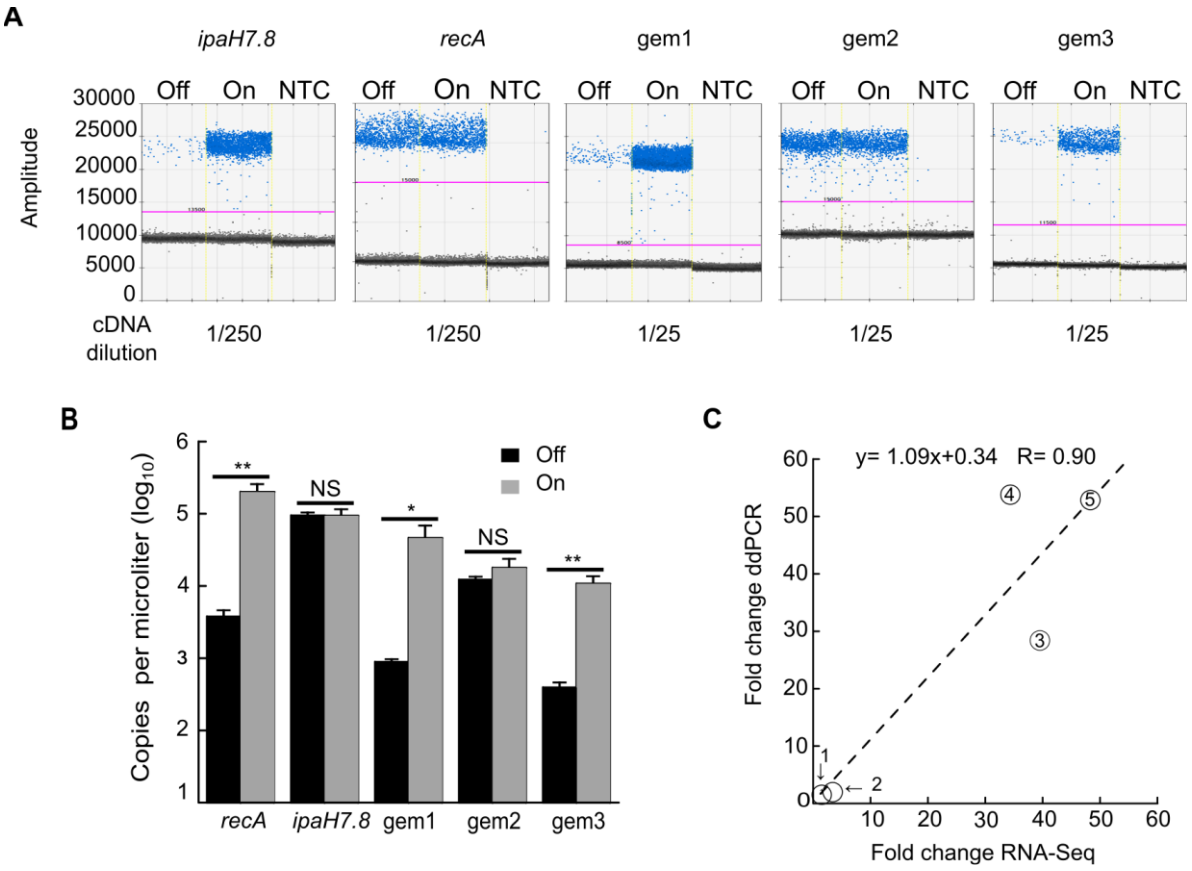


Figure 2. 5. Gem1 and gem3 are upregulated in the on-state. (**A**) Representative droplet intensities graph obtained for the five tested genes in the off-state and on-state as well as the corresponding non-template control (NTC); the signal of tested genes gem1, gem2 and gem3 are compared to *ipaH7.8* (positive control), and the housekeeping gene *recA* (negative control). The NTC of each primer pair was used to fix the threshold (pink line) that allowed discriminating positive (blue dots) from negative (grey dots) droplets. (**B**) Histogram representation of the number of copies of each transcript per microliter of cDNA input as calculated from the droplet intensities graph shown in A. The mean value and standard deviation of biological triplicates, as well as the result of Student's t-tests for unpaired data are shown; *p< 0.05, ** p< 0.01, *** p< 0.001, NS: non-statistically significant (p> 0.05). (**C**) Correlation plot of the fold-change expression on/off of the ddPCR and RNA-Seq for the set of 5 genes mentioned above. The two datasets were linearly fitted (dashed line). The data points indicated correspond to: 1, *recA*; 2, gem2; 3, gem3, 4, *ipaH7.8*; 5, gem1. The quality of this fitting was tested with a Chi-square goodness of fit (R) as indicated on the graph.

The T3SA regulon is composed of genes that are upregulated in the on-state by the activation of the transcription activator MxiE through binding a DNA sequence called the MxiE box (Le Gall et al., 2005; Parsot et al., 2005; Pilonieta and Munson, 2008; Bongrand et al., 2012). We identified a putative MxiE box within the 5' untranslated region of gem1 and gem3, but not of gem2 (Figure 2.6A). To determine whether MxiE was implicated in the regulation of gem1 and gem3, we performed ddPCR on cDNA from the BS176 (virulence plasmid-cured M90T), *ipaB4* (on-state) and *ipaB4 mxiE* (on-state *ΔmxiE*) strains. The expression of our positive control *ipaH7.8* was upregulated in the on-state but downregulated in the on-state in the absence of MxiE or in the plasmid-cured strain, as expected. Interestingly gem1 and gem3, but not gem2, behaved similarly to *ipaH7.8*, while the expression of the negative control *recA* was equivalent in the three strains (Figure 2.6B)[7]. These observations support the notion that the upregulation of gem1 and gem3 in the on-state depends on the transcription activator MxiE.

[7] Figure 2.6B was added.

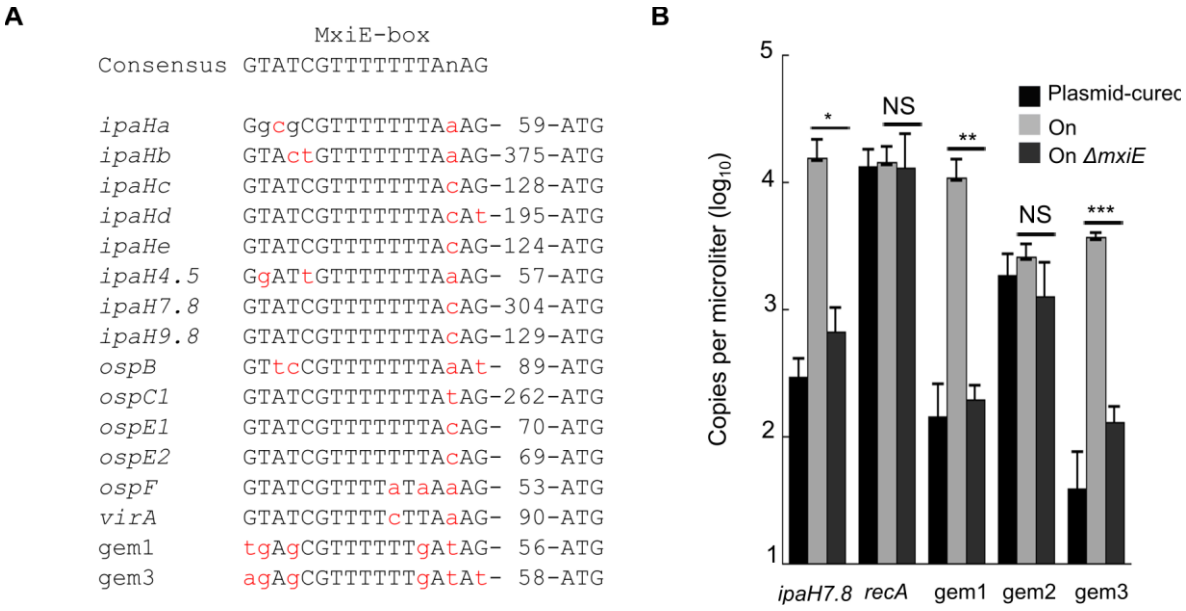


Figure 2. 6. The expression of gem1 and gem3 is dependent on the transcription activator MxiE. (**A**) The putative MxiE box of gem1 and gem3 is aligned with the MxiE-box of genes that are recognized members of the T3SA regulon. The red color lowercase nucleotides highlight the difference with the consensus MxiE box represented at the top of the alignment. The numerical values represent the number of nucleotides between the 3' end of the MxiE box and the translation start site (ATG). (**B**) Histogram representation of the number of copies per microliter for five genes in the cDNA from the BS176 (virulence plasmid-cured M90T), *ipaB4* (on-state) and *ipaB4 mxiE* (on-state Δ*mxiE*). The mean value and standard deviation of biological triplicates of gem1, gem2, gem3, ipaH7.8 (positive control), and *recA* (negative control) are represented, as well as the result of Student's t-tests for unpaired data between the on-state and on-state Δ*mxiE*; *p< 0.05, ** p< 0.01, *** p< 0.001, NS: non-statistically significant (p> 0.05).

2.5 Discussion

Herein, we have investigated the T3SA regulon of *Shigella flexneri*. Previous work had been limited to genes encoded on the virulence plasmid. To our knowledge this study is the first to explore the T3SA regulon encoded on the chromosome. In doing so, we have identified two novel genes that are upregulated when the T3SA is in its on-state. These genes are the first putative members of the T3SA regulon that have no homologs on the virulence plasmid.

RNA-seq is becoming the preferred method to analyze transcriptional programs on a large scale. The depth and sensitivity of RNA-Seq allows it to supersede pioneering approaches such as

microarray (Zhao et al., 2014; Lowe et al., 2017). In this study, RNA-Seq provided an opportunity to revisit the T3SA regulon with greater depth and scale than the seminal microarray study on the virulence plasmid genes (Le Gall et al., 2005). Before we undertook this study, we reasoned that the amount of additional information that the RNA-Seq would provide would be relatively small since all but seven of known genes associated with the T3SS are encoded on the virulence plasmid. To minimize the cost of this analysis without compromising the depth of the sequencing, we thus performed a duplex RNA-Seq consisting of one set of each on- and off-state samples, and tested the validity of this dataset by comparing the fold change of expression of genes that were also probed in the microarray study (Le Gall et al., 2005; Parsot, 2009). The demonstration that FC of genes measured in both studies were well correlated by a linear regression confirmed that the RNA-Seq dataset was of high quality despite relying on a single set of samples. To assess the amount of novel information embedded in the RNA-Seq dataset, we counted the number of chromosomal genes, which expression was over a 1.7-fold threshold. This allowed us to identify 10 genes that were upregulated and 86 that were downregulated in the on-state. Only two previously unknown upregulated genes dubbed gem1 and gem3 had fold change that qualified them as members of the T3SA-regulon. We were then able to validate these findings through ddPCR experiments performed on three biological replicates, and by testing their correlation with the matching values of the RNA-Seq. Hence, we think that our method will be useful to study comparable bacterial systems for which medium- to large-scale transcriptomics data are already available. If this condition is met, it may represent an alternative to the multiplexing of biological replicates (Shishkin et al., 2015; Poulsen and Vinther, 2018), which nevertheless remains the gold standard to validate RNA-Seq data in a resource-efficient manner in the absence of prior studies.

Only few programs have been specifically designed to analyze bacterial RNA-Seq data (McClure et al., 2013; Xia, 2017). In this study, we used the software Rockhopper (McClure et al., 2013). It can perform either *de novo*, or genome-guided transcriptome assembly using reference genome. In cases such as *Shigella flexneri* strain M90T for which reference genome are not available, one may submit fasta, .ptt and .rnt files to Rockhopper for reads mapping. The .ptt and .rnt files are not readily available from gene repositories, but .gff file may be used to generate them. The ability of one to write code would be useful as we are not aware of any freely available web resource to generate .ptt and .rnt files. Apart from this, the use of Rockhopper was overall straightforward. In the analysis of the output of Rockhopper, we noticed that it failed to report statistical uncertainties in the RPKM of identical or highly similar genes. Indeed, Rockhopper mapped matching reads to all identical genes, thereby artificially equalizing their expression level. This was demonstrated by measuring the expression of IpaH1.4 and IpaH2.5 from their natural promoters by immunoblotting. This pitfall of RNA-Seq analysis had certainly an impact on the RPKM values of other *ipaH*s, which have identical 3' fragments in their coding sequence and on other families of effectors such as *ospD1-3*, *ospE1-2* and *ospC1-3*. For this reason, the activity of the promoter of these genes is better approximated by the *lacZ* transcriptional assay previously reported (Bongrand, Sansonetti and Parsot, 2012). Secondly, we also observed that the operon prediction function of Rockhopper did not perform well for operons of the T3SA regulon most likely because of the presence of IS, which introduced large spacing between constituting cistrons (McClure et al., 2013). By contrast, the software WoPPER accurately predicted most operons of the T3SA regulon, probably because it induced a lower penalty for large intercistronic spacing (Puccio et al., 2017). Using RT-PCR, we demonstrated the existence of six operons. Four of them are likely regulated by a unique MxiE-box containing promoter located upstream of their 5'

cistron. Furthermore, RNA-seq data on *ospB* and *phoN2* (O1) are in agreement with previous studies (Santapaola et al., 2002; Le Gall et al., 2005), which had suggested the existence of *ospB-phoN2* and *phoN2* transcripts expressed from a MxiE-dependent and a MxiE-independent promoter, respectively. Prior to this work, it was not clear whether *ospD2* was cotranscribed with *ospF*, and its expression was deemed independent of MxiE (Le Gall et al., 2005; Parsot, 2009). Our data suggest that *ospD2* expression is regulated by both a MxiE-independent and a MxiE-dependent promoter. We also showed by qualitative RT-PCR that an *ospF-ospD2* transcript exists. Since the MxiE box closest to *ospD2* is immediately upstream of *ospF*, we reasoned that the *ospF-ospD2* transcript is regulated by MxiE. Since the expression of both *ospF* and *ospD2* is also controlled in a MxiE independent manner, we still do not know if they are so through a single promoter upstream of *ospF* or through independent promoters. Importantly, the much higher expression of *ospF* in both the off- and the on-state suggests that it may also be transcribed by itself. Taken together, these observations suggest that *ospD2* is a member of the T3SA regulon. The upregulation of *ipaH1.4* in the on-state depends on a MxiE box located upstream of the neighboring gene *ospE1*. The genetic unit *ospE2-ipaH2.5* was obtained by duplication of *ospE1-ipaH1.4*, thus its expression should have been governed by the same mechanism. However, we have demonstrated that *ipaH2.5* is poorly expressed; this is probably due to the presence of two large IS sequences located between it and *ospE2*. In addition, *ipaH2.5* has 12 codons deleted in 3' due to the insertion of another IS sequence. These observations suggest that *ipaH2.5* is a pseudogene. This would explain the lack of phenotype of *ΔipaH2.5*, and accordingly the dominant effect of *ΔipaH1.4* on the stabilization of the host protein HOIP whose degradation is thus likely induced only by the E3 ubiquitin ligase IpaH1.4 in *S. flexneri* M90T strain (De Jong et al., 2016).

This is likely general since the presence of IS downstream and upstream of *ipaH2.5* is common in other serotypes of *S. flexneri*, as well as in *S. sonnei*, *boydii* and *dysenteriae*.

Interestingly, the RNA-Seq analysis also identified 86 transcripts that were downregulated in the on-state. Of those only *fadL* is in the immediate vicinity of an upregulated gene (gem1), which superior promoter activity in the on-state may have had an indirect effect on its expression. Of the 86 downregulated transcripts, 45 corresponded to partly overlapping non-coding transcripts that are antisense to the 16S and 23S rRNA genes. It is unclear whether these antisense transcripts are genuine or are an artefact due to the rRNA removal kit or to the RNA-Seq itself. There is little evidence in the literature of the existence of antisense RNA overlapping with the rRNA, although they have recently been hypothesized to play a role in translation regulation (Garcia-Mazcorro and Barcenas-Walls, 2016). Interestingly, among the 40 downregulated transcripts mapped to coding genes, seven encoded proteins that are known or suspected to regulate translation. By contrast, they did not encode cytosolic chaperones or other proteins that would have been indicative of cellular stress induced by secretion. The protein encoding by the remaining genes were involved mostly in three distinct metabolic processes: galactose transport and metabolism, respiration, the TCA and associated transporters and enzymes and UMP de novo biosynthesis and purine metabolism. In addition, RpsT and PlaP are involved in regulating the cytosolic concentration of putrescein (Panagiotidis et al., 1995; Kurihara et al., 2011). Polyamines had previously been implicated in *Shigella* pathogenesis (Maurelli et al., 1998; Campilongo et al., 2014), but not directly through the T3SA activity. Nevertheless, care must be taken in the interpretation of data about chromosomal genes because the *ipaD* strain used to model the on-state grows more slowly than the WT, which may have indirectly affected their expression. Therefore, these data warrant

further studies before one may conclude on the role, if any, of these downregulated genes in the physiological adaptation of *Shigella* to the on-state.

Finally, we found that gem1 and gem3 are upregulated when the T3SA is active in a MxiE-dependent manner. The presence of a MxiE box upstream from the coding sequence of gem1 and gem3 suggests that their expression is independently controlled. Prior to this work, the seven chromosome-encoded *ipaHs* were the only members of the T3SA regulon not encoded on the virulence plasmid. Our model is thus that gem1 and gem3 are previously unrecognized members of the T3SA regulon. Interestingly, gem1 and gem3 coding sequence share no homology with T3SS-associated genes. However, their GC content is similar, suggesting that gem1 and gem3 may have originated from the virulence plasmid, and share the same evolutionary origin as the T3SS. Previously known members of the T3SA regulon are late substrates that act as effectors once transferred into host cells. Hence the proteins encoded by gem1 and gem3 might be effectors although this remains purely speculative for the moment. Gem1 and gem3 are temporary names for these genes that we suggest keeping until the properties of the corresponding proteins are determined.

2.6 Materials and Methods

2.6.1 Bacterial strains and plasmids

Strains used in this study are derivatives of *Shigella flexneri* serotype 5a strain M90T (accession number for the chromosome: CM001474), which harbors the virulence plasmid pWR100 (accession number for the virulence plasmid: NC_024996, or AL391753) (Sansonnetti et al., 1982), or BS176 , the plasmid cured counterpart of the original M90T strain. In addition, we have used the wild-type *S. flexneri* 5a strain M90T-Sm harboring a streptomycin resistance

mutation , and its derivative mutant strains *ipaD* (Ménard et al., 1994; Mavris et al., 2002a), *ipaB4* and *ipaB4 mxiE* (Ménard et al., 1994; Mavris et al., 2002a). The plasmids pUC18Δ ospE1-Intergenic (Ig)-*ipaH1.4-2xMyc*, pUC18Δ *ospE2*-Ig-*ipaH2.5-2xMyc*, pUC18Δ *ospE1-Flag*-Ig-*ipaH1.4-2xMyc* and pUC18Δ *ospE2-Flag*-Ig-*ipaH2.5*-2xMyc were generated in the following manner. First, pUC18Δ *2xMyc* was generated by inserting a *2xMyc* tag (EQKLISEEDLGGSEQKLISEEDL) by mutagenesis PCR in the plasmid pUC18Δ (Campbell-Valois et al., 2014). Second, genetic fragments from pWR100, comprising the sequence located between the MxiE box of *ospE1* and *ospE2* (Le Gall et al., 2005; Bongrand et al., 2012) and the 3'end of the coding sequence of *ipaH1.4* and *ipaH2.5* (without their stop codons), respectively, were cloned into EcoRI and XbaI sites of pUC18Δ *2xMyc* using MfeI or EcoRI and XbaI, respectively. Finally, a sequence containing the Flag tag (DYKDDDDK) was inserted in 3' of *ospE1* and *ospE2* by mutagenesis PCR. Sequences of the resulting plasmid were confirmed by Sanger sequencing. All restriction enzymes were obtained from Thermo Fisher.

2.6.2 RNA extraction

The wild-type, BS176*, ipaD, ipaB4* and *ipaB4 mxiE* strains were streaked on tryptic soy agar (BD Biosciences, #236920) with 0.01% Congo red (Fisher Scientific, C580-25), and incubated at 37°C for 16 hours (Bahrani et al., 1997). After incubation, red centered T3SS-positive bacteria were picked and inoculated in 2 mL of Difco Tryptic Soy Broth (TSB) (BD Biosciences, #211822) and incubated at 30°C for approximately 16 hours under constant shaking (225 rpm). The overnight cultures were then sub-cultured 1/200 in 4 mL of LB broth and incubated at 37°C for 4 hours with shaking (OD600≈ 1). Immediately after incubation was interrupted, 400 µL of culture was transferred into 800 µL of RNAprotect Bacterial Reagent (Qiagen, #76506). The

resulting cell suspension was vortexed at maximum speed for 5 seconds and incubated at room temperature (15°C -25°C) for 5 minutes. In the meantime, the OD600 was measured for all cultures; the amount of cellular extract used to perform the RNA isolation of each strain was normalized according to these measurements if necessary. Total RNA was extracted using the RNeasy Mini Kit (Qiagen, #74104). In brief, cell suspensions described above were centrifuged at 5,000xg for 10 minutes and the supernatant was discarded. The cellular pellet was resuspended with 200 µL Tris EDTA buffer containing 15 mg/mL lysozyme and 10µL Qiagen Proteinase K and vortexed at maximum speed for 10 seconds. The cells were incubated for 10 minutes at room temperature; during this incubation the cell suspension was briefly vortexed at maximum speed every 2 minutes. After this incubation, 700µL of buffer RLT (buffer RLT supplement with β-mercaptoethanol at 10:1) was added to the cells. Then, suspensions were centrifuged at 17,000xg for 2 min. The resulting supernatant was mixed with 500 µL of 100% ethanol. 2x700 µL of the resulting solution was successively transferred to a RNeasy spin column in separate binding steps. The column washing steps were performed as indicated by the supplier. Finally, elution was performed twice with 40 µL of RNAse-free water.

2.6.3 Library preparation, RNA-Seq and data analysis

Total RNA was quantified using a NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies) and its integrity was assessed using a 2100 Bioanalyzer (Agilent Technologies). The ribosomal RNA (rRNA) was depleted from 400 ng of total RNA using Ribo-Zero rRNA Removal kit specific for bacteria rRNA (Illumina, #MRZMB126). Residual RNA was cleaned up using the Agencourt RNACleanTM XP Kit (Beckman Coulter, #A63987) and eluted in water. Libraries were generated using the KAPA Stranded RNA-Seq Library Preparation Kit

(Kapa Biosystems, #07962142001), as per the manufacturer's recommendations. TruSeq adapters and PCR primers were purchased from IDT. Libraries were quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (ThermoFisher Scientific, #P11496) and the Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems, #07960166001). The average size of fragments was determined with a LabChip GX instrument (PerkinElmer). The RNA-Seq was performed at the McGill University and Génome Québec Innovation Center (Montréal, QC) on the Illumina HiSeq 2500 PE125 platform, which produced paired-end reads of 125 bp. 85,667,365 and 81,908,777 reads were obtained respectively for WT and *ipaD* strains, respectively. The corresponding read files in .fastq format are accessible on the GEO repository as GSE123435. The software Rockhopper (version 2.0.3) was used to map the sequencing reads to S. flexneri M90T's chromosome and pWR100 (McClure et al., 2013). To do this, Rockhopper required a fasta nucleotide file of the chromosome and pWR100 that are retrievable from Genbank, as well as .ptt and .rnt files. The .ptt and .rnt files for the chromosome genes were generated by converting the .gtf annotation file obtained from Ensembl bacteria (accession: ASM25289v1; (Onodera et al., 2012)) using a python script kindly provided by Brian Tjaden. The chromosomal gene names were not extracted correctly from this .ptt file; they were therefore reinserted manually only for genes that displayed significant changes in RPKM values between the off-state and the on-state. The .gff file for the plasmid (AL391753; (Buchrieser et al., 2000; Venkatesan et al., 2001)) was generated with the SeqRet tool from EMBOSS (Rice et al., 2000). Fasta entries were deleted from the .gff file before its conversion into .ptt and .rnt files using the same python script as above. Rockhopper computes the Read per Kilobase per Millions mapped reads (RPKM) to normalize the transcripts abundance. A q-value was also computed for each gene to test the significance of the variation in the abundance of its cognate transcript in the two conditions tested

(McClure et al., 2013). Rockhopper was run on a PC equipped with an Intel Core i5-4690 CPU @ 3.50GHz, 8 gigabytes of RAM and operating with Window 7 64-bit. The on/off fold change (FC) obtained from the RNA-Seq data was compared with previously published microarray data (Le Gall et al., 2005), using linear regression performed with the software Kaleidagraph 3.5 (Synergy Software). The software Snapgene (GSL Biotech) was used to generate the plasmid virulence map. The software CIRCOS was used to represent the expression of genes encode on the chromosome and pWR100 (Krzywinski et al., 2008). The protein-protein interaction network of the products of downregulated genes made using STRING (Szklarczyk et al., 2017). Snapgene, CIRCOS and STRING generated figures were formatted as vector objects using Inkscape or Illustrator CS6 (Adobe).

2.6.4 cDNA synthesis for RT-PCR and droplet digital PCR (ddPCR)

To remove residual genomic DNA, 20µL of purified total RNA (obtained from step 2.2) previously adjusted at 120 ng/µL was mixed with 2.5 µL of DNase I (Thermo Scientific, #8936) diluted with 2.5 µL of 10x reaction buffer and incubated at 37°C for 30-60 minutes. Then, DNAse I was inactivated at 75°C for 10-15 minutes. Using 4 µl of the resulting RNA as template, the cDNA was synthesized with the Improm II Reverse synthesis kit (Promega, #A3800), using 1 µl (0.5 µg) of random primers mix (Promega, #C118B). The no RT control was obtained by replacing the RNA with pure water. The reverse transcription reaction was then carried out at 42°C in 20 µL reaction mix according to the manufacturer's recommendations. The cDNA or the no RT control obtained was used to perform PCR for operons characterization or ddPCR as described below.

2.6.5 Characterization of *mxiE*-regulated operons

The software WoPPER was used to predict operons in pWR100 based on the RNA-Seq gene expression data (Puccio et al., 2017). The existence of these operons and others were then tested by RT-PCR. Total RNA was isolated and treated with DNAse that was inactivated by heating and reverse transcribed, as described above. The amplification of the region between two cistrons of a predicted operon was tested using primers flanking the intercistronic region (Table A.1). To prepare the template, the RT reaction (+), or the no RT negative control (-) was diluted 1/10 with ultrapure water. The DNA positive control (D) consisted of one colony of *S. flexneri* M90T diluted into 100 µl of tryptic soy broth (BD Biosciences, #211823) and incubated during 60 min at 37°C; this resulting culture was diluted 1/12.5 with ultrapure water. The PCR reaction mix was composed of: 2.5 µl of 10x green buffer, 1 µl of 10 mM dNTP, 1.25 µl of each primer (10µM), 13.9 µl of ultrapure water and 0.1µl of DreamTaq Green DNA polymerase enzyme (ThermoScientific, #EP0712) and 5 µl of either one of the three diluted templates described above. The thermal cycler was programmed as follows: 5 min at 95°C, followed by 35 cycles of a three-step PCR consisting of 30 s at 95°C, 30 s at 51.4°C and 30 s at 72 °C and a final step of 5 min at 72 °C. PCR products were analyzed by agarose gel electrophoresis (1.5%) and ethidium bromide staining. WT and *ipaD* cells harboring plasmids described above were grown overnight at 37ºC in tryptic soy broth. Next, 100 µL of that overnight culture was mixed with 100 µL of laemmli 2x and resulting samples were run in polyacrylamide gradient (4-15%) SDS-PAGE gels (BioRad, #456-8086). Typically, 10 µl of samples were run along with 5µl of molecular weight marker (BioRad, #161-0374). The transblot Turbo RTA Transfer kit (BioRad, #170-4272) and Transblot Turbo Transfer system (BioRad) were used to transfer protein onto a PVDF membrane using the fast transfer program. The membrane was blotted with anti-Myc (Genescript, #A00704), or anti-

FLAG (Sigma, #F3165) primary antibodies diluted at 1/10,000 and 1/5,000, respectively and an anti-mouse IgG-HRP secondary antibody (Jackson Immunoresearch, #115-035-003) diluted at 1/25,000. The chemiluminescence signal was acquired with the Clarity western ECL Substrate (BioRad, #170-5060) and the BioRad ChemiDoc MP imaging system (BioRad). Finally, we used the software ImageJ (FIJI) for western blot (WB) quantification by dosimetry (Schindelin et al., 2012; Schneider et al., 2012).

2.6.6 ddPCR

The online-tool Primer3 (v 0.4.0) was used to design all ddPCR primers (Ye et al., 2012) (Table A.1). The PCR reaction mix was prepared as follows: 11.5µl of QX200 EvaGreen supermix 2x (BioRad, #1864033), 0.23 µL of each primer at 10 µM to obtain a final concentration of 100 nM, 6.04 µl of sterile water and 5 µl of a 1/250 dilution (*ipaH7.8* and *recA*), 1/100 (*ospD2*, *ospF*) and 1/25 (gem1,gem2, gem3) dilution of the cDNA reaction described in section 2.4 or 5µl of sterile water for the No Template Control (NTC). To generate the droplets 20 µl of the PCR reaction mix and 65 µL of oil were respectively loaded into separate wells of a DG8 cartridge that was inserted into the QX200 droplet generator (Bio-Rad). Then, 40 µl of the resulting droplets emulsion was transferred to a 96-well PCR plate, which was then sealed and loaded in a T100^{TM} Thermal cycler (Bio-Rad) for amplification. The thermal cycler was programed as follows: 5 min at 95°C, followed by 40 cycles of a three-step program consisting of 30 s at 95 °C for denaturation, 60 s at 57ºC for annealing, and 30s at 72 ºC for extension; the samples were finally incubated for 5 min successively at 4 °C and 90 °C to stabilize the droplets. The number of positive droplets (P_d) per tube was counted using the droplet reader and the software Quantasoft (version 1.6.6.0320) provided with the ddPCR system. Then, for each gene (primer pair), we determined the threshold

to discriminate positive from negative droplets by determining the minimal amplitude that results in no positive droplets in the no template control (NTC). Finally, we computed the concentration of the transcript (Ct) in copy number per microliter of the cDNA reaction with Eq. (1) (Pinheiro et al., 2012; Dreo et al., 2014):

$$\text{Eq. (1) } C_t = {}^{-1}/_{V_d}\left(\ln\left(1 - {}^{P_d}/_{T_d}\right)\right) \times D$$

T_d is the total number of droplets per PCR reaction, V_d is the volume of droplets and D the dilution factor. V_d was assumed to be 9.1×10^{-4} µL as in (Dreo et al., 2014). We carried out the ddPCR on three biological replicates and tested the statistical significance of the data with Student's t-tests with the software KaleidaGraph 3.5.

Acknowledgements

We are grateful to Claude Parsot for sharing raw data of his microarray study and numerous discussions that constituted the foundation of this study. We would like to thank Philip Pelletier and Sofia Perin for technical assistance in setting up the ddPCR experiments and the Core Molecular Biology and Genomics Laboratory.

Funding

This work was funded by the Faculty of Science of uOttawa and the Discovery grant No. 05587 from the Natural Sciences and Engineering Research Council of Canada (NSERC). NS was funded by a stipend from the NSERC CREATE in Technologies for Microbiome Science and Engineering (TECHNOMISE) program.

Chapter 3: *icaR* and *icaT* are ancient chromosome genes encoding substrates of the Type III Secretion Apparatus in *Shigella flexneri*

Running Title: IcaR and IcaT are T3SA substrates in *S. flexneri*

Navoun Silué[a,b], François-Xavier Campbell-Valois[a,b,*]

[a] Host-Microbe Interactions Laboratory, Center for Chemical and Synthetic Biology, Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada
[b] Centre for Infection, Immunity and Inflammation, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON K1N 6N5, Canada

*Corresponding author.
E-mail address: fcampbel@uottawa.ca (F-X Campbell-Valois)

3.1 Preface

This chapter includes the full text edited from the paper published in mSphere (Silué and Campbell-Valois, 2022). I did all the experiments in this paper, prepared all the figures, and contributed to the redaction of the text. As authors in this journal, we have the copyright to use this article, which may be reused and quoted, provided that the original published version is cited. The complete Supporting Material is available on the journal's website. Most supplemental figures and tables are available in the Appendix 2.

3.2 Abstract (250 words)

Shigella is an *Escherichia coli* pathovar that colonizes the cytosol of mucosal cells in the human large intestine. To do this, *Shigella* uses a Type III Secretion Apparatus (T3SA) to translocate several proteins into host cells. The T3SA and its substrates are encoded by genes of the virulence plasmid pINV or by chromosomal genes derived thereof. We recently discovered

two chromosome genes, which seem unrelated to pINV, although they are activated by MxiE and IpgC similarly to some of the canonical substrates of the T3SA. Herein, we show that the production of the corresponding proteins depended on the conservation of an upstream MxiE box. Furthermore, both proteins are secreted by the T3SA in a chaperon-independent manner through the recognition of their respective amino-terminal secretion signal. Based on these observations, we named these new genes *icaR* and *icaT*, which stands for invasion chromosome antigen with homology for a transcriptional regulator and a transposase, respectively. *icaR* and *icaT* have orthologs in commensal and pathogenic *E. coli* strains belonging mainly to phylogroups A, B1, D and E. Finally, we demonstrate that *icaR* and *icaT* orthologs can be activated by the coproduction of IpgC and MxiE in strains MG1655 K12 (phylogroup A) and O157:H7 ATCC43888 (phylogroup E). By contrast, the coproduction of EivF and YgeG, which are homologs of MxiE and IpgC in the *E. coli* T3SS 2 (ETT2), failed to activate *icaR* and *icaT*.

3.3 Importance (150 words)

icaR and *icaT* are the latest members of the MxiE regulon discovered in the chromosome. The proteins IcaR and IcaT, albeit produced in low amount, are nonetheless secreted by the T3SA comparably to canonical substrates. The high occurrence of *icaR* and *icaT* in phylogroups A, B1, D and E coupled with their widespread absence in their B2 counterparts agree with the consensus *E. coli* phylogeny. The widespread conservation of the MxiE box among *icaR* and *icaT* orthologs supports the notion that both genes had already undergone co-evolution with transcriptional activators *ipgC* and *mxiE*- harbored in pINV or a relative- in the last common ancestor of *Shigella* and of *E. coli* from phylogroups A, B1, D and E. The possibility that *icaR* and *icaT* may contribute

to *Shigella* pathogenesis cannot be excluded, although some of their characteristics suggest they are fossil genes.

3.4 Introduction

Shigella is an *Escherichia coli* pathovar infecting the large intestine of humans and one of the major causes of diarrheal diseases. Its pathogenesis is characterized by the invasion of the epithelial cells forming the colonic mucosae and some immune cells found in the *lamina propria* (Kang et al., 2018). The type III secretion system (T3SS) of *Shigella* is essential to the invasion of the cytosol of host cells and the resistance to cell autonomous immunity (Campbell-Valois and Pontier, 2016).

The T3SS is chiefly composed of genes necessary for assembly of the type III secretion apparatus (T3SA), a megadalton protein complex resembling a syringe (Hu et al., 2015, 2017), which is able to distinguish its substrates among thousands of cytosolic proteins and secrete them in an orderly fashion. Early substrates are components of the needle; intermediate substrates are translocators, which forms the translocon that will allow the transfer of effectors from the T3SA needle into the cytosol. Finally, the late substrate consists of effectors that hijack host cell processes (Bajunaid et al., 2020). Besides, the T3SA and its substrates, the T3SS also include chaperones that are essential for the stability and secretion of some substrates and transcriptional regulators that control the expression of the T3SS components (Bajunaid et al., 2020).

The T3SS of *Shigella* spp. and related enteroinvasive *Escherichia coli* (EIEC) is encoded on the virulence plasmid pINV (Buchrieser et al., 2000), which is named differently in each strain (e.g. pWR100 in strain M90T). The T3SA of *Shigella flexneri* strain M90T has currently 35 known or suspected protein substrates (Pinaud et al., 2017a). In *Shigella*, late substrates form two main classes (Bajunaid et al., 2020). The late substrate A are prestored effectors, which must bind to

one of three chaperones IpgA, IpgE and Spa15 in order to be optimally secreted. The late substrates B are effectors that are produced at significant levels only when T3SA are active (Bajunaid et al., 2020). Late substrates B are thus secreted in a chaperon-independent manner as soon as they are produced (Ernst et al., 2018). The expression of genes encoding late substrate B depends on the transcription activator MxiE and the co-activator IpgC. These proteins form a complex that activate the transcription of their target gene through the binding of a MxiE box located at the 5' end of their promoter (Bajunaid et al., 2020). The formation of the MxiE-IpgC complex peaks when the cytosolic store of their inhibitors OspD1 and IpaBC is depleted through their T3SA-mediated secretion, thereby coupling the expression of the MxiE regulon to the activity of the T3SA (Bajunaid et al., 2020). *In vitro*, it is possible to induce the secretion of WT cells with the dye Congo red (CR); alternatively, one can study constitutive secretion of mutants that have defects in the *ipaD* or *ipaB* genes, which encode proteins forming the tip complex that acts as a repressor of constitutive secretion in the WT (Bajunaid et al., 2020). The transcription activation mediated by the T3SA is so exquisitely well controlled that the coupling of a MxiE-regulated promoter to a fast-maturing variant of the GFP afforded a reporter highlighting bacteria that are actively secreting in tissue culture cells or animal tissues (Campbell-Valois et al., 2014; Nigro et al., 2019). Genes that are activated by MxiE-IpgC encode effectors such as OspF, VirA, and the IpaH family (Bongrand et al., 2012). These genes are located in pINV, with the notable exception of the 7 *ipaH* genes located in the chromosome (Ashida et al., 2007; Bongrand et al., 2012).

Recently, we exhaustively probed the genome of *S. flexneri* str. M90T to uncover unknown members of the MxiE regulon by comparing the transcriptome of WT (inactive T3SA) and Δ*ipaD* (constitutively active T3SA) using RNA sequencing (RNAseq). This led to the identification of two new chromosome genes, which transcription regulation resembled those of canonical member

of the MxiE regulon, albeit being expressed one order of magnitude lower (Silué et al., 2020; Bajunaid et al., 2020). These genes are annotated as coding, but because the properties of the corresponding proteins were unknown, we temporarily named them gem1 and gem3. Their transcription start sites (2,469,433 and 4,386,701, respectively) are at least 200 kbp from other chromosome-encoded T3SS genes (Onodera et al., 2012). In addition, gem1 and gem3 possess a consensus MxiE box, suggesting that their transcription regulation is autonomous (Silué et al., 2020). Their GC content is significantly lower than the chromosome, suggesting they were acquired by horizontal gene transfer. Nevertheless, both genes have no sequence homology with pINV. Interestingly, orthologs of both genes are found in pathogenic and non-pathogenic *E. coli* strains. The primary structure of the protein encoded by gem1 has low homology with transposases associated with insertion sequences (IS). Its homolog in *E. coli* K12 is annotated *yfdF* and encodes a hypothetical protein. Instead, the primary structure of the protein encoded by gem3 is annotated as DNA binding protein. Indeed, the Predict Protein server gives a high DNA binding score to segments of this protein located between residues 40 and 160 (Bernhofer et al., 2021). The primary structure of this protein is highly similar to the N-terminal region of the YjgL hypothetical protein of *E. coli* K12. The introduction of an IS sequence at the 3' end of gem3, however, truncated its coding sequence in *Shigella*. Herein, we show that the production of the protein products of these genes depends on the MxiE box in their promoter and that they are substrates of the T3SA. By analogy with the invasion plasmid antigen genes (*ipa*; e.g. *ipaH*s), we renamed them *icaT* (gem1) and *icaR* (gem3), which stands for Invasion Chromosome Antigen with homology for a Transposase and transcription Regulator, respectively. Finally, we also demonstrated that *E. coli* orthologs of *icaR* and *icaT* can be activated by the coproduction of MxiE and IpgC.

3.5 Results

3.5.1 The production of IcaR and IcaT requires MxiE, IpgC and a MxiE box

We have previously shown that the transcription of *icaR* and *icaT* was upregulated in a MxiE dependent manner. Typically, genes belonging to the MxiE regulon are expressed to high levels only when T3SA were active (Silué et al., 2020; Bajunaid et al., 2020). This active state allows the association of an upstream promoter element name the MxiE box with a MxiE-IpgC complex, which favors the initiation of transcription. To verify whether this model applies to *icaR* and *icaT*, we analyzed the production of the corresponding proteins. To do this, we subcloned the promoter region and the coding sequence of *icaR* and *icaT* into a promoterless plasmid, which allowed the detection of their cognate proteins with a 3xFLAG epitope. This plasmid was then transformed into WT and *ipaB4* or its derivatives *ipaB4* Δ*mxiE* and *ipaB4* Δ*ipgC*. The *ipaB4* strains harbor a non-functional allele of *ipaB*, making these cells constitutively secrete *in vitro*, in contrast with the WT, which is non-secreting in the same conditions. The Western blot (WB) of the total cell lysate (TCL) confirmed that IcaR and IcaT, similarly to the endogenous IpaH proteins (IpaHs) were expressed in constitutively secreting *ipaB4* in a *ipgC* and *mxiE* dependent manner, whereas they were not expressed in the WT and RecA was invariable in all strains (Figure.3.1A and B). IcaR-3xFLAG migrated according to its expected MW of 31 kDa. IcaT-3xFLAG migrated around 50 kDa although its expected MW is 44 kDa. It is noteworthy that the pI of IcaT is acidic (5.1), which is notorious for reducing electrophoretic mobility. In addition, we identified a putative MxiE box within the promoter of *icaR* and *icaT* that aligned well with the consensus MxiE box (Figure 3.1C). Guided by this alignment (Bongrand et al., 2012), we introduced point mutations at key positions of the MxiE box of *icaT* and *icaR*. As expected, all mutations reduced the production

plasmid borne IcaR and IcaT in *ipaB4*, whereas RecA and IpaHs produced from their chromosome loci were invariable (Figure 3.1D and E)[8]. In the case of *icaR*, we noted the presence of a second putative MxiE box partly overlapping with the first one. Mutations of this box, however, did not impact the production of IcaR, suggesting it was not involved in the regulation of the promoter (Figure 3.S.1). Taken together, these data indicated that *icaRp* and *icaTp* allowed the expression of their cognate proteins in response to secretion as previously shown for other MxiE-regulated genes.

[8] Corrected from published paper

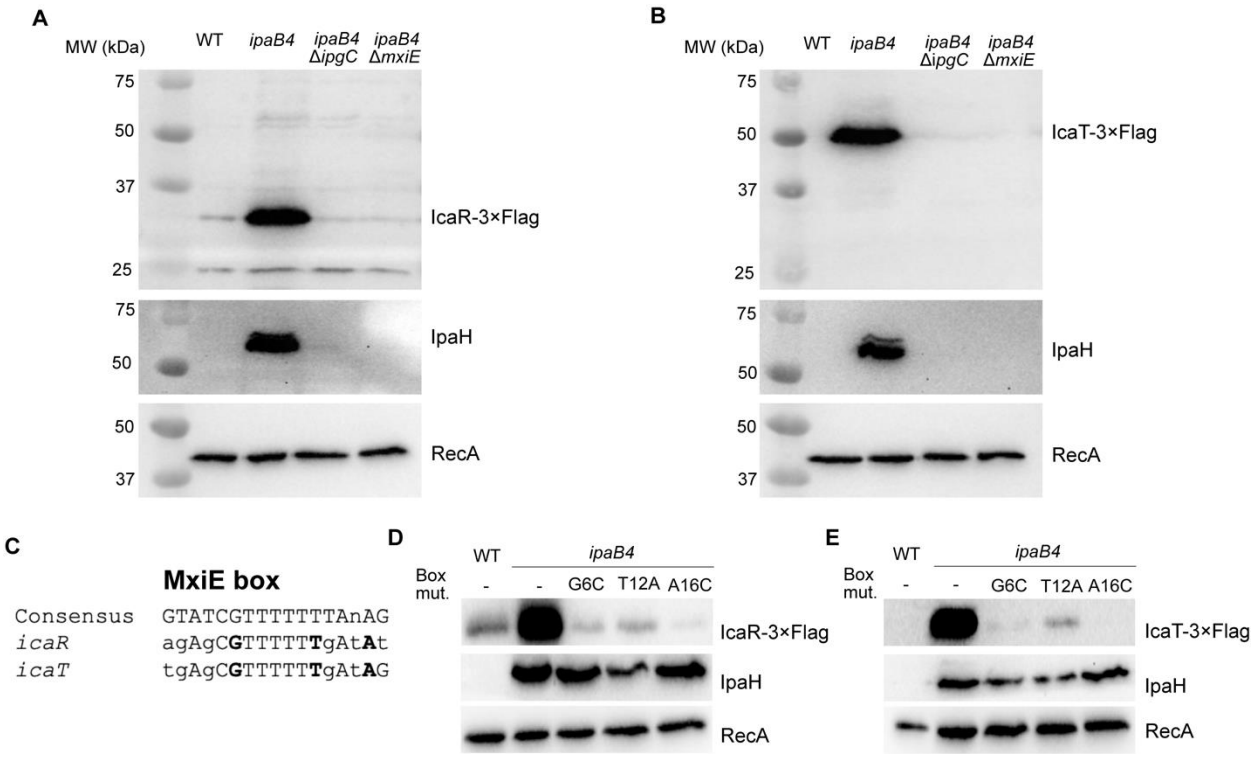


Figure 3. 1. MxiE and IpgC are required for the production of IcaR and IcaT. (**A** and **B**) Detection of IcaR and IcaT by immunoblotting with a FLAG-antibody in the total cell lysate (TCL) of the indicated *S. flexneri* M90T strains harboring plasmid-born *icaR* and *icaT* placed under the control of their endogenous promoters and grown in TSB at 37°C. IpaH and RecA, which production are respectively dependent and independent of MxiE and IpgC, were used as controls; (**C**) Alignment of *icaR* and *icaT* putative MxiE boxes with the consensus MxiE box. The nucleotides in bold are mutated to validate putative MxiE box; (**D** and **E**) Detection of IcaR and IcaT, using the method described above, in strains harboring plasmid born *icaR* and *icaT* in which the indicated mutations were inserted into their MxiE box. In the experimental conditions used here the MxiE regulon is activated in *ipaB4* and its derivatives, but not in the WT strain. The results are representative of three independent experiments.

3.5.2 IcaR and IcaT are secreted in a T3SA-dependent manner

Since the proteins encoded by canonical MxiE-regulated genes are substrates of the T3SA (Bajunaid et al., 2020), we wondered whether the same was true for *icaR* and *icaT*. To test this hypothesis, we used two well-validated approaches that measure the secretion of T3SA substrates in the extracellular medium. The first approach consists in the comparison of the TCL and secreted fraction (SF) of the *ΔmxiD* (T3SA-negative), WT (T3SA are inactive in the absence of inducer) and Δ*ipaD* (T3SA are constitively active, as in *ipaB4*) (Ménard et al., 1994). The second approach compared TCL and SF of Δ*mxiD* and WT in the presence or absence of Congo red (CR), a chemical inducer of secretion, using a classical assay that measures the secretion of substrates that are pre-stored in the cytosol (Bahrani et al., 1997). Since the activity of MxiE-regulated promoters are low in Δ*mxiD* and in the WT in the absence of CR, the use of endogenous promoters would lead to diverging level of expression the various strains required for these assays. To circumvent this issue, the endogenous promoters were replaced with *lacZ*p, which is a constitutive promoter in *Shigella*, and the expression of IcaR and IcaT were probed by WB as described in Figure 3.1. We collected the TCL and SF of Δ*mxiD* (T3SA-negative), WT, and Δ*ipaD* strains in the absence or presence of CR as indicated (Figure 3.2). In the absence of CR, IcaR was invariable in the TCL of the three strains (Figure 3.2A). As expected, the SF of Δ*ipaD,* however*,* contained greater amount of IcaR and of the positive control IpaC than those of WT and Δ*mxiD* cells. To detect their expression from their chromosome locus, we raised polyclonal antibodies against these proteins. The serum raised against IcaT was unreactive against the purified recombinant protein. Fortunately, the serum raised against IcaR was more promising. Indeed, a band with the predicted MW of 27 kDa was specifically observed in the TCL of *ΔipaD* using this serum (Figure 3.S.2A). The Δ*ipaD* harboring the plasmid with *lacZ*p::*icaR*-3xFLAG used in Figure 3.2 yielded a second

band with an apparent molecular weight of 31 kDa as expected, thus demonstrating that the serum recognized IcaR. The intensity of bands at 27 kDa and 31 kDa suggests the recombinant protein was produced comparably to the endogenous protein. Finally, immunoblotting of the SF revealed that endogenous IcaR is indeed secreted in a T3SA-dependent manner (Figure 3.S.2B).

Furthermore, using the classical CR assay (Bahrani et al., 1997), the secretion of IcaR was detected in the WT, but not in Δ*mxiD* (Figure 3.2B). Whereas IcaT was detectable in the TCL and secreted fraction of Δ*ipaD* as expected (Figure 3.2C), it was undetectable in both the TCL and secreted fraction of WT and Δ*mxiD*. Since IcaT was well expressed in the constitutive secretion state established in Δ*ipaD*, we hypothesized that IcaT was unstable in the cytosol of Δ*mxiD* and WT, thus precluding the detection of its secretion using the classical CR assay that detect secretion of stored effectors. We reasoned that WT cells grown in the presence of CR should be able to secrete the protein before it might be degraded in the cytosol. Using this protocol, the level of IcaT in the TCL of the WT strain was increased, and this protein as well as IpaC were detected in the secreted fraction of the WT strain but not in that of the Δ*mxiD* strain (Figure 3.2D).

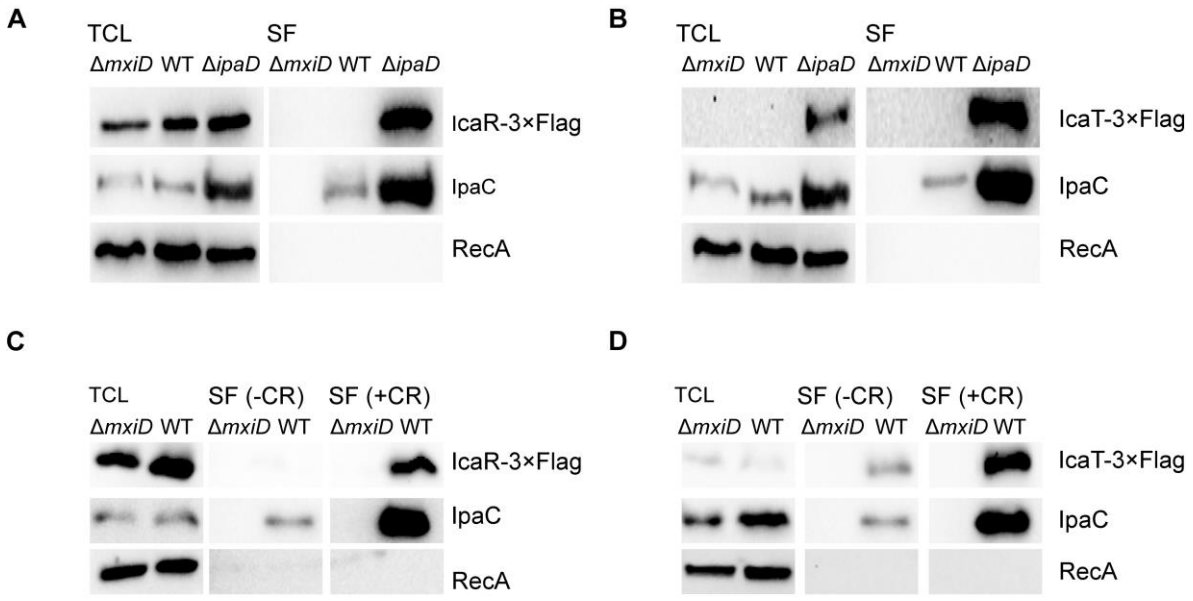


Figure 3. 2. IcaR and IcaT are secreted in a T3SA-dependent fashion. (**A** and **B**) Detection of IcaR and IcaT by immunoblotting with a FLAG-antibody in the total cell lysate (TCL) and the secreted fraction (SF) of the indicated *S. flexneri* 5a str. M90T strains harboring plasmid-born *icaR* and *icaT* placed under the control of the lacZ promoter and grown in TSB at 37°C. IpaC, which is a T3SA substrate, and RecA, a housekeeping protein, were used as controls. (**C** and **D**) Detection of IcaR and IcaT by immunoblotting with a FLAG-antibody in the TCL and SF, as described above, and in the presence or absence of the secretion inducer Congo red (CR). Δ*mxiD* is T3SA-deficient. T3SA of the WT strain can be activated by CR whereas they are constitutively active in Δ*ipaD*. The results are representative of three independent experiments.

We wondered whether the secretion signal of both IcaR and IcaT was located at the amino-terminal end, as observed for bona fide T3SA substrates. Indeed, deletion of 5, 10 or 20 residues at the N-terminus abrogated CR-induced secretion and constitutive secretion of Δ*ipaD* (Figure 3.3A-B and Figure 3.S.3). The Δ20 IcaT was expressed to higher level than the full-length IcaT (Figure 3.S.4). This suggests the N-terminus, which is predicted by Alpha fold and RoseTTAFold to be random coil (Baek et al., 2021; Jumper et al., 2021), destabilizes the protein. Next, we asked whether the N-terminus of IcaR and IcaT are sufficient to induce the secretion. To test this, we

measured the nitrocefin hydrolysis activity of the TCL and SF of Δ*mxiD* and Δ*ipaD* strains producing cytosolic bla_{TEM3} M182T as C-terminal fusion to residues 1-20 of IcaR and IcaT. First, the TCL containing IcaR and IcaT fusions yielded comparable nitrocefin hydrolysis activity, indicating their production in Δ*mxiD* and Δ*ipaD* were similar, and stemming from this, that difference in the SF would correlate with the respective secretion level of the fusion proteins (Figure 3.3C). Second, nitrocefin hydrolysis was detected in the SF of Δ*ipaD*, but not in Δ*mxiD*, suggesting the secretion of bla_{TEM3} mediated by 1-20 was T3SA-dependent. In addition, the efficacy of the secretion mediated by 1-20 of IcaR and IcaT was comparable to 1-80 of OspD1 (Pinaud et al., 2017b), a validated late substrate A of the T3SA (Bajunaid et al., 2020). Because IcaR and IcaT share the same type of N-terminus and transcription regulation as the late substrates B, we reasoned that their secretion should be likewise chaperon independent. To test this hypothesis, we used the CR assay on WT *S. flexneri* 2a str. 2457T and isogenic strains Δ*ipgA*, Δ*ipgE*, Δ*spa15* and the corresponding triple knockout strain ΔΔΔ (Ernst et al., 2018). The amount of IcaR and IcaT in the secreted fraction of these strains was equivalent to the WT, suggesting their secretion was indeed chaperon independent (Figure 3.3D-E). Thus, these data support the notion that IcaR and IcaT are late substrates B of the T3SA in *Shigella* (Bajunaid et al., 2020).

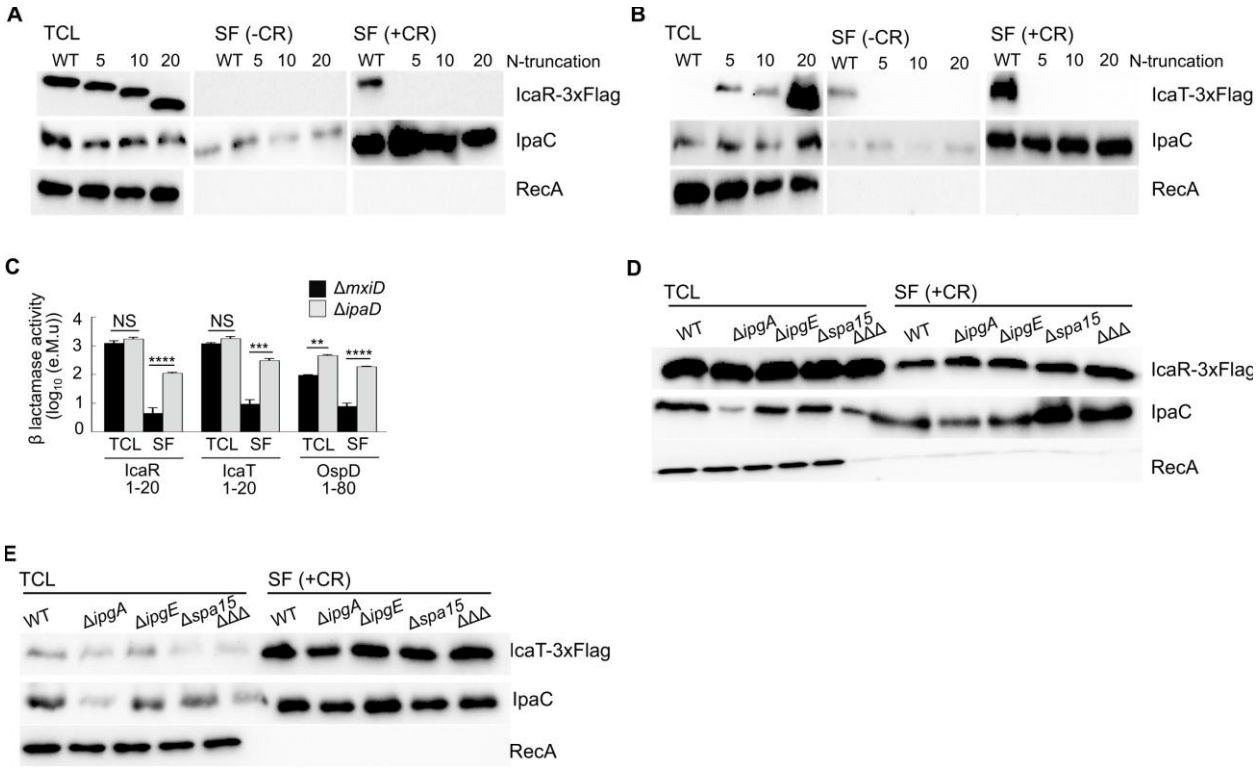


Figure 3. 3. The secretion of IcaR and IcaT depends on their amino-terminus and is chaperone-independent. (**A** and **B**) Detection of IcaR and IcaT N-terminal truncation mutants (0, 5, 10 or 20 residues truncation) by immunoblotting with a FLAG-antibody in the total cell lysate (TCL) and the secreted fraction (SF) of the indicated *S. flexneri* 5a str. M90T strains harboring plasmid-born *icaR* and *icaT* placed under the control of the lacZ promoter and grown in TSB at 37°C in the presence or absence of the secretion inducer Congo red. IpaC, which is a T3SA substrate, and RecA, a housekeeping protein, were used as controls. (**C**) Nitrocefin assays performed on the TCL and SF of Δ*mxiD and* Δ*ipaD* strains producing IcaR (residues 1-20), IcaT (residues 1-20) or OspD1 (residues 1-80) N-terminus and their corresponding deletion mutants fused to bla_{TEM3} M182T; Student's t-tests for unpaired data with a 95% confidence interval are shown. NS: no statistical significance *P ≤ 0.05; **P ≤ 0.01; ***P ≤ 0.001; ****P ≤ 0.0001. (**D** and **E**) Detection of full-length IcaR and IcaT by immunoblotting with a FLAG-antibody in the TCL and SF of *S. flexneri* 2a str. 2457T harboring plasmid born *icaR* and *icaT* under the control of the lacZ promoter and cultivated as above. The ΔΔΔ is a triple knockout strain devoid of the three chaperones (e.g. IpgA, IpgE and Spa15) required for the secretion of some of the late substrates in *Shigella*. The results are representative of three independent experiments.

3.5.3 Orthologs of IcaR and IcaT are present in several phylogroups of *E. coli*

Using BLAST, we identified orthologs of *icaR* and *icaT* in several *Shigella* subgroups and *E. coli* phylogroups (Table 3.S.1). Because *E. coli* strains abound in databases, we performed another BLAST search that excluded *E. coli* and *Shigella*, to probe the occurrence of *icaR* and *icaT* in other species. We found that *icaT* was only present in two poorly studied *Salmonella* spp. By contrast *icaR* also occurred frequently in *Escherichia albertii*. In addition, *icaR* and *icaT* seems absent in *Escherichia marmotae* and *Escherichia fergusonii*. Given the uncertainties in the phylogenetic relationships between species in the *Escherichia* genus (Retchless and Lawrence, 2010), we decided to focus on *E. coli*. Indeed, the locus of *icaR* and of *icaT* are conserved across *E. coli* (Figure 3.S.5), suggesting these genes appeared in this lineage through two independent chromosomal insertion events in the vicinity of *argF* and *fadL*, respectively. The locus of *icaT* is relatively well-preserved, whereas the locus of *icaR* is disrupted to variable extent by IS that are phylogroup or even strain specific. The disruption of *icaR* culminates in *Shigella* with a large 3' deletion (Figure 3.S.5). Indeed, a nucleotide BLAST (BLASTn) with the full-length *icaR* from MG1655 yielded *S. flexneri* str. C32 as the unique hit with 100% query coverage, suggesting that *icaR* truncation is widespread in *Shigella*. The size of the 3' deletion in *icaR* varies in *flexneri*, *sonnei* and *boydii*, suggesting that IS landed on this locus in several independent occasions (Figure 3.S.5). To quantify the disruption of these genes, we counted the occurrence of integral *icaR* and *icaT* using BLASTn within each *Shigella* subgroups using M90T orthologs as queries (Table 3.S.2). Overall, *icaR* and *icaT* appeared severely disrupted in *dysenteriae*, whereas they are less disrupted in the other subgroups, particularly in *flexneri* in which a large proportion of strain seems to possess integral *icaR* and *icaT* (≈60% and 66% of hits, respectively). We next wondered about the status of these genes in EIEC, which harbors its own version of pINV. Using the PATRIC

database, we identified EIEC strains 8-3-DC15 and 8-3-Ti3, which genomes were completely sequenced (Wattam et al., 2014). Both strains harbored identical copies of *icaR* and *icaT* that mapped to loci homologous to those described above (Figure 3. S.5). EIEC *icaR* and *icaT* were devoid of internal insertion or deletion, and respectively displayed 89% and 98% nucleotide identity toward their M90T counterpart. Nonetheless, the MxiE box of *icaR* in EIEC carries a G6T mutation suggesting the activity of its promoter might be diminished (Figure 3.1 and Figure 3.S.6). Notwithstanding this mutation, the status of *icaR* and *icaT* in EIEC appears similar to their M90T orthologs. Furthermore, we observed the presence of *icaR* and *icaT* in all strains of phylogroups A, B1, and E that we scrutinized (Sims and Kim, 2011; Dunne et al., 2017). Likewise, three out of the four strains tested in phylogroup D harbored *icaR* and *icaT*, whereas IAI39 only harbored *icaT*. By contrast, five out of six strains belonging to phylogroup B2 were devoid of both genes, whereas the outlier strain O127:H6 E2348/69 only harbor *icaR* (Table 3.S.1). The *icaR*-negative B2 and D strains have in common a duplication of *argF* and loss of *tabA*, *bdcR* and *bdcA*, which set them apart from other *E. coli* scrutinized. Furthermore, the primary structure of IcaR and IcaT is rather well conserved as the pairwise sequence identities between *E. coli* homologs and *Shigella* M90T was 65-80% and 71-91%, respectively. In *E. coli*, IcaR orthologs are overall similar in length to YjgL from K12 MG1655 (median of 475 residues versus 604 residues, respectively) (Figure 3.S.3). The strains with the lowest molecular weight IcaR (e.g. W, UMN026 and E2348/69) display C-terminal truncations due to IS. The orthologs with higher molecular weight are due to internal and carboxy-terminal additions, which tend to be conserved among strains of the same phylogroup. On the other hand, the conservation of IcaT is high and its length is homogenous across the full dataset (median≈ mean ≈ 352) as well as upon comparing *Shigella* and *E. coli* sequences (means of 314 and 354, respectively), Sb227 being the single low-end outlier (Figure

3.S.7). IcaT from all strains of phylogroup E and from one strain of phylogroup D have a similar 71 residues carboxy-terminal truncation. A single internal addition of similar length and composition is observed at the end of the first half of the primary structure in all strains of phylogroups D and E. This suggests that some of the variations in the sequence of *icaR* and *icaT* are phylogroup specific, whereas others, albeit less frequent, are shared between some phylogroups or are strain specific. Taken together, these results indicate that *icaT* is more conserved than *icaR*.

3.5.4 The introduction of *mxiE* and *ipgC* in *E. coli* activates expression of *icaR* and *icaT*

It is noteworthy that the 5' UTR and the promoter region, notably the MxiE boxes identified in *S. flexneri* 5a str. M90T, are highly conserved in both *icaR* and *icaT* alignments, suggesting that these genes are functional in most non-*Shigella E. coli* (Figure 3.4). Because *E. coli* does not harbor *mxiE* and *ipgC*, we reasoned that they should not express *icaR* and *icaT*. Therefore, we hypothesized that the ectopic production of MxiE and IpgC in *E. coli* should activate the expression of these genes. To test this, we selected a representative of phylogroup A, the commensal K12 strain MG1655, a representative of phylogroup E, the O157:H7 strain ATCC 43888, and BS176, a virulence pINV-cured strain derived from M90T that does not produce MxiE and IpgC. Then, we introduced plasmids allowing the production of MxiE-2xMyc and IpgC-3xFLAG into these strains and measured the expression of *icaR* and *icaT* by ddPCR. These data indicated that *ipgC*$^{+}$ *mxiE*$^{+}$ cells expressed higher levels of *icaR* and *icaT* than the control strains, whereas the expression of the housekeeping gene *recA* was unaffected (Figure 3.5A-C).

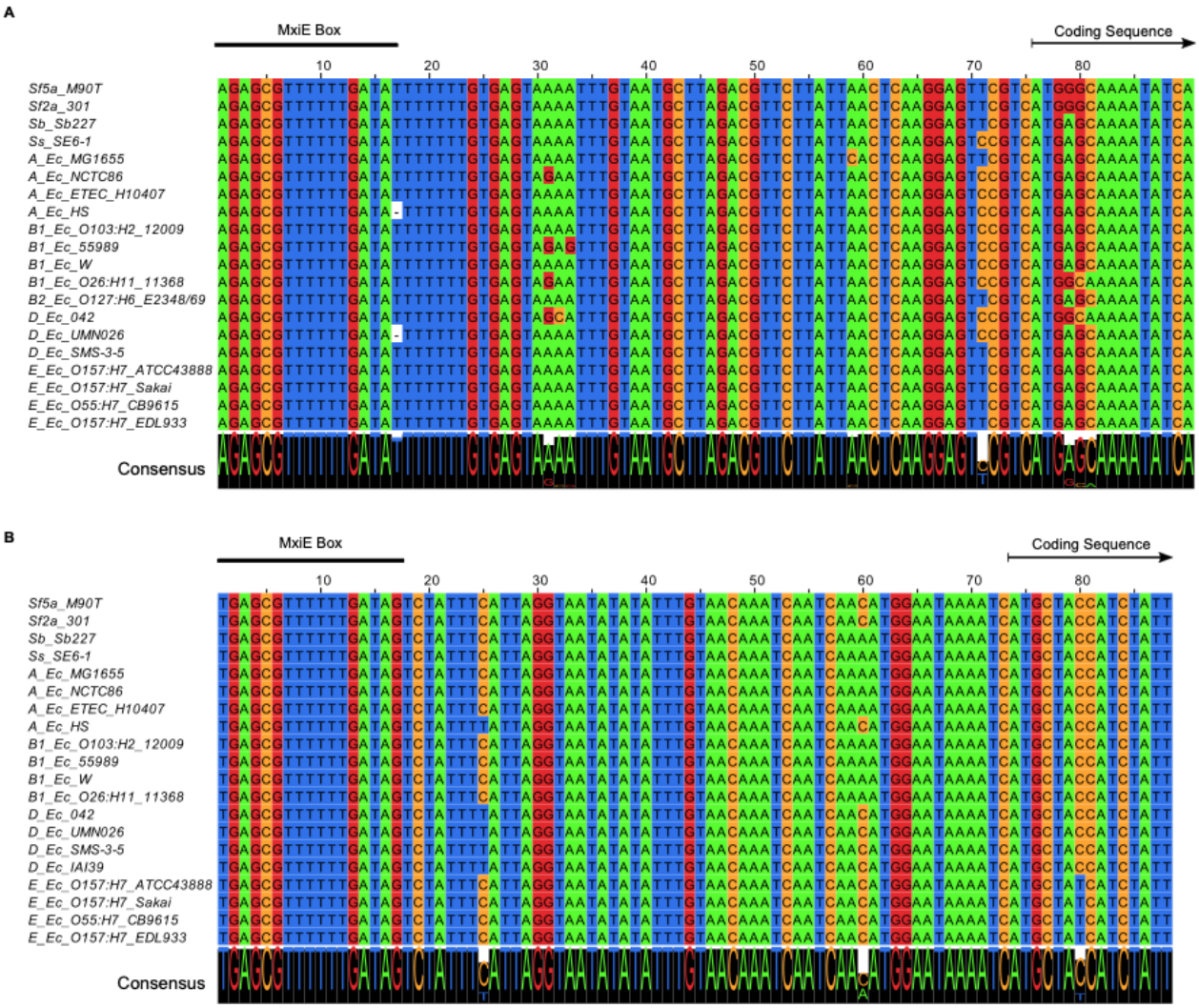


Figure 3. 4. The MxiE box in *icaR* and *icaT* is well conserved. Alignment of the promoter, the 5'UTR and the 5' end of the coding sequence for (**A**) *icaR*; (**B**) *icaT*. The prefix of each sequence indicates the *Shigella* subgroups or the *E. coli* phylogroups to which each corresponding strain belongs. The region matching the MxiE box and the beginning of the coding sequence are indicated. The consensus allows to appreciate the few positions where mutations are occurring.

Using the polyclonal antibody introduced earlier, we confirmed the increased production of IcaR from its endogenous locus in *ipgC*$^{+}$ *mxiE*$^{+}$ cells (Figure 3.5D-F). In both *E. coli* strains, the apparent MW of the protein encoded by *icaR* was approximately 70 kDa, as expected due to the integrity of their coding sequences, which contrast with their truncated *Shigella* counterpart (Figure 3.S.7 and 3.S.8).

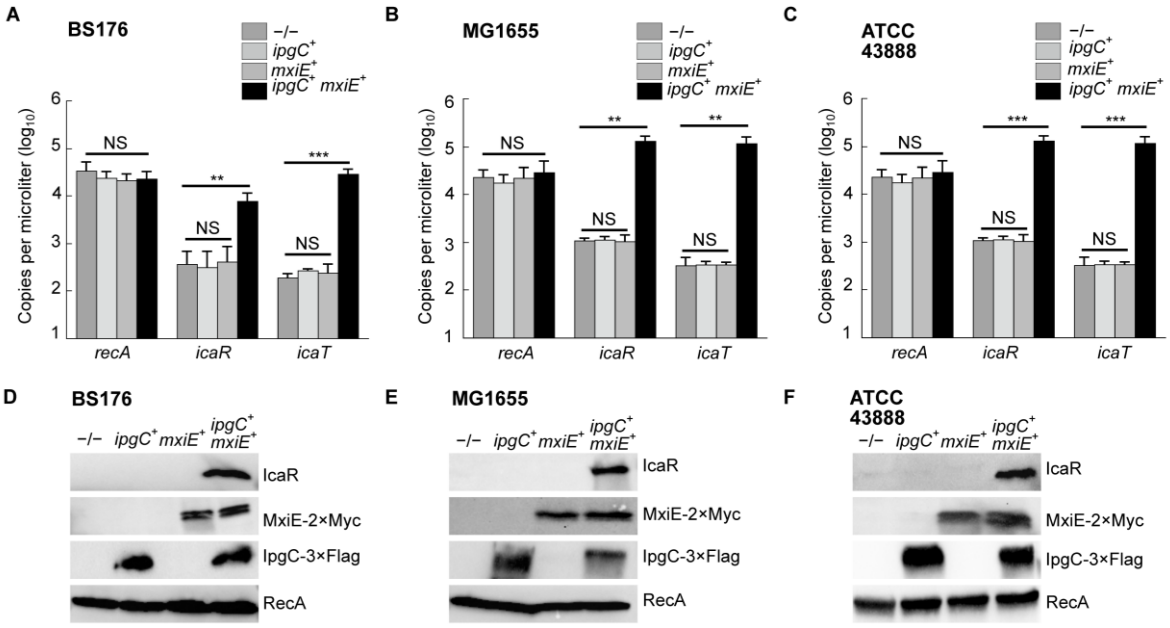


Figure 3. 5. The coproduction of MxiE and IpgC activates the expression of *icaR* and *icaT* in *E. coli*. Quantification of the expression, as the number of copies of transcript per microliter, of *icaR*, *icaT* and *recA* by ddPCR in -/-, *ipgC*$^{+}$, *mxiE*$^{+}$ or *ipgC*$^{+}$ *mxiE*$^{+}$ strains obtained by transformation with the relevant plasmids. (**A**) BS176, a plasmid cured derivative of *S. flexneri* 5a str. M90T; (**B**) *E. coli* K12 str. MG1655, a representative of phylogroup A; (**C**) *E. coli* O157:H7 str. ATCC43888, a representative of phylogroup E. The mean values and standard deviations based on three biological replicates are represented. The statistical significance of these data were tested with a one way ANOVA and a Bonferroni correction for pairwise post-hoc tests with a 95% confidence interval; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; NS (non-statistically significant), $p > 0.05$; (**D** to **F**) Detection of the production of MxiE and IpgC and of IcaR from its endogenous locus by immunoblotting the TCL of strains described in panels A-C. These results are representative of three independent experiments.

The *E. coli* T3SS 2 (ETT2) is present in part or in whole in the chromosome of most *E. coli* phylogroups. The ETT2 harbors distant homologs of *mxiE* and *ipgC* named *eivF* and *ygeG*, respectively. Thus, we wondered whether they might also be endowed with the capacity to activate *icaR* and *icaT*. To test this hypothesis, we introduced *eivF* and *ygeG* in MG1655 and ATCC 43888 (Figure 3.6). The data indicated that the coproduction of EivF and YgeG did not activate the expression of *icaR* and *icaT*, thereby highlighting the unique capacity of MxiE and IpgC to do so.

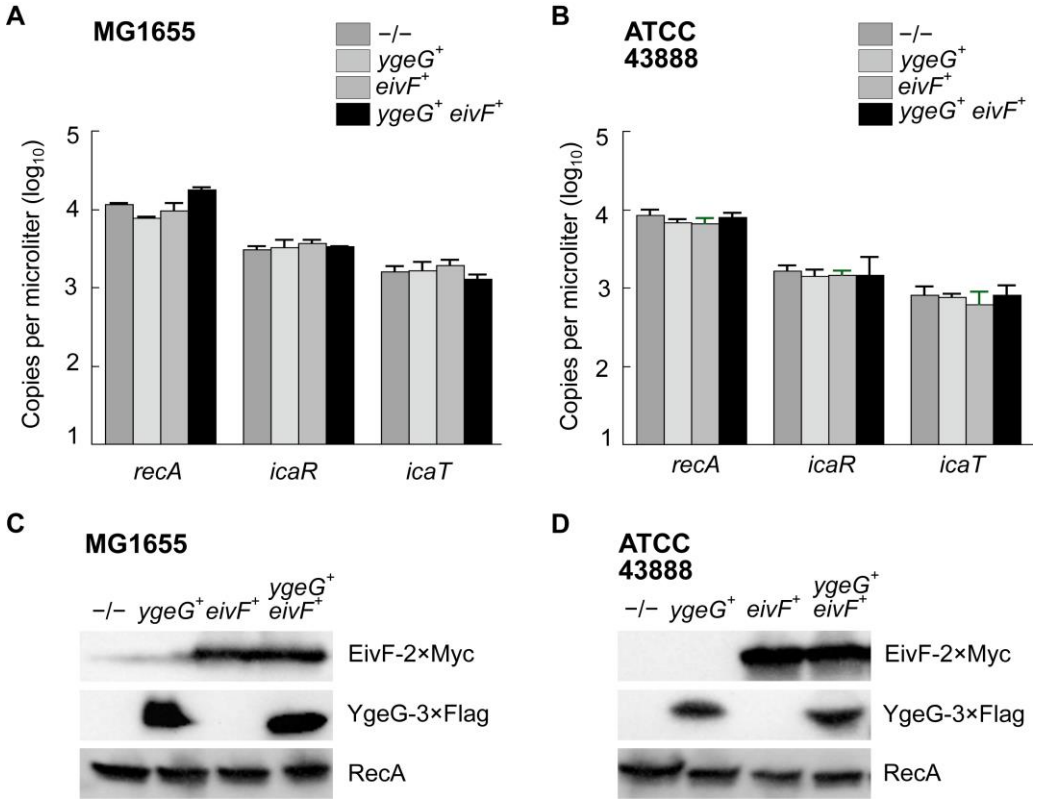


Figure 3. 6. The coproduction of EivF and YgeG does not activate the expression of *icaR* and *icaT* in *E. coli*. Quantification of the expression, as the number of copies of transcript per microliter, of *icaR*, *icaT* and *recA* by ddPCR in -/-, *ygeG*$^+$, *eivF*$^+$ or *ygeG*$^+$ *eivF*$^+$ strains obtained by transformation with the relevant plasmids. (**A**) *E. coli* K12 str. MG1655; (**B**) *E. coli* O157:H7 str. ATCC43888. The mean values and standard deviations based on three biological replicates are represented. The statistical significance of these data were tested with a one way ANOVA and a Bonferroni correction for pairwise post-hoc tests with a 95% confidence interval; they were no statistically significant variations in the data within any given group. (**C** to **D**) Detection of the production of YgeG and EivF by immunoblotting the TCL of strains described in panels **A**-**B**. These results are representative of three independent experiments.

3.6 Discussion

The data presented here suggest that *icaR* and *icaT* are regulated by MxiE and IpgC in *Shigella*. Both genes possess a functional MxiE box that is essential for their expression. Furthermore, the corresponding proteins IcaR and IcaT are secreted by the T3SA in a chaperone-independent manner similarly to other late substrates B (Bajunaid et al., 2020). As expected of T3SA-mediated secretion, the secretion signal of IcaR and IcaT is located at their N-termini. Therefore, the chromosome genes *icaR* and *icaT* encode T3SA substrates that were previously unknown. Prior to this work, a subset of IpaH was the only known case of chromosome-encoded T3SA substrates in *Shigella* (Bajunaid et al., 2020). Chromosomal *ipaH* are thought to have resulted from the duplication of pINV *ipaH* genes with which they share striking similarities. Because they do not have homologs on the plasmid, the origin of *icaR* and *icaT* is difficult to pinpoint and deserving of a separate discussion. Nevertheless, it is noteworthy that the 3' end of the *icaR* open reading frame is truncated by an IS sequence in *Shigella*, suggesting that this happened after their divergence from other phylogroups. Because *E. coli* strains are devoid of chromosomal *ipaH*, we reasoned that the acquisition of *icaR* and *icaT* preceded the transfer of *ipaH* genes to the chromosome of *Shigella*. Generally speaking, *icaR* and *icaT* are frequently disrupted by the landing of additional IS in boydii, dysenteriae and sonnei. In flexneri, both genes are more frequently integral, notwithstanding the 3' deletion of *icaR* discussed earlier. The resulting truncated protein, however, is sufficiently stable to be detected by immunoblotting, suggesting it might still be partly functional. In brief, *icaR* and *icaT* are expressed to low levels (Silué et al., 2020; Bajunaid et al., 2020), and conserved mostly in flexneri, which indicates they are not under strong evolutionary pressure in *Shigella*. Hence, a role for IcaR or IcaT in the pathogenesis of *S. flexneri* cannot be discarded, but it is likely minor.

Furthermore, *icaR* and *icaT* are present in all the tested *E. coli* strains that belong to phylogroups A, B1 and E. Our data suggest that *icaR* and *icaT* can be activated by the ectopic expression of *mxiE* and *ipgC* in a strain of phylogroup A or E. We expect this phenomenon to be widespread because *icaR* and *icaT* orthologs are associated with a conserved MxiE box irrespective of their phylogroups. By contrast, most strains of phylogroup B2 harbor neither *icaR* nor *icaT*. Several independent phylogenetic studies indicate that B2 is the most ancient phylogroup of *E. coli* (Gonzalez-Alba et al., 2019; Touchon et al., 2020; Abram et al., 2021). Instead, phylogroups A, B1 and E constitute a more recent radiation. The presence of *icaR* and *icaT* in these three phylogroupe coupled with their absence in the B2 are in agreement with this model of the evolution of *E. coli* (Gonzalez-Alba et al., 2019). The strain O127:H6 E2348/69 in phylogroup B2, however, harbor *icaR*, suggesting the picture is more complex. Furthermore, *icaT* is present in the four D strains scrutinized, whereas *icaR* is present in three of them. This discrepancy might stem from the genetic diversity of this phylogroup considered to be the closest to the primordial *E. coli* (Gonzalez-Alba et al., 2019; Touchon et al., 2020; Abram et al., 2021). It is also noteworthy that the *icaR*-negative B2 and D strains have in common a remodeled *argF* locus. Therefore, an alternative model is that *icaR* was present in the *E. coli* lineage before the radiation of the B2 phylogroup but was later lost in most B2 strains and some D strains through shuffling of the *argF* locus. By contrast, the specific absence of *icaT* in the B2 phylogroup coupled with the integrity of its *fadL* locus suggest that *icaT* appeared in the last common ancestor of the other phylogroups after B2 had branched off. Taken together, these data suggest that *icaR* and then, *icaT* were inserted in the chromosome through two distinct horizontal transfer events.

Comparative phylogeny of pINV genes and chromosomal genes suggested that the ancestor of phylogroup A and B1 harbored an ancestor of pINV (Escobar-Páramo et al., 2003). This makes

pINV a potential source of T3SS associated genes in *E. coli*. An alternative source is the ETT2, which is ubiquitous in *E. coli,* except in the B2 phylogroup (Ren et al., 2004). Whereas the ETT2 locus is moderately or severely disrupted within phylogroups A and B1, it is often integral in the phylogroups D and E, but evidence supporting its secretion activity is lacking (Zhou et al., 2014; Slater et al., 2018). Besides, the GC content of *icaR* and *icaT* (35.7%, 34.9%, respectively) is lower than that of the chromosome (50.9%), whereas it is comparable with those of the entry region of pINV (34,2%) and the ETT2 (36,9%) (Zhou et al., 2014). Indeed, the expression of *icaR* and *icaT* is controlled by a MxiE box whose occurrence and that of the corresponding transcription activator MxiE and its co-activator IpgC has been solely associated with the entry region of pINV and of the related pEM148 (Buchrieser et al., 2000; Liu et al., 2019). The ETT2 harbors distant homologs to MxiE and IpgC named EivF and YgeG, highlighting the possibility that they might also regulate *icaR* and *icaT*. In contrast with this hypothesis, we found the co-expression of EivF and YgeG did not activate the expression of *icaR* and *icaT*. Furthermore, our data indicate that IcaR and IcaT are substrates of the T3SA encoded by pINV. This argument does not favor pINV over ETT2 as the origin of these genes, because substrates of one type of T3SA were shown to be substrates for others (Subtil et al., 2001; Du et al., 2016). The protein encoded by *icaT* has low homology with transposases associated with IS (Siguier et al., 2015). Nonetheless, the top scoring match of IcaT in the IS finder is not toward the transposase of an IS family present in pINV, but rather toward a clostridia transposase of the IS256 family (Siguier et al., 2006). In eukaryotes, transposases distantly related to those found in IS are implicated in gene fusion events that contributed to the emergence of new biological functions (Cosby et al., 2021). Transposases also play a role in the evolution of new gene function in prokaryotes, as suggested for CRISPR/Cas (Altae-Tran et al., 2021). Similarly, *icaT* might have resulted from the capture of a transposase domain by the T3SS.

On the other hand, PFAM domains search with full-length IcaR (YjgL) from K12 MG1655 revealed statistically insignificant match with a short stretch of the central domain of the *Salmonella* T3SS effector SopA. This region is, however, absent of *Shigella* IcaR due to the carboxy-terminal truncation described above. In light of these observations, we cannot discard that *icaR* and *icaT* originated from neither pINV nor ETT2, but rather from an unknown mobile element. Given the paucity of functional MxiE boxes in the chromosome (Ashida et al., 2007), we reasoned that whatever may be the origin of *icaR* and *icaT*, they nonetheless coevolved to some extent with the T3SS encoded in pINV or a relative, which eventually favored the installation of a MxiE box in their promoter. The conservation of *icaR* and *icaT* in *E. coli* is intriguing. Indeed, the absence of *mxiE* and *ipgC* silenced these genes in *E. coli* devoid of pINV. One could argue that precisely due to this silencing, these genes are under neutral selection pressure in *E. coli*. By contrast, in *Shigella* their cognate proteins could potentially interfere with virulence by competing with other substrates for T3SA-mediated secretion, thus introducing a negative selection pressure that might have contributed to the disruption of *icaR* and *icaT* in this pathovar. Taken together, these observations bring additional evidence that *icaR* and *icaT* did not acquire functions worthy of conservation.

In summary, *icaR* and *icaT* are ancient T3SS-associated genes that were already present in the chromosome of the last common ancestor of *Shigella* and several *E. coli* strains belonging mainly to phylogroups A, B1, D and E. *icaR* and *icaT* are expressed to low levels in *Shigella* and only when T3SA are active, explaining why they had escaped attention thus far. Because IcaR and IcaT are T3SA substrates, they may act as effectors within host cells, although a function in the bacterial cytosol cannot be ruled out. In all cases, their potential role in *Shigella* would be

restrained by their limited conservation and by their low production. Alternatively, as some of their properties suggest, *icaR* and *icaT* might be fossil genes or on the path to becoming so.

3.7 Materials and Methods

3.7.1 Bacterial strains

Shigella flexneri str. M90T WT, and its isogenic mutants *ipaB4*, *ipaB4 ΔmxiE*, *ipaB4* Δ*ipgC*, Δ*ipaD,* Δ*mxiD*, or the plasmid cured strain BS176 were obtained from Philippe Sansonetti and Claude Parsot (Mavris et al., 2002a; Ménard et al., 1994; Sansonetti and Mounier, 1987; Allaoui et al., 1992, 1993). *Shigella flexneri* 2a str. 2457T and its isogenic mutants Δ*ipgA*, Δ*ipgE*, Δ*spa15*, and Δ*ipgA* Δ*ipgE* Δ*spa15 (*ΔΔΔ*)* were obtained from Cammie F. Lesser (Ernst et al., 2018). *E. coli* K12 str. MG1655 (DSM18039) and O157:H7 (ATCC43888) were obtained from the DSMZ and ATCC, respectively. *Shigella* strains and *E. coli* O157:H7 strain ATCC43888 were routinely grown on tryptic soy agar (TSA) and tryptic soy broth (TSB) with or without antibiotics. All other *E. coli* strains were grown on Luria-Bertani agar or broth supplemented with the appropriate antibiotics when required.

3.7.2 Plasmids

Plasmids used in this study are described in Table S3. Briefly, the promoter and the coding region of *icaT* and *icaR* were obtained by PCR with primer pairs HMIO117/HMIO118 and HMIO119/HMIO120 (Table S4), respectively, using a small volume of liquid culture of *S. flexneri* 5a str. M90T. The resulting amplicons were digested by KpnI and Kpn2I, and inserted by ligation into pUC18Δ-3xFlag (Silué et al., 2020). Several derivatives of these constructs were made using

mutagenesis PCR, Gibson assembly (NEBuilder HiFi, New England Biolab) or both (Table S4): 1) three punctual mutations at conserved positions of the MxiE box (G6C, T12A and A16C) were independently introduced (Bongrand et al., 2012); 2) the endogenous promoters were replaced with the *lacZ* promoter *(lacZp);* 3) the 5' end of the coding sequence downstream of the start codon was deleted to generate amino-terminal truncations of 5, 10 and 20 residues; 4) using NEBuilder, *icaR* and *icaT* were inserted at the 5' end of the coding sequence of bla_{TEM3} M182T devoid of the region encoding its signal peptide (Pinaud et al., 2017b). The 3'end of the coding sequence (upstream of bp 63) of *icaR* and *icaT* was then deleted by mutagenesis PCR. The coding sequence of IpgC was amplified by PCR from a colony of *S. flexenri* str. M90T and cloned by restriction digest with BglII and BamHI into pUC18.1 a derivative of pUC18; the 3' end 3xFLAG was then inserted by mutagenesis PCR (FXCV unpublished work). MxiE-Myc was subcloned by restriction digest with EcoRI and XbaI into pSU2.1 a derivative of pSU2718 (FXCV unpublished work). EivF and YgeG were amplified from O157:H7 str. ATCC43888 and cloned by Gibson assembly in place of *mxiE* and *ipgC*, respectively. All constructs were verified by Sanger sequencing (Génome Québec).

3.7.3 Detection of the expression of IcaR and IcaT by immunoblotting

M90T WT, *ipaB4*, *ipaB4* Δ*ipgC* and *ipaB4* Δ*mxiE* isogenic strains harboring either one of plasmids pNS1-8 (Table S3) were inoculated in TSB supplemented with ampicillin (100µg/mL) and incubated overnight at 37ºC with shaking at 250 rpm. Next, the resulting cultures were diluted 1:1 in laemmli 2x and heated at 95ºC for 5 minutes and ran into 4-15% polyacrylamide gradient SDS-PAGE gels (BioRad, #456-8086). Immunoblotting were performed as described in (Silué et al., 2020), using the following antibodies as indicated; primary antibodies: 1/5000 mouse anti-

FLAG (Sigma. #F3165), 1/10000 rabbit anti-IpaH, and 1/1000 mouse anti-RecA (MBL ARM191); secondary antibodies: 1/25000 (except for RecA 1/10000) anti-Mouse IgG-HRP (Jackson Immunoresearch #115-035-003), and 1/10000 anti-Rabbit IgG-HRP (Jackson Immunoresearch #111-035-003).

3.7.4: Immunoblotting assay to measure constitutive secretion of IcaR and IcaT

The secretion assay was performed as previously described with slight modifications indicated below (Ménard et al., 1994). M90T WT, Δ*mxiD* and Δ*ipaD* isogenic strains harboring plasmids pNS9 or pNS10 were inoculated in TSB supplemented with ampicillin (100µg/mL) and incubated overnight at 30ºC with shaking at 250 rpm. Next morning, these outgrowths were subcultured 1:100 into 4 mL of TSB and incubated for 4 h at 37ºC. The optical density at 600 nm (OD600) was used to normalize the volume of each culture used to prepare the TCL, as described above, and the secreted fraction (SF). To isolate the SF, 2 mL of subcultures was centrifuged 10 min at 17000xg. The supernatant was transferred to a fresh microfuge tube and the centrifugation repeated a second time to remove residual cells. Then, 1.4 mL of the resulting supernatant was incubated overnight at 4ºC in 10% (vol/vol) trichloroacetic acid (TCA) and centrifuged during 10 min at 14000xg. The resulting protein pellet was washed twice with 200 µL of cold acetone (-20°C) and centrifuged 5 min at 14000xg. The pellet was air dried for 5 minutes using a dry bath set to 95ºC in order to evaporate acetone traces. The pellets were resuspended in 50 µL 1x laemmli buffer, loaded on a SDS PAGE and immunoblotted as described above. The membranes were blotted with anti-FLAG and anti-RecA antibodies described above, and with 1/1000 mouse anti-IpaC (clone N9) (Phalipon et al., 1992); this antibody was obtained from Armelle Phalipon. The secondary antibody anti-Mouse IgG-HRP was used as described above.

3.7.5 Immunoblotting assay to measure Congo red-induced secretion of IcaR.

The secretion assay was performed as previously described with slight modifications indicated below (Bahrani et al., 1997). Briefly, M90T WT and Δ*mxiD* harboring plasmids pNS9, 11, 13, or 15 were incubated overnight at 30ºC and subcultured at 37ºC, as described for the constitutive secretion assay. Using OD600, an equivalent number of cells from each culture was pelleted, resuspended in 2 mL of 25 µM Congo red (CR) diluted in phosphate buffered saline (PBS), and incubated for 30 min at 37ºC. The TCL, SF, SDS-PAGE and immunoblotting were performed as described above. The only notable change was that 1.8 mL of culture supernatant was precipitated with the TCA in order to isolate the SF.

3.7.6 Immunoblotting assay to measure Congo red-induced secretion of IcaT.

M90T WT and Δ*mxiD* harboring plasmids pNS10, 11, 12, 14, or 16 were incubated overnight as described in the previous section. The resulting outgrowths were subcultured 1/100 in 100 mL TSB and incubated 2h at 37ºC. Then, the culture of each strain was split into two groups of 50 mL each. In the first group, CR was adjusted to 10 µM to induce the secretion, whereas in the second group, serving as a negative control, no CR was added. Both groups were then incubated at 37ºC for 4h. The resulting cultures were centrifuged during 20 min at 4500xg; 40 mL of the resulting supernatant was used to obtain the SF using TCA precipitation, as described above. The protein pellets were resuspended in 400 µL of 1x laemmli and 4 µL of NaOH (8M) to neutralize the pH. Samples were analyzed by immunoblotting as described above.

3.7.7 β-lactamase secretion assay

Δ*mxiD* and Δ*ipaD* were transformed with pNS17, pNS19, or pSU2.1tt ospD1sh M31L bla_{TEM3} M182T; the latter described in (Pinaud et al., 2017b), serving as a control. The nitrocefin β-lactamase secretion assay was performed as previously described (Pinaud et al., 2017a, 2017b). The absorbance of nitrocefin hydrolysis was measured at 486 nm into a 96 well microplate (Greiner #655096) with a BioTek Synergy H4 plate reader.

3.7.8 Activation of *icaR* and *icaT* in *E. coli*

The M90T derivative plasmid-cured strain BS176, *E. coli* K12 str. MG1655 and O157:H7 str. ATCC43888 harboring plasmids pNS21 (*ipgC*$^+$), pNS22 (*mxiE*$^+$), both (*ipgC*$^+$ *mxiE*+), or none (-/-) or pNS23 (*ygeG*$^+$), pNS24 (*eivF*$^+$), both (*ygeG*$^+$ *eivF*$^+$), or none (-/-) were cultivated, as described in the previous section, except that pNS22 and pNS24 were maintained with 30 µg/mL chloramphenicol. Primers design, RNA extraction, cDNA synthesis, and ddPCR were performed as previously described (Silué et al., 2020). We used the primers ddO1/ddO2, ddO7/ddO8 and qPCR87/qPCR88 to amplify *icaT*, *icaR* and *recA*, respectively. Care was taken to select primers annealing to regions that were identical across the strains analyzed. To perform the ddPCR, template cDNA of BS176, ATCC43888, MG1655 were diluted 1/75, 1/100 and 1/250, respectively. The optimal annealing temperature of 58ºC was determined through a temperature gradient experiment. The culture used for preparing the cDNA was also used for the detection of IcaR by immunoblotting. The TCL were prepared as described above, but 1 mL of BS176 and ATCC43888 cultures were resuspended in 100 µL of laemli (10x concentrate). Primary antibodies used: 1/500 rabbit polyclonal anti-IcaR antibody raised against residues 143-158:

SDGSTNRYEGKSFERK (MÉDIMABS, custom made #pAb170-Anti P2); 1/20000 mouse anti-Myc (Genescript, #A00704); other primary and secondary antibodies are described above.

3.7.9 Bioinformatics

Strains representative of each *Shigella* subgroup and *E. coli* phylogroup with available whole genome sequencing were mined for *icaR* and *icaT* using protein and nucleotide BLAST searches (Altschul et al., 1990), and *S. flexneri* 5a str. M90T *icaR* and *icaT* as queries. Nucleotide BLAST (BLASTn) searches were also performed by excluding *Shigella* and *E. coli* to identify other species that harbored the genes. These were complemented with BLASTn against species related to *E. coli* such as *Escherichia marmotae*, *Escherichia albertii*, *Escherichia fergusonii* or *Salmonella* spp. The loci of selected *E. coli* strains were obtained through the NCBI Sequence Viewer (Rangwala et al., 2021), and represented using Snapgene (Insightful Science). Protein and nucleotide alignments were done with MUSCLE in Jalview using ClustalX coloring (Edgar, 2004; Waterhouse et al., 2009). The region spanning from the MxiE box to the end of the coding sequence of *icaR* and *icaT* from str. M90T were used as query in successive nucleotide BLAST searches within each *Shigella* subgroup. We categorized *icaR* and *icaT* in each strain identified in our list as integral, moderately, or highly disrupted. The threshold used to establish those categories were as follows. Integral: number of hits with greater than or equal to 99% query coverage, greater than or equal to 98% sequence identity, and gap intolerance; moderately disrupted: number of hits with greater than or equal to 80% query coverage subtracted by integral genes; highly disrupted: total hits subtracted by integral and moderately disrupted genes. The percentage in each category was computed using the number of total hits as the denominator.

Acknowledgements

We would like to acknowledge the role of the support staff of the Faculty of Science to the realization of this work, particularly G Hatch for help in the realization of the ddPCR. We are grateful to JW Keillor, CN Boddy and JP Pezacki and their research group for sharing their equipment, CF Lesser and A Phalipon for sharing strains and reagents. This work was funded by the Faculty of Science of uOttawa and the Discovery grant no. 05587 from the Natural Sciences and Engineering Research Council of Canada (NSERC), the CFI John R. Evans Leaders fund no. 34789, and CIHR project grant no. 159517. NS received a stipend from the NSERC CREATE program Technologies for Microbiome Science and Engineering (TECHNOMISE).

Chapter 4: Identification of Two Chemical Inhibitors of the Type Three Secretion System in *Shigella*

Navoun Silué[a,b] ; Silei Bai[c]; France Ourida Manigat[a,b]; Endrei Marcantonio; Xinxin Feng[c]; F-X Campbell-Valois[a,b]*

[a] Host-Microbe Interactions Laboratory, Center for Chemical and Synthetic Biology, Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[b] Centre for Infection, Immunity and Inflammation, Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, ON K1N 6N5, Canada

[c]Hunan University, China

*Corresponding author.

E-mail address: fcampbel@uottawa.ca (F-X Campbell-Valois)

4.1 Preface

The l-TSAR reporter plasmid was constructed initially by Endrei Marcantonio. The preliminarily small molecules screening was done by Silei Bai, supervised by Xinxin Feng (Hunan University, China). Cell-to-cell spread in Caco2 cells was performed by France Ourida Manigat, a graduate student in my laboratory. They also wrote the material and methods relative to their experiment parts. The sections 4.8.3 to 4.8.4 by Silei and 4.8.7 to 4.8.8 by France. Professor Francesco Gentile helped us identify structural analogs for our compounds. I performed all the experiments and prepared this manuscript in collaboration with my thesis supervisor. This work has not published prior to the submission of this dissertation. Supplemental figures and tables are available in the Appendix 3.

4.2 Abstract

Infectious diseases caused by enteropathogenic bacteria profoundly impact human health worldwide and are associated with diarrheal diseases, which remain a significant cause of infant mortality. Antibiotic resistance continues to grow, causing significant concerns for treating bacterial infections. Antibiotics can also disrupt the microbiome and adversely affect human health. Many enterobacteria harbor a T3SA to infect their host, whereas the resident microbiota usually lacks T3SA. Therefore, drugs based on T3SA inhibition would be helpful against some infections while reducing antibiotic resistance. This project aims to find chemical compounds capable of inhibiting the T3SA without killing or arresting the growth of commensal bacteria.

We screened compounds from the National Cancer Institute (NCI) and the Food and Drug Administration (FDA) for T3SA inhibition in *Shigella flexneri*. We used the luciferase reporter and dose-response assay in the initial screen and selected two candidate inhibitors with IC50 < 20μM, which we would refer to as C5 and C6. Next, we validated the inhibitory activity of these compounds on secretion by analyzing the secreted fraction by Coomassie Blue and immunoblotting of T3SA substrates.

Our data suggested that C5 reduced the transcription of T3SA-associated genes by inhibiting *virB*. This repression of *virB* is dependent on the small RNA *ryhB* upregulated by C5. Instead, C6 seems to affect the activity or assembly of the T3SA. Further investigations are needed to characterize the mechanism of action of C6. However, this work proves that compounds structurally distinct from the salicylhydene hydrazide (SAH) family could be designed to treat infections caused by gram-negative bacteria by repressing the T3SS.

4.3 Introduction

Shigella spp. causes millions of yearly infections (Watkins and Appiah, n.d.; Kotloff et al., 2018). Like several of its Enterobacteriaceae relative, *Shigella* is among the bacteria for which novel antibiotics are most urgently needed (WHO, 2017). An alternative to antibiotics is antivirulents, which blunt the infectious capacity of microbes without killing them. It is reasoned that this property will reduce the selection pressure on the development of resistance. This and other competitive advantages have sparked interest in antivirulents (Dickey et al., 2017).

The pathogenesis of *Shigella* is characterized by its intracellular niche in the colonic mucosae of humans (Kang et al., 2018). This lifestyle relies on two unique characteristics: 1. a Type III Secretion System (T3SS) that encodes an injectisome [aka the Type III Secretion Apparatus (T3SA)], its substrates and various regulatory factors required for the invasion of the cytosol of host cells; 2. the outer membrane protein IcsA that enables *Shigella's* cytosolic movement required for the extension of the infection foci through cell-to-cell spread (Campbell-Valois and Pontier, 2016; Bajunaid et al., 2020). These components are encoded on the virulence plasmid pINV but have distinct evolutionary origins (Buchrieser et al., 2000). Of the two, the T3SA is the most attractive therapeutic target because it is conserved in numerous bacteria that colonize plants or animals.

Molecules inhibiting the T3SS can intervene at multiple steps (reviewed in (Duncan et al., 2012; McShan and De Guzman, 2015; Hotinger et al., 2021). For example, some members of the salicylhydene hydrazide (SAH) family downregulate the expression of T3SS genes. This, of course, blocks the assembly of the T3SA due to the lack of its proteinaceous parts. How some SAH affects the expression of T3SS genes is still debated and may vary according to the SAH class and bacteria targeted (McShan and De Guzman, 2015). In *Shigella*, the genetic expression of

the T3SS is regulated at the transcriptional level by a cascade comprising the transcription activators VirF, VirB, and MxiE and repressed when the temperature is below 32°C by the nucleoid factor *hns* (Bajunaid et al., 2020). Interestingly, a quinoline was reported to inhibit *Shigella* T3SA by disrupting the DNA binding capacity of VirF and other members of the AraC family (Koppolu et al., 2013; Skredenske et al., 2013). Other environmental factors such as the pH (VirF), iron (VirB), and physical contact with host cells (MxiE) are also implicated in the transcription regulation of the T3SS (Nakayama and Watanabe, 1995; Murphy and Payne, 2007; Broach et al., 2012; Bajunaid et al., 2020; Demers et al., 1998), but small molecules targeting these pathways have not been identified as far as we know. Alternatively, antivirulence may block the assembly of the T3SA. For example, this could prevent the sec-dependent export of parts making the inner or outer rings of the T3SA or the T3SA-dependent export of the early substrates composing the needle (Deng et al., 2017). For example, thiazolidinones are suspected of inhibiting the assembly of the T3SA (Felise et al., 2008). Finally, molecules may prevent the function of crucial parts of the fully assembled T3SA by directly inhibiting the secretion of substrates. Indeed, following assembly and activation, the T3SA secretes middle and late substrates essential to invading the host cells (Deng et al., 2017). The ATPase is likely fuels this phenomenon at the base of the sorting platform and by the proton motive force through the export apparatus that may act as a proton-protein antiporter (Bajunaid et al., 2020). Indeed, inhibitors of the Spa47 ATPase of the T3SA of *Shigella* have been reported (Case et al., 2018, 2020).

Here we used a luminescent screening assay in living *Shigella* to identify candidate inhibitors of the T3SS in a set of bioactive compounds. We identified two molecules with novel structural characteristics that inhibited the T3SS *in vitro*. During host cell invasion, one of these inhibitors was also active in the mid-micromolar range. Characterization of its mechanism of

action indicated it repressed the expression of T3SS genes through the perturbation of the iron homeostasis pathway implicating the small RNA RyhB.

4.4 Results

4.4.1 Screening for small molecules inhibiting the T3SA with the l-TSAR

We previously described a fluorescent transcription-based activation reporter of the T3SA (f-TSAR). We obtained a luminescent derivative reporter based on the *lux* operon from *Photorhabdus luminescens* (l-TSAR) (Figure 4.1A). As expected, when *Shigella* harboring l-TSAR were grown in broth, Δ*ipaD,* which displays constitutively active T3SA (Ménard et al., 1994), emitted light, whereas Δ*mxiE* in which the l-TSAR is inactive, did not. The data indicated the assay had an excellent discrimination power (z around 0.69) and a stable kinetic performance (Figure 4.S.1). We next screened 2,962 small molecules for their T3SA inhibitory potential in 96-well plates containing individual small molecules by directly measuring the light emitted from overnight cultures of the reporter strain Δ*ipaD* l-TSAR (Figure 4.1B). We discarded from further analyses the molecules that reduced the optical density at 600 nm of bacterial culture below 70% of that of the untreated sample. Of the remaining 2913 molecules, 19 reduced luminescence by greater than 70%. Using the l-TSAR, we determined their IC50 and focused on those with IC50< 20 µM. The selected molecules originated from the NCI library and were labeled candidates 1 to 7 (C1-C7) (Figure 4.1C). C1-C4 and C7 were nitrogen-based analogs and shared similarities with known luciferase inhibitors (Auld and Inglese, 2004). On the other hand, C5 and C6 were structurally distinct from the other molecules. C5 is a dipyridyl acetone and a distant curcumin analog (Cao et al., 2012), whereas C6 is a quinone. To distinguish whether the candidates were inhibiting the T3SA or the luciferase used in the screening, we compared the reduction of the light

emitted by Δ*ipaD* harboring the l-TSAR (T3SA dependent) to that of Δ*ipaD* harboring the lux operon put under the control of the *rpsM* promoter (T3SA independent) (Figure 4.1D). The data suggested that only C4, C5, and C6 specifically reduced the signal of the l-TSAR. Thus, we decided to investigate the T3SA inhibitory properties of these molecules. To do this, we tested their capacity to block the secretion of the T3SA in two assays *in vitro*: 1. the Congo red (CR) induced secretion in the WT and T3SA deficient Δ*mxiD* (Allaoui et al., 1993); 2. the constitutive secretion in the Δ*ipaD*. The analyses of the secreted fraction (SF) by Coomassie blue indicated that C5 and C6 decreased protein secretion in both assays (Figure 4.1E).

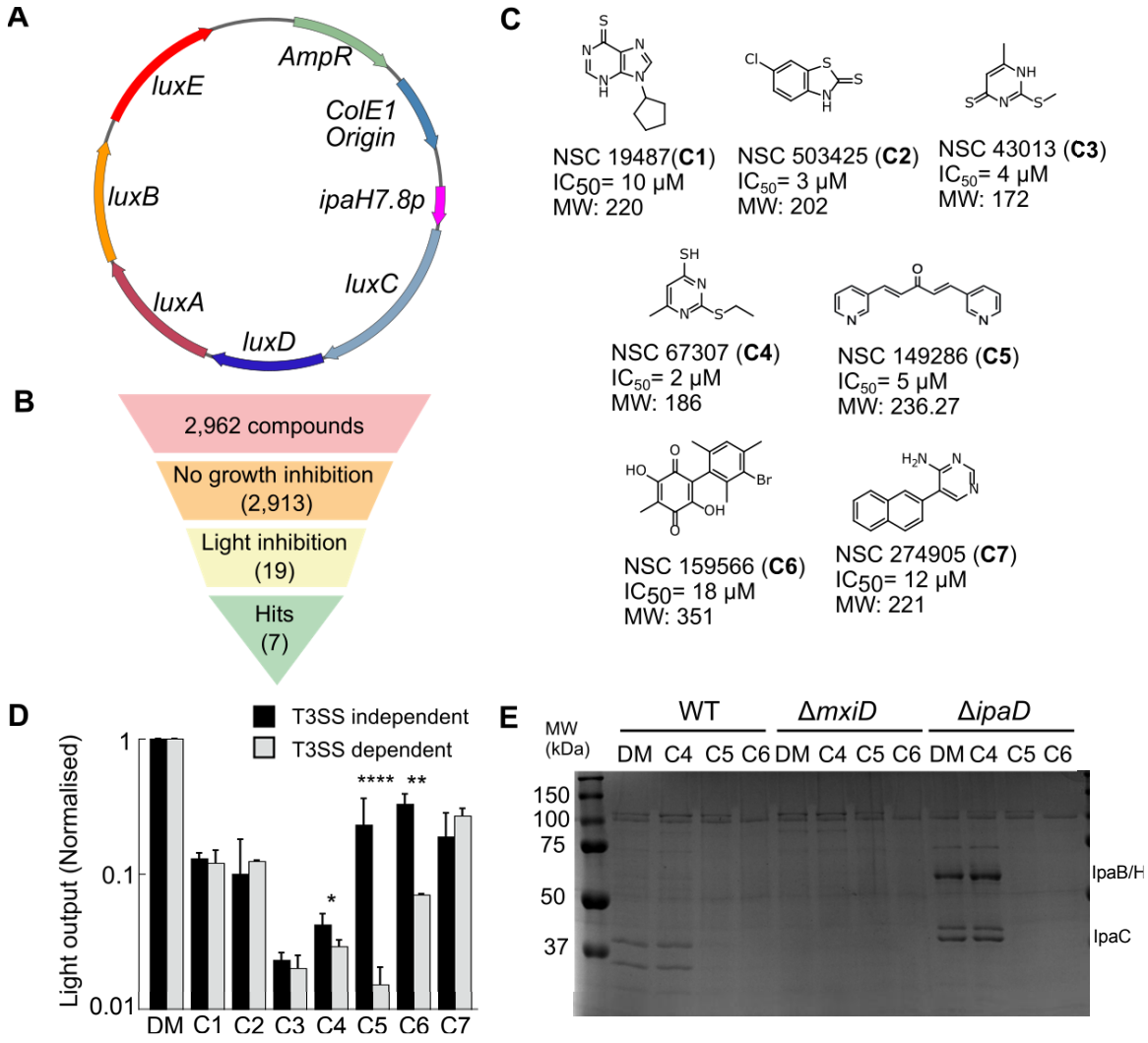


Figure 4. 1. Screening inhibitory compounds. (**A**) Luciferase transcription-based activation reporter (l-TSAR under the secreted gene ipaH7.8 promoter. (**B**) 1584 compounds from the NCI and 1378 compounds from the FDA, which yielded culture with OD600> 0.1, Light inhibition greater than or equal to 70%, and IC50 smaller than or equal to 20 µM. (**C**) Compounds structure after preliminary screening. (**D**) Secondary screening to discard compounds that inhibited the luciferase by comparing the light output of a T3SA-independent promoter to that of a T3SA-dependent promoter. DM (DMSO); The mean values and standard deviations based on three biological replicates, as well as the result of Student's t-tests for unpaired data, are represented; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001. (**E**) Coomassie blue staining to assess the effect of the secondary screening compounds on the secreted fraction of WT, Δ*mxiD,* and Δ*ipaD* strain. Δ*ipaD* secreted fraction was diluted 4-fold.

4.4.2 Inhibition of cell-to-cell spread in Caco2 cells.

Human colon epithelial cells are routinely used to investigate *Shigella* virulence because *Shigella* naturally infects this tissue (Headley et al., 1997)**.** To address the effect of the small molecules C5 and C6 to repress cell-to-cell spread, C2BBe1 (Caco2) cells were infected by *Shigella* WT and its isogenic mutant Δ*icsA*, which is deficient in cell-to-cell spread. Both strains harbored pTSAR3.4t plasmid-producing GFP (T3SA dependent) or mCherry (T3SA independent). Caco2 cells were infected by these bacteria and incubated at 37ºC with 5% CO2 for 30 minutes, then treated with gentamicin containing the appropriate compounds of C5 (12.5µM), C6 (300 µM) or DMSO (0.5%) for 180 minutes.

Our results suggested that C5 reduced Shigella secretion to levels comparable with Δ*icsA,* thus much lower than WT treated with DMSO. In contrast, the infection level measured by mCherry was similar in all treated conditions. (Figure 4.2A-B). Besides, the repression of secretion was also observed with C6, but at much higher concentration than in the in vitro assay (300 µM) (data not shown). This confirmed that C6 had a poor drug potential, and we did not investigate it further.

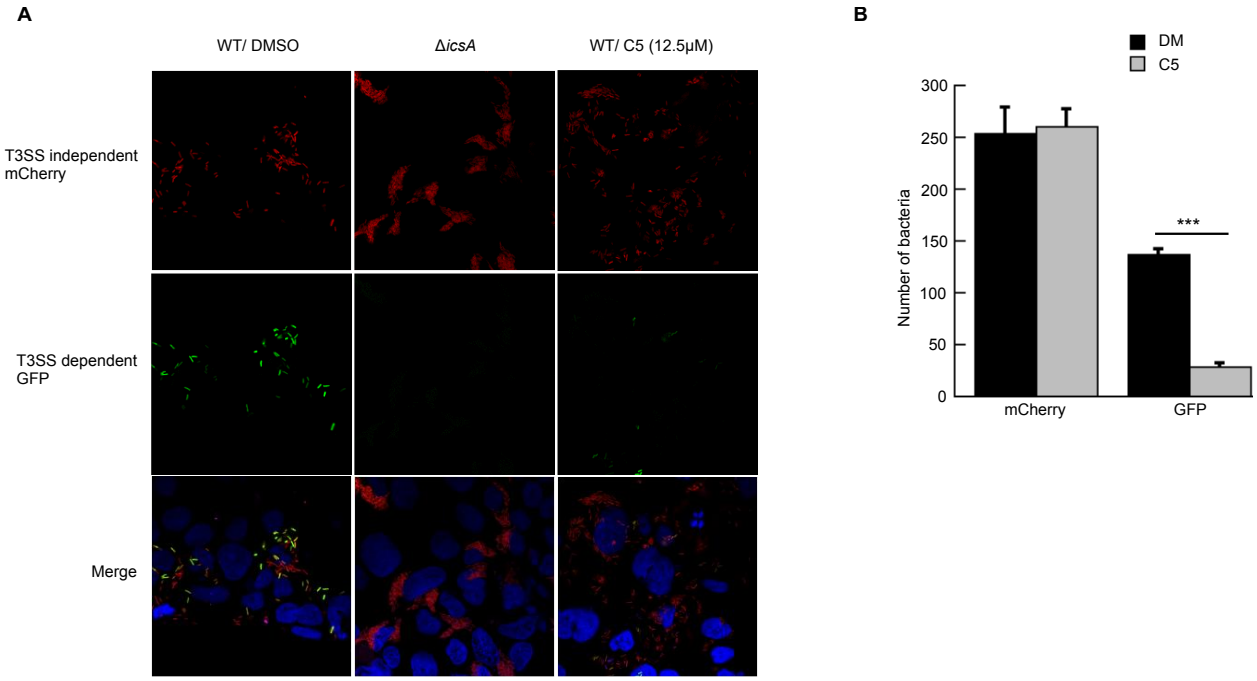


Figure 4. 2. C5 inhibits cell-to-cell spread. (**A**) Image of confocal microscopy; C5 at 12.5µM represses the secretion (GFP) of WT *Shigella* into C2BBe1 (Caco2) cells compared to no spreading mutant Δ*icsA* and WT/DMSO. (**B**) Histogram represented the number of bacteria expressing mCherry or GFP. The mean values and standard deviations based on three biological replicates, as well as the result of Student's t-tests for unpaired data, are represented; *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

4.4.3 Initial assessment of the mechanism of action of the inhibitors

To delineate their mechanism of action, we measured the production and induced secretion of the middle substrates and translocators IpaB and IpaC in WT *Shigella* grown in the presence of C5 and C6 prior to secretion activation by Congo red (Figure 4.3A-B) or in Δ*ipaD* strain (Figure 4.S.2). The production of IpaB and IpaC in the total cell lysate (TCL) was reduced in a dose-dependent manner in the cells treated with C5. This was mirrored by a reduction of IpaBC in the secreted fraction (SF) (Figure 4.3A, Figure 4. S.2A). By contrast, C6 only reduced IpaBC in the secreted fraction (SF) (Figure 4.3B, Figure 4.S.2B). Interestingly, both C6 and C5 were not active

when added after T3SA assembly and only 5-30 minutes prior to adding Congo red (Figure 4.3C). These data suggested that C5 might dampen the T3SS at the expression level, whereas C6 might act post-translationally, perhaps at the T3SA assembly level.

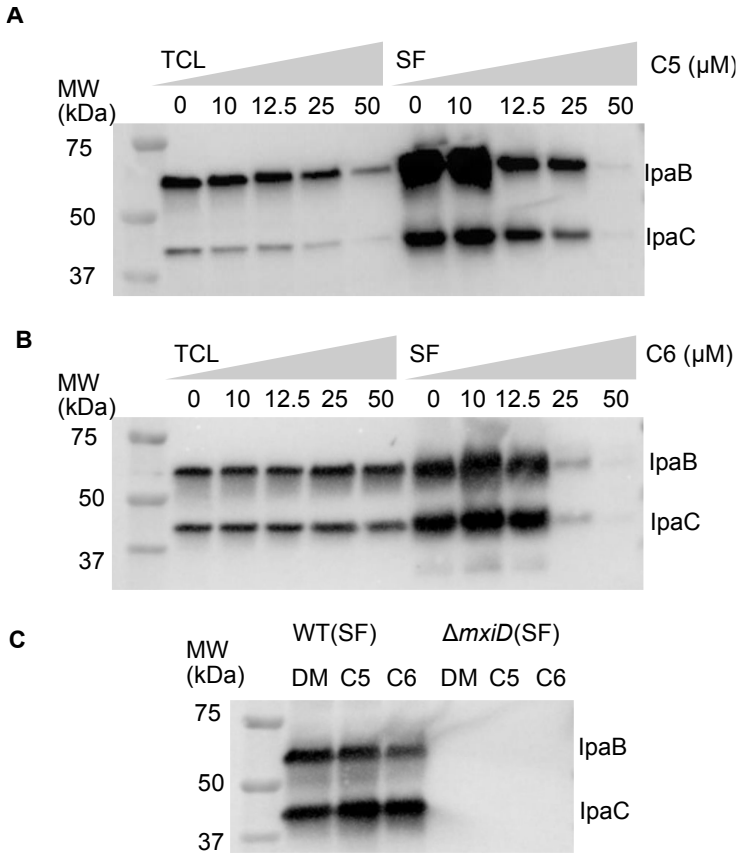


Figure 4. 3. Dose-dependent repression of IpaB and IpaC production and secretion in WT by C5 and C6, respectively. (**A**) Dose-response effect of C5 on Total cell lysate (TCL) and the secreted fraction (SF) when C5 is added before the T3SS assembly with WT strain. (**B)** Dose-response effect of C6 in the secreted fraction when C6 is added before the T3SS assembly with WT strain. (**C**) C5 or C6 does not affect the secretion fraction of IpaB and IpaC when added after T3SS assembly.

4.4.4 C5 downregulates T3SA genes through RyhB and VirB

The T3SS and cell-to-cell spread in *Shigella* are controlled at the gene expression level by a central module composed of transcription activators VirF and VirB (Figure 4.4A) (Bajunaid et al., 2020). Hence, we tested the effect of C5 or C6 on the expression of these transcription activators and some of their targets using qRT-ddPCR (Figure 4.4B). DMSO-treated samples were undistinguishable from those exposed to C6, suggesting this compound did not inhibit the T3SA through transcription, as the data described above suggested. By contrast, C5 reduced the expression of *virB*, *icsA*, *ipaBCD*, mxiD, *spa33,* and *spa40*, whereas *virF* was unaffected, similarly to the housekeeping gene *recA*. We thus reasoned that C5 could either downregulate *virB* through an alternative pathway (Figure 4. 4A) or inhibit the VirF protein. Furthermore, VirF belongs to the ubiquitous AraC/XylS family of transcription activators (Skredenske et al., 2013). We thus evaluated the possibility that C5 could be a general inhibitor of this family. C5 did not affect the AraC arabinose-induced expression of *arap*::sfGFP (Figure 4.S.3).

We observed, however, that C5 did not affect the VirF-dependent increase of the luminescence produced by *icsAp*::*lux* and *virBp*::*lux* in *E. coli* (Figure 4.4C), hence discarding the hypothesis that C5 inhibited VirF. Thus we tested by qRT-ddPCR the effect of C5 on the expression, on the one hand, of *ryhB* and *hns* (Figure 4.4D), and, on the other hand, of *rnaG*, which are respectively negative regulators of the central module and *icsA* (Murphy and Payne, 2007; Broach et al., 2012; Bajunaid et al., 2020). The expression of *hns* and *rnaG* was unchanged or moderately reduced by C5, suggesting they are not involved in its mechanism. By contrast, *ryhB* was activated 22-fold. RyhB is a non-coding RNA that downregulates the expression of *virB* through annealing with a complementary site in its coding sequence (Broach et al., 2012). Furthermore, *ryhB* is in the chromosome and conserved in *E. coli*. This enabled us to decipher the

role of *ryhB* by placing the coding sequence of *virB* under the control of *lacp* in *E. coli*, which is devoid of VirF. Interestingly, C5 reduced the production of VirB from *lacp*::*virB*. As expected, the mutation of the RyhB-binding region in *virB* abolished the effect of C5 (Figure 4.4E).

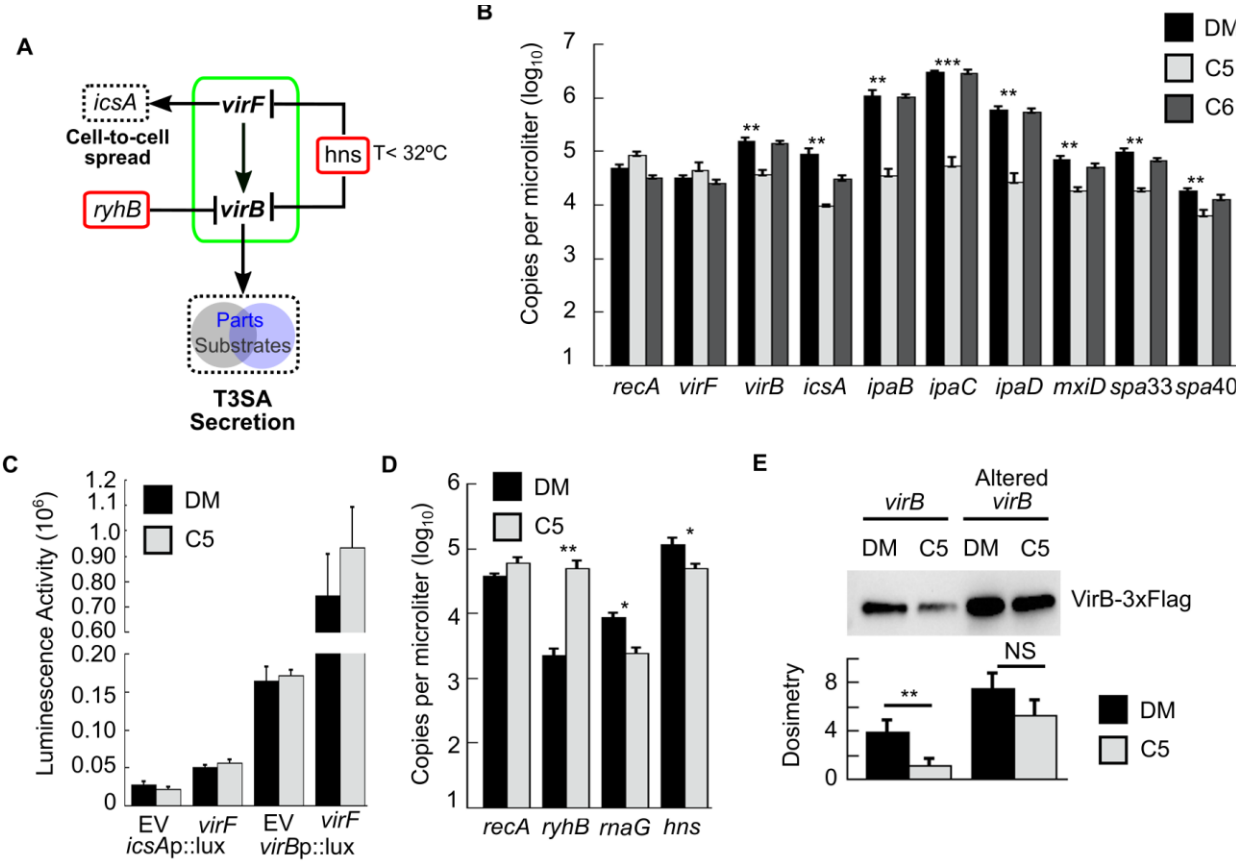


Figure 4. 4. C5 represses gene expression in the T3SA Parts by small RNA RyhB through the *virB* pathway. (**A**) Cascade transcriptional regulation of T3SA. (**B**) C5 downregulates *virB*, *icsA*, and *virB*-regulon genes (*ipa*, *mxi*, and *spa*). (**C**) Repression of *icsA* and *virB* by C5 was promotor independent. (**D**) C5 upregulates and downregulates, respectively, *ryhB* transcript and *rnaG*, *hns* determined by ddPCR analysis. (**E**) C5 reduces the production of VirB, whereas the ryhB binding site altered into *virB* restores VirB production. A histogram plot of the relative output of VirB estimated by dosimetry of WB bands corresponding to biological triplicates. The mean values and standard deviations based on three biological replicates and the result of Student's t-tests for unpaired data are represented; *p<0.05; **p<0.01; ***p<0.001.

4.4.5 C5 is interfering with iron homeostasis

The transcription of *ryhB* is repressed by the ferric uptake regulator (*fur*) in its holo form, i.e., when it is bound to ferrous iron (iron(II) or Fe^{2+}) (Seo et al., 2014). In addition to controlling the expression of one part of its regulon through RyhB, holo-Fur also directly repressed several genes and activated a few genes by antagonizing the *hns* repression (Seo et al., 2014) (Figure 4.5A). Therefore, to determine whether C5 acted on or upstream of *fur* or directly on *ryhB*, we measured by qRT-ddPCR the expression of a set of genes regulated by holo-Fur through those three main regulatory mechanisms (Figure 4.5B). As expected, *sodB,* which is repressed by *ryhB* as *virB*, was downregulated in the presence of C5. Similarly*, ftnA, whose* expression is regulated by the interplay of *hns* and *fur*, was downregulated in the presence of the compound. By contrast, *cirA/iucD,* repressed by holo-Fur, were upregulated in the presence of C5. These data indicate that C5 is acting on or upstream of *fur* in a fashion that phenocopies iron deprivation. Compounds containing two or three dipyridyl groups are iron chelators (Figure 4.5D). Therefore, we wondered whether the pyridyls of C5 were essential for its activity. To test this, we performed a structure-activity relationship with several available analogs off the shelves (Figure 4.5C-D). We found that none of the phenyl (DBA), methoxyphenyl, pyrrole, or thiophene analogs retained the capacity to downregulate the production of IpaBC. This suggested that the pyridyl groups of C5 were essential for its activity. In addition, the nature of the spacer between the pyridyl groups was also crucial as analogs replacing the ketopentadiene spacer of C5 with hydrazones or an amide were also inactive. Finally, bipyridine (BPY) and terpyridine (TPY) containing di- or tripyridyls were active, although less potent than C5. Besides, the activity of phenanthroline, which is a dipyridyl compound, was similar to C5 (Figure 4.S.4). However, it reduced bacterial growth approximately 5-fold, a phenomenon also observed with TPY and, to a lesser extent with BPY. Notably, to reduce the bias

induced by the toxicity of these compounds, the loading of the samples on the gel was normalized with the optical density of each sample.

Interestingly, a dipyridyl ketone close to BPY was inactive, whereas a fluorinated and mildly toxic analog was weakly active. The latter possessed a ketopropene spacer which, albeit smaller, shares some of the properties of C5's spacer. These data suggest that polypyridyl compounds can repress the VirB-dependent production of IpaB and IpaC. Because BPY, TPY, and phenanthroline are known iron(II) chelators (Breuer et al., 1995; Vilà and Walcarius, 2020), we reasoned that C5 might also act through iron chelation. The effect of chelators can be reversed by providing an excess of the titrated metal in the growth medium (Layton et al., 2010). Thus, we checked whether the effect of C5 could be compensated by adding iron to the medium. We observed a partial restoration of IpaBC production in cells treated with C5 and 100 μM ferric (Figure 4.5E) or ferrous iron (Figure 4.S.5). The restoration was shallow at the higher concentration of C5 tested, highlighting the potency of this compound. Iron chelators are used to treat iron overload in humans. Deferiprone, deferasirox, and deferoxamine are most often employed. They act through the chelation of ferric iron in the circulation. Quercetin is a plant flavonoid with antioxidant and iron chelation properties. It also preferred ferric iron and was previously shown to inhibit the T3SS in Salmonella (Tsou et al., 2016). The data indicate that the clinical iron chelators are inactive, whereas C5 is more potent than quercetin (Figure 4.5F). We next wondered whether C5 was endowed with iron chelation. Because pyridyls prefer coordinating ferrous iron [iron(II)], the standard ferrozine assay is an excellent approach to measuring their chelation activity (Dinis et al., 1994). (Figure 4.5G). As expected, BPY displayed strong chelator activity, whereas BPA was inactive. C5, however, was also inactive. This suggested that C5 was not an iron(II) chelator in these conditions. This lack of chelation could be due to the stiffness of

the ketopentadiene spacer, which would prevent the pyridyl groups from coming together to coordinate the iron atom.

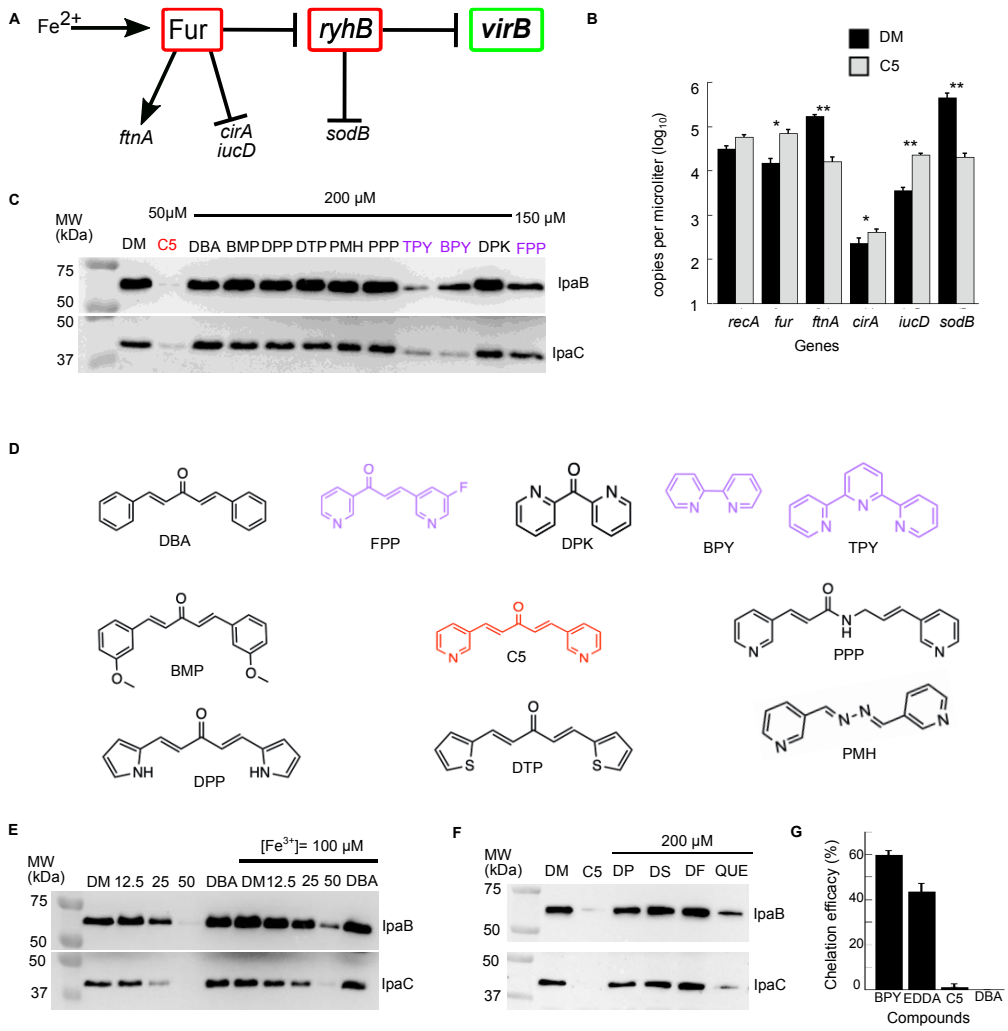


Figure 4. 5. C5 perturbs iron homeostasis. (**A**) Holo-Fur (Fe^{2+}-Fur) represses *ryhB*, *cirA,* and *iucD* genes. (**B**) C5 upregulates *fur*, *cirA*, and *iucD* and decreases *ftnA* and *sodB* gene expression. (**C-D**) Structure-Activity Relationship. WT *Shigella* were cultured in the presence of DM or C5 (50 µM) or 200 µM of DBA, BMP, DPP, DTP, PMH, PPP, TPY, BPY, DPK and 150 µM of FPP (Toxique at 200 µM). C5 is more potent than TPY, BPY, and FPP. (**E)** Partial complementation by iron(III) (100 µM) of IpaBC productions in the presence of gradient concentration of C5. (**F**) Clinical iron chelators are inactive on IpaBC production, whereas QUE has a negligible effect but is less potent than C5. (**G**) Competition of ferrozine with 2,2′-Bipyridine (BPY), Ethylene Diamine-N, N'-Diacetic Acid (EDDA), C5, or DiBenzyldeneAcetone (DBA) to bind Iron(II). The results are representative of three independent experiments. DMSO (DM); Deferiprone (DP); Deferasirox (DS); Deferoxamine mesylate salt (DF); Quercetin (QUE). All abbreviation compounds are listed in the section Materials and Methods.

4.5 Discussion

In this work, we identified two molecules that inhibited the T3SS in *Shigella*. They interfered with the assembly of the functional T3SA through two distinct mechanisms. Both compounds inhibited cell-to-cell spread in human epithelial cells, but C6 was active only at concentrations detrimental to host cells' fitness. For this reason, we focused the rest of this study on the more promising C5 (1E,4E)-1,5-dipyridin-3-ylpenta-1,4-dien-3-one (NSC 149286) that repressed the expression of the regulon of *virB*, which included T3SA parts and substrates such as the translocators IpaB and IpaC. Thus, C5 likely reduced the functional T3SA at the bacterial surface by restraining the production of proteins needed for its assembly and function. We also showed that *virB* is probably repressed by the C5-mediated upregulation of its repressor RyhB. It was reported that the chromosomal small RNA RyhB upregulation generally occurs in iron-restricted environments and represses *virB* and, consequently, genes within the VirB-regulon (Murphy and Payne, 2007; Africa et al., 2011; Broach et al., 2012). In agreement with the notion that C5 interferes with iron homeostasis, we observed that C5 phenocopied the effect of iron deprivation on the expression of several genes of the *fur* operon, including on genes not regulated by RyhB such as *ftnA* and *cirA* (Seo et al., 2014).

We observed that C5 was mildly toxic to human cells at our in vitro assay concentration. We thus used C5 at 12.5 µM during epithelial cell infection and found this treatment almost completely abolished cell-to-cell spread. Although the effect of C5 on the production of IpaB and IpaC is already detectable at this concentration in LB broth, it is much reduced compared to its effect at 50 µM in vitro. For the in vitro experiments, we used the iron-rich LB broth (~17 µM of iron) to grow *Shigella* (Abdul-Tehrani et al., 1999; Yang et al., 2009). Thus, the discrepancy in the concentration of C5 required to observe a maximal effect in both assays may arise from a

combination of factors. The iron requirement for many bacterial pathogens is at least 0.1 μM, but iron bioavailability is approximately 10^{-18} μM in the cytosol of human cells (Andrews et al., 2003; Raymond et al., 2003; Wei and Murphy, 2016), thus much lower than in LB broth.

Furthermore, the expression of invasion genes is already downregulated in cytosolic *Shigella* due to the low concentration of iron in this cellular compartment (Lucchini et al., 2005; Murphy and Payne, 2007). Therefore, it is possible that the poor availability of iron in the cytosol of host cells rendered *Shigella* susceptible to a lower concentration of C5 in this context. Finally, our data suggest that *icsA* was also downregulated by C5, although the mechanism of the repression of *icsA* by C5 is not yet understood (Figure 4.4B). Notably, the eventual downregulation of *icsA* by C5 in the cytosol of infected epithelial cells would further restrict cell-to-cell spread and diminish the signal of the f-TSAR (Campbell-Valois et al., 2014).

The mechanism used by *ryhB* to repress *virB* independently of *virF* relied on their interaction through the base pairing (Broach et al., 2012). By disrupting the *ryhB* complementary binding sites on the coding sequence of *virB* by introducing silent mutations (Broach et al., 2012), we restored the production of VirB in the presence of C5. To decipher whether C5 acted directly on RyhB or through Fur, we investigated the effect of C5 on other genes of the *fur* regulon. We observed that C5 has a widespread impact on several categories of genes in the Fur regulon and was not limited to RyhB and its target genes *virB* and *sodB* (Broach et al., 2012; Seo et al., 2014). For example, C5 led to the repression of *ftnA,* whose expression is repressed by *hns* and derepressed by holo-Fur. In addition, C5 upregulated the expression *cirA* and *iucD*, which are directly repressed by holo-Fur (Porcheron et al., 2014; Seo et al., 2014). However, what is intriguing is the slight upregulation of the gene *fur* in the presence of C5. This small increase may be a compensatory mechanism arising from the disruption of Fur function by C5. These data

suggested that C5 acted through or upstream of Fur. The structural activity relationship indicated that the bipyridyl groups of C5 were essential for its activity and that other bipyridyl or tripyridyl compounds, such as bipyridine and terpyridine, were also capable of reducing the production of IpaB and IpaC. Bipyridine and terpyridine being iron(II) chelators led us to test whether C5 could also be a chelator. According to the ferrozine assay, however, C5 was not an iron(II) chelator in vitro. It was not entirely surprising as the spacing between the pyridyl groups of C5 seems too large for the bidentate intramolecular coordination of iron(II). An alternative mechanism of action that should be tested is the direct binding of C5 to *fur*. It was reported that Fur carries three essential cysteines (Cys93, Cys96, and Cys133), which are required for binding the [2Fe-2S] cluster to Fur (Fontenot et al., 2020; Lill, 2020), a crucial step to obtain holo-Fur. We speculated that C5 could react through the double bonds of carbon in its spacer with one or a combination of Cys93, 96, and 133 in Fur. This could then disrupt the binding of Fur to the [2Fe-2S] cluster, thereby preventing holo-Fur formation. In that circumstance, Fur protein is inactivated directly by C5, which may disrupt the regulation of the Fur regulon, including RyhB. Further work will be needed to determine the complete mechanism of action of C5.

Nevertheless, the data proved that targeting iron homeostasis is a valid pharmacological approach to inhibit the T3SS in *Shigella*. Furthermore, quercetin is a natural product iron chelator that inhibits the T3SS in *Salmonella* (Tsou et al., 2016), suggesting that this approach could be extended to other microbes. Furthermore, iron (III) chelators are already approved to treat human iron overload conditions. Therefore, we wondered whether they could be repurposed to inhibit the T3SS in *Shigella*. Unfortunately, our data suggest otherwise, as they cannot reduce the production of IpaB and IpaC, in contrast with C5 and, in a lesser measure, with quercetin.

In conclusion, C5 has unique properties as a potent dysregulation of iron homeostasis. Indeed, C5 suppresses the expression of genes of T3SA, which is superior to bipyridine, terpyridine, and quercetin effects. However, the challenges for applying C5 as an antivirulence drug still need to be addressed, such as its narrow pharmacological window hinted by its toxicity for human cells and the incomplete delineation of its mechanism of action.

4.6 Significance

Many gram-negative bacteria use the T3SA to invade host cells and cause diseases. Because iron homeostasis regulates the T3SS and other virulence traits, C5 or one of its derivatives may help treat other bacterial infections, the most obvious candidate being *Salmonella*, which also uses a T3SS. This compound does not affect *Shigella's* growth while disarming its virulence in vitro or in human epithelial cells. This strategy for treating infection could circumvent the antimicrobial resistance of these pathogens against traditional antibiotics.

4.7 Materials and Methods

4.7.1 Bacterial strains and media

Shigella flexneri str. M90T and its corresponding isogenic mutants Δ*mxiD* and Δ*ipaD* were previously described (Silué and Campbell-Valois, 2022). *Shigella flexneri sf301* Δ*ipaD* and Δ*mxiE* were used for preliminary screening. *Shigella* strains were routinely cultured in tryptic casein soy agar (TSA), tryptic casein soy broth (TSB), and *E. coli* in Luria-Bertani agar or broth. Unless otherwise indicated, antibiotics were used at the following final concentrations: Ampicillin, 100 µg/ml, Chloramphenicol 30 µg/mL.

4.7.2 Plasmids

The luminescent reporters based on the *lux* operon from *Photorhabdus luminescens* (l-TSAR) were derived from the fluorescent transcription-based activation reporter of the T3SA (f-TSAR) (Campbell-Valois et al., 2014). The lux operon activation was T3SA dependent under the promoter *ipaH7.8* or T3SA independent under the promoter *rpsM*. Also, NEBuilder hifi DNA assembly was used to replace the promoter *ipaH7.8* with the promoter of *icsA* or *virB* to generate derivative l-TSAR *icsAp*::*lux* and *virBp*::*lux,* respectively. The coding sequences of *virB* or *virF* were amplified from *S. flexneri* M90T and cloned by NEBuilder hifi DNA assembly to generate pUC18Δ_pLac_virB_3xFlag and pSU2.1-*virF*-2xMyc respectively. The plasmids pUC18Δ-pLac-3xFlag and pSU2,1-2xMyc were previously described (Silué and Campbell-Valois, 2022). Mutagenesis was performed using primers (HMIO730/HMIO731) to alter the binding site of *ryhB* (Broach et al., 2012) to generate pUC18Δ_pLac_Altered-*virB*_3xFlag. pTSAR3.4t was obtained by subcloning through Gibson assembly the *rpsMp*-mCherry-*ipaH7.8p*-GFPsfm2 fragment from pTSAR1ud2.4s (Campbell-Valois et al., 2014)) obtained by PCR with primers HMIO254 and HMIO255 into the backbone of pUdO1a (Unpublished). The tandem Trp terminator separating *rpsMp*-mCherry and *ipaH7.8p*-GFPsfm2 was replaced by the strong terminator separated by the L3S3P21 terminator (Chen et al., 2013).

All constructs were confirmed by Sanger sequencing (Génome Québec). The oligonucleotides (Integrated DNA Technologies Coralville, IA) used in this study are listed in Table 4.S.1.

4.7.3 Compounds screening

The approximately 3,000 compounds used in the screening were broken down as 1584 compounds from the Diversity Set VII of the National Cancer Institute (NCI) and 1378 from the

Food and Drug Administration-approved compound library (Selleck Chemicals LLC, Houston, TX). An overnight culture of *Shigella* sf301Δ*ipaD* and l-TSAR plasmid was diluted in fresh Tryptic soy broth (TSB) to obtain a 5×10^5 CFU/mL working solution. First, 100 µL of the working solution was added to a sterile 96-well plate. Then, 1 µL of each compound was added to each cell in duplicate, such that the final screening concentration of each compound was 50µM. Three wells with *Shigella sf301ΔipaD* and l-TSAR plasmid were used as a positive control. The wells with *Shigella sf301ΔmxiE* and l-TSAR plasmid were used as negative controls, and wells with TSB only were set as blank. Then, the plate was incubated at 37 °C, 220 rpm. Then the OD_{600nm} and luminescence intensity of each well was determined using a microplate reader after 24 h. The OD_{600nm} values were measured to determine bacterial growth inhibition. Moreover, the luminescence intensity was measured to determine T3SS expression.

Z'-factor : The Z'-factor (a statistical factor used for screening purposes) was calculated using the following equation, where the term $|\mu_{c+} - \mu_{c-}|$ represents the difference between the means of the positive and negative controls and σ_{c+} and σ_{c-} indicates the standard deviations for the positive and negative controls, respectively.

$$Z' - \text{factor} = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{|\mu_{c+} - \mu_{c-}|}$$

The coefficient of variation (CV) was calculated using the equation CV=σ/µ, where σ represents the standard deviation and µ represents the mean.

4.7.4 T3SS expression inhibition assay

An overnight starter culture of *Shigella sf301ΔipaD* and l-TSAR plasmid was diluted 1000-fold in fresh TSB and grown to an OD_{600} value of ~0.3 (approx. 3.5 hours at 37 °C). As a working solution, this log-phase culture was diluted to 5 × 105 CFU/mL. 100µL of this working solution

was transferred into every well in a flat-bottom 96-well plate, except for the first column. Inhibitors were added to the first column at specific starting concentrations (50 µM) with a total volume of 200 µL (diluted with working solution). The inhibitors were then sequentially diluted two-fold across 8 wells. Plates were incubated at 37 °C, shaking at 200 rpm for 24 hours. The luminescence intensity was measure with a microplate reader.

4.7.5 Small molecules

Additional lots of C5 and C6 were reordered from the NCI to perform the validation experiments. The compounds' stock solutions (5 mM) were prepared in dimethyl sulfoxide (DMSO) and stored in a small aliquot at -20ºC. The compounds were added at 50 µM in a culture medium except when indicated otherwise. DMSO was used as a control, adding the same volume as the stock solutions. The chelator compounds and structural analogs of C5 selected were purchased (Sigma-Aldrich, St-Louis, MO) or synthesized on demand (Enamine, Ukraine). Stock solutions (20 mM) were also prepared with DMSO. These latter compounds were added at 200 µM except when indicated. All compounds used in this study are listed: DBA, BMP, DPP, DTP, PMH, PPP, TPY, BPY, DPK, FPP, DP, DS, DF, QUE, EDDA, and DMSO.

4.7.6 RNA extraction, cDNA synthesis, and ddPCR

RNA extraction, cDNA synthesis, and ddPCR were performed as previously described (Silué et al., 2020) with slight modifications. First, samples for RNA extraction were treated with DMSO or Compound C5 (50 µM) or C6 (50 µM).

4.7.7 Cell culture and cell-to-cell spreading assay

C2BBe1(clone of Caco2) cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Wisent, 320-026-CL) supplemented with 20% fetal bovine serum (FBS) (Wisent, 081-105) and 1% Pen Strep (Wisent, 450-200-EL). Coverslips were incubated with 10 µg/mL polylysine in 1xPBS for 1h, washed three times with PBS, and seeded with C2BBe1 cells at a 2.2 x 10^5 per well in 24 wells plate. Cells were grown at 37ºC, 5% CO_2 to 80-90% confluency. Wild-type *Shigella* harboring pTSAR3.4t were grown overnight at 30ºC and sub-cultured for 3 hours in tryptic soy broth (TSB) at 37ºC with shaking, then collected by centrifugation (6000xg, 2 min, RT) and washed with PBS. Bacterial cells were incubated with 10 µg/mL polylysine/PBS for 15 min, washed three times with 1xPBS, and resuspended in DMEM, 20 mM HEPES. C2BBe1 cells were washed three times with DMEM, 20 mM HEPES, and infected with *Shigella* strains of choice at a multiplicity of infection (MOI) of 100. Infected cells were allowed to incubate at room temperature for 10 min; the infecting medium was then removed and replaced with DMEM, 20 mM HEPES, and incubated at 37°C with 5% CO_2 for 30 min. Coverslips were washed three times with PBS, DMEM (1% FBS, 20 mM HEPES, 50 µg/mL gentamicin) with C5 12.5 µM or C6 300 µM was added, and the infected cells were incubated at 37° with 5%CO_2 for 180min.

4.7.8 Imaging of *Shigella* infection and quantification of GFP/mCherry pixel intensities

Coverslips containing infected C2BBe1 cells were washed twice with PBS and fixed with 4% PFA. Cells were permeabilized in PBS, 0.1% Triton X100 and blocked in PBS, 1% BSA (30min, RT). Coverslips were washed thrice with PBS and incubated with DAPI diluted in 1xPBS, 1% BSA (15 min, RT). Stained coverslips were washed thrice with PBS and mounted on microscope slides (VWR) with homemade MOWIOL mounting medium. Samples were imaged with a 63X/1.46 NA oil

objective on a Zeiss LSM880 confocal microscope. Images were analyzed with FIJI (Schindelin et al., 2012). The mean pixel intensity of bacteria in the GFP (TSAR) and mCherry (*rpsM*p) channels were measured. The channels were split, and the RFP channel was used to auto-threshold the image using the MaxEntropy setting and then converted into a binary mask that was next used to quantify the pixel intensity of the original image. Objects with areas smaller than 0.5 μm^2 or larger than 7 μm^2 were discarded.

4.7.9 Total cell lysate and secretion assays

The measures of secretion of T3SS effectors were performed as described (Silué and Campbell-Valois, 2022), except for the following modifications. First, for secretion induced by Congo red (CR) and constitutive secretion assays, overnight cultures were diluted (1/100) in a fresh TSB medium. Then small-molecule compounds were added prior to T3SA assembly or after. Next, the cultures were incubated at 37°C for four hours with shaking at 250 rpm. Next, cell pellets were harvested and resuspended in phosphate-buffered saline (PBS) for secretion induction by CR (Silué and Campbell-Valois, 2022). Next, the constitutive secretion of the Δ*ipaD* strain was performed as in the CR induction protocol but without adding the CR dye (Silué and Campbell-Valois, 2022). Finally, the total cell lysate (TCL) samples and the secreted fraction (SF) were prepared as described (Silué and Campbell-Valois, 2022) previously. The TCL and the SF were analyzed by SDS-PAGE followed by Coomassie staining or Western blotting and immunodetection using mouse and rabbit antibodies directed against IpaC (1/2000) and IpaB (1/4000), respectively. These antibodies IpaBC were a kind gift from Armelle Phalipon.

4.7.10 Detection of the production of VirB by Immunoblotting

E. coli DH10B strain was transformed with pUC18Δ-*virB*-3xFlag or pUC18Δ-Altered_*virB*-3xFlag containing the coding sequence of *virB* or altered binding site of ryhB, respectively. These strains were cultured overnight, and subcultures (1/100) were performed in the presence of DMSO (DM) or C5 at 37ºC for 4h with shaking (250 rpm). The total cell lysate (TCL) was prepared with normalized OD_{600} and loaded in SDS-PAGE, followed by western blotting with an anti-Flag antibody (Silué and Campbell-Valois, 2022).

4.7.11 Iron chelating assay

100 µM of C5; BPY; EDDA, and DBA were incubated with 20 µM of Fe^{2+} (ammonium Iron (II) sulfate) in 5% ammonium acetate, pH 6.9. The reaction was initiated by adding 100µM ferrozine and, after 10 minutes of incubation at room temperature to let the mixture reach equilibrium, the absorbance at 562 nm was measured (Dinis et al., 1994; Viollier et al., 2000). It is well known that BPY and EDDA are iron chelators and were used as positive controls, but DBA, not an iron chelator, was used as a negative control. The percentage of inhibition of the ferrozine–Fe2+ complex formation was calculated as previously determined (Dinis et al., 1994).

Acknowledgements

We acknowledge the role of the Faculty of Science's support staff in realizing this work, particularly those working at the Core Molecular Biology and Genomics Laboratory (Hailey Quigley, Gary Hatch) by providing the ddPCR reagents. We thank JW Keillor, JP Pezacki, and their group for generously sharing their equipment. This work was funded by the Faculty of Science of Ottawa and the Discovery grant no. 05587 from the Natural Sciences and Engineering

Research Council of Canada (NSERC), the CFI John R. Evans Leaders fund no. 34789, and the CIHR project grant no. 159517. NS received a stipend from the NSERC CREATE program Technologies for Microbiome Science and Engineering (TECHNOMISE).

Chapter 5: General Discussion and Future Direction

5.1 MxiE mediated genes regulation in the chromosome of *Shigella* and *E. coli*

With studies detailed in Chapter 2 of this document, we investigated the T3SA regulon in *Shigella flexneri* by RNA-Seq (Silué et al., 2020). We compared the differential gene expression in *Shigella flexneri str.* M90T WT to its isogenic mutant Δ*ipaD*. This mutant is a constitutively secreting strain mimicking the "On-state" encountered during the invasion of human cells whereas the WT is not secreting and is therefore said to be in the "Off-state" in the absence of host cells. At the time of preparing my thesis, the article presented in Chapter 2 was cited in five publications that I did not co-author, e.g. (Cervantes-Rivera et al., 2020; Daswani et al., 2020; Solaiyappan Mani and Reinhold-Hurek, 2021; McKenna et al., 2022; Hall et al., 2022). Some of these studies will be discussed below.

The pioneering method to investigate the transcriptome of bacteria was DNA microarray, as illustrated in a seminal study in *S. flexneri* (Le Gall et al., 2005). The disadvantages of the microarray method are it requires knowledge of the coding sequences to detect transcripts, suffers from a lower sensitivity than RNA-Seq, and from possible specificity issues such as cross-hybridization and non-specific hybridization (Solaiyappan Mani and Reinhold-Hurek, 2021). The RNA-Seq overwhelms drawbacks of the microarray by analyzing both known and unknown transcripts on a large scale at high sensitivity and specificity (Solaiyappan Mani and Reinhold-Hurek, 2021). In addition, RNA-Seq, although facilitated by full genome sequencing that allows direct mapping of the transcripts, can also work with de novo assembly (Tjaden, 2015), which is amenable to study species with poorly characterized genomes. So, RNA-Seq has become the gold

standard method to analyze transcriptomes in bacteria, including for complex microbiota (Giannoukos et al., 2012).

The study presented in Chapter 2 is among the pioneer to investigate the transcriptome of *Shigella flexneri* by RNA-Seq (Silué et al., 2020). Our data unveiled other details and regulatory features in addition to previous data from microarray by examining the T3SA regulon encoded in the chromosome and the virulence plasmid (pINV) simultaneously. In addition to canonical pINV genes and seven *ipaHs* in the chromosome with homology with *ipaHs* in the pINV, three uncharacterized chromosomal genes *icaT* (gem1), gem2 and *icaR* (gem3) were upregulated in the "On-state". Interestingly, they did not share any homology with genes from pINV. Specifically, our data enriched the number of the T3SA substrates, and the most salient discovery of this study was the validation of *icaT* (gem1) and *icaR* (gem3) as novel members of the T3SS/MxiE regulon as they harbored a functional MxiE box in contrast to gem2. Indeed, this is the first research revealing that genes in the chromosome without homologs in the virulence plasmid carried an active MxiE box. In addition to these upregulated genes, we observed that 86 genes were downregulated in the constitutively secreted strain Δ*ipaD*. Additional studies must be realized to determine the significance of those downregulated genes. These investigations are rendered difficult by the indirect effect on the growth rate of the knockout of *ipaD* and may be better addressed by an *in vivo* model of infection. Also, in this chapter, we characterized six operons activated in a MxiE-dependent manner by RT-PCR and suggested that the expression of *ospD2*, as suggested previously for *phoN2* (Santapaola et al., 2002; Le Gall et al., 2005), is regulated by two independent promoters; a first promoter whose activation is MxiE-dependent, and a second promoter whose activation is independent of MxiE. The E3 ubiquitin ligase-encoding *ipaH1.4* is identical to *ipaH2.5* and are part of the operon O7 and O3 (Figure 2.3.A), respectively, but the

corresponding protein IpaH1.4 was produced at higher levels than IpaH2.5 (Silué et al., 2020), likely due to the presence of insertion sequences between *ospE2* and *ipaH2.5*. Indeed, by increasing the distance between the coding sequence of *ipaH2.5* and its promoter, these insertions likely induced premature transcription termination. Indeed, *Shigella* expressing both E3 ubiquitin ligases or only *ipaH1.4* could degrade HOIP, whereas *ipaH*2.5 could not (De Jong et al., 2016). This was surprising has both IpaHs were active when overexpressed in mammalian cells. Our data suggest that this discrepancy was due to the poor endogenous production of *ipaH2.5*.

Hall et *al*. have examined the transcriptome of the "Off-state" and "On-state" in *Shigella flexneri*, and they discovered three genes in the chromosome that were upregulated in the "On-state," harboring each an active MxiE box (Hall et al., 2022). This study also found that *icaT* (gem1) and *icaR* (gem3) were regulated by MxiE, as we had reported. They also studied a third gene named *yccE* that was not reported in our study. In light of these findings, we reinspected the raw data provided in the supporting information of the study presented in Chapter 2 (Silué et al., 2020). Among genes that were below the cutoff q-value, we found *yccE* as well as two other chromosomal genes forming an operon (*yhaBC*), which were upregulated in the "On-state." This uncharacterized operon harbored a MxiE box similar to that of *icaT*. *yhaC* is frameshifted in the strain M90T and could be a pseudogene. Since the first of six internal stop codons occur at the 239th codons, *yhaC* might still encode a stable protein similarly to what we have reported for *icaR*, which is also frameshifted in M90T. On the other hand, the 179 codons-long coding sequence of *yhaB* is complete. As previously reported for *icaT*, *icaR* and *yccE* (Silué and Campbell-Valois, 2022; Hall et al., 2022), *yhaB* and *yhaC* are also widespread in many *E. coli* phylogroups in which they have conserved their MxiE box. Further investigations are needed to characterize the MxiE box of *yhaBC* and to determine whether the proteins encoded by *yccE* and *yhaBC* are T3SA

substrates. The weakness of our RNA-Seq analyses was to use a single biological replicate. This likely led to reducing q-values for lower expressed genes such as *yccE* and *yhaBC*. Indeed, these genes displayed q-values above the suggested significance threshold <0.1, which led us to discard them from further analyses at the time. This is clearly a drawback of using one replicate of RNA-Seq, which cannot be rigorously addressed solely by confirming the data by qRT-ddPCR as we did for *icaR* and *icaT*.

In Chapter 3, we confirmed that gem1 and gem3 were members of T3SA substrates by demonstrating they shared the common characteristics of canonical late substrates B. Considering these findings, we renamed gem1 as *icaT*, standing for invasion chromosome antigen with homology for a transposase, and gem3, as *icaR,* standing for invasion chromosome antigen with homology for a transcription regulator. The genes encoding canonical late substrate B are regulated by an upstream MxiE Box located immediately upstream of the coding sequence as for *icaT* or *icaR* or further away when the gene is found in an operon. The MxiE regulon in the chromosome of *Shigella flexneri* is composed of ten genes with seven *ipaHs* and three novel partly characterized genes (*icaR, icaT, yccE*) without any homolog in the pINV (Silué et al., 2020; Silué and Campbell-Valois, 2022; Hall et al., 2022). As noted above, *yhaBC* should be considered as additional candidates to the MxiE regulon. Homologs of *icaR* and *icaT* were also found in many *E. coli* phylogroups (A, B1, D, E), but were absent from the B2 phylogroup. As reported previously, the B2 group seems older than the A and the B1 groups, which have more recently diverged (Le Gall et al., 2007; Gonzalez-Alba et al., 2019). Accordingly, we hypothesized that *icaR* and *icaT* moved to the genome post-radiation of phylogroup B2. As discussed in Chapter 3, more uncertainties remain concerning *icaR* as its presence in the *E. coli* lineage seems more ancient than *icaT*.

E. coli and pINV cured *Shigella* such as strain BS176 are devoid of endogenous *mxiE* and *ipgC*, which in effect silenced genes whose expression is regulated by a MxiE-box. *E. coli* MG1655, *E. coli* ATCC 43888 and *Shigella* str. BS176 were transformed with plasmid harboring *mxiE* or *ipgC* under the control of the promoter Lac. The strains with both plasmids displayed increased expression *icaR* and *icaT* (Silué and Campbell-Valois, 2022). Furthermore, it was reported that *hns* represses the MxiE regulon, including genes *yccE*, *icaT (yfdF*), *and icaR (yjgL)* (Lamberte et al., 2017; Hall et al., 2022). Therefore, according to these papers and our results, one can conclude that the loss of *hn*s repression, whether it be through genetic manipulations, the production of MxiE and IpgC or an alternative mechanism, would result in the activation of the MxiE regulon, including *icaR*, *icaT* and *yccE* (Lamberte et al., 2017; Hall et al., 2022; Silué and Campbell-Valois, 2022).

These results paved the way to interrogate the evolutionary aspect of *Shigella* in the *E. coli* lineage. It was indicated that new genes horizontally transferred to a receptor genome undergo an amelioration process over time which may conduct in convergence toward the GC content of the host genome (Lawrence and Ochman, 1997). As a result, the base composition of foreign genes converges to the new host base proportion slowly over time. Recently acquired genes, however, retain the base composition of the donor, a way of labeling them as foreign genes (Lawrence and Ochman, 1997). Indeed, *icaR* and *icaT* have a GC content of around 35%, like the genes of pINV, in contrast to neighboring genes in the chromosome, which have a GC content of about 50%. Therefore, we speculated that the transfer of *icaR* or *icaT* from pINV, an ancestor thereof or an unrelated mobile element to the chromosome preceded the emergence of the modern pINV but is sufficiently recent to have maintained a low GC content. Additional investigations will be required to precisely estimate the transfer period of these genes.

From the estimations of the phylogeny of *Shigella* based on all the genome or a set of core genes emerged two opposites hypotheses (Sims and Kim, 2011). Regarding the whole genome approach, a commensal *E. coli* acquired pINV by horizontal transfer before progressively adopting a pathogenic lifestyle (Sims and Kim, 2011). It was reported that the acquisition of pINV was combined with the loss of some metabolic functions encode in the chromosome of *Shigella* or EIEC (Maurelli et al., 1998; Lan et al., 2001; Prosseda et al., 2012). These genes deletions adapting the metabolism of *Shigella* to pathogenesis were confirmed by the observation that the reinsertion of the gene *cadA* to restore the Lysine decarboxylase (LDC) activity in *Shigella flexneri* attenuated virulence (Maurelli et al., 1998). This hypothesis raises the question of interaction between pINV and the chromosome of *Shigella* throughout evolution. By contrast, based on the core genes approach, commensal *E. coli* and *Shigella* derived from the same ancestor (Escobar-Páramo et al., 2003, 2004). By comparing the phylogeny of *Shigella* and EIEC, which both carried a pINV, these authors inferred that *E. coli* commensal strain emanated from a common ancestor with *Shigella*, then the commensal lost the pINV and recovered the metabolic function previously lacking in the genome of the ancestor harboring pINV (Escobar-Páramo et al., 2003). The mechanism involved in the loss or acquisition of pINV and the relation between the chromosome and pINV remains unknown.

The first model of the evolutionary origin of pINV states that it derived from a common ancestor of *Shigella* and EIEC before their divergence into the different subgroups. This model is supported by the observation that the pINV plasmid is not self-transmissible by conjugation due to a large deletion in the *tra* region locus, which is responsible for forming the pilus required for conjugation (Escobar-Páramo et al., 2003; Pasqua et al., 2017). However, the second and most favored model is that it was acquired multiple times by convergent evolution from different

ancestral *E. coli* strains. This view is supported by phylogenetic analyses of chromosomal genes and genome comparisons, showing a high diversity of *Shigella* and EIEC pathovars (Pupo et al., 2000; Lan and Reeves, 2002; The et al., 2016; Pilla et al., 2022). Two forms of pINV named A and B were described based on the analysis of three virulence genes located in the entry region (*ipgD*, *mxiA*, and *mxiC*). The authors showed that all strains in the phylogenic cluster 1 have the pINV A form, whereas those in cluster 3 have pINV B, suggesting they derived from distinct ancestors, respectively (Lan et al., 2001). The difference in chromosomal location of *icaR* and *icaT* and the conservation of their MxiE box across the *E. coli* lineage support the notion that they were transferred before the acquisition of the modern pINV in *Shigella*. Based on our data, we can infer the order in which *icaR* and *icaT* were transferred to the chromosome. The most recent transfer seems to be *icaT* because it is absent in all the B2 *E. coli* phylogroup strains that we have inspected, suggesting it was transferred after the radiation of the B2 phylogroup (Silué and Campbell-Valois, 2022). This is a clear indication that *icaR* and *icaT* were acquired independently, although we cannot exclude that they came from the same mobile genetic element. In fact, the high conservation of their MxiE box suggests it may be the case.

The acquisition of properties such as the presence of a MxiE box in genes like *icaR*, *icaT*, *yccE*, and *yhaBC* by commensal or pathogenic strains of *E. coli* may have been directed by an evolutionary selection process rather than by chance. Did this happen through a one-step horizontal transfer of these genes, including the MxiE-box, or through two distinct transfers comprising the coding sequence of these genes and then their respective MxiE box? We favor the former hypothesis for its simplicity. This led us to speculate that these four genes had co-evolved with *mxiE* and *ipgC* or homologs thereof prior to their transfer to the chromosome. The absence of any trace of these genes in modern pINV suggests indeed that they likely originated from an extinct

version of pINV or an unrelated mobile element. Further investigations on these genes and potentially other genes sharing the same properties will be required to answer these interesting questions.

The data presented in Chapters 2 and 3 provide the first evidence of MxiE-mediated gene regulation in the chromosome of *Shigella flexneri* without homology in the pINV genes. These studies contributed to borne-out *Shigella* as an *E. coli*'s pathovar and enrich the understanding of the relationship between the pINV and the chromosome of *Shigella*.

5.2 Small molecules repress T3SA by inhibiting the secretion of proteins.

In chapter 4, we presented two small molecules (C5 and C6) capable of interfering with the T3SA and of inhibiting the secretion of proteins in vitro and in tissue culture cells similarly to the family of salicylhydene hydrazides family (McShan and De Guzman, 2015). McShan et *al*. also mentioned the importance of transcription factors as putative drug target (McShan and De Guzman, 2015). Likewise, C5 gives us an unprecedented mechanism to repress the T3SS at the transcriptional level by downregulating the expression of *virB* in a *virF*-independent manner. It is the first time it was established that a small molecule targets the regulator *virB* through the small noncoding RNA RyhB leading to the repression of the T3SA in *Shigella*. We also established that C5 acted through the upregulation of the expression of *ryhB*. Indeed, the resulting RNA RyhB directly bind to a partly complementary region that was previously identified in base residues 403 to 420 in the ORF of *virB* (Broach et al., 2012). In this model, RyhB would gain access to the ORF of *virB* during the dynamic opening of DNA bubbles, a.k.a. DNA “breathing” similar to protein gain access to the double-stranded (ds) DNA interior (Beyerle et al., 2021). The binding of RyhB

to *virB* DNA prevents the polymerase from binding to *virB*, and this interference conducts to repress its transcription.

The upregulation of *ryhB* depends on the low iron environment in classical conditions. Our data also showed that C5 is not an iron(II) chelator according to the ferrozine assay. The master controller of iron homeostasis is the ferric uptake regulator (Fur), which acts as a repressor when bound to iron(II). By qRT-ddPCR, we revealed that C5 targeted many genes regulated by *fur*, suggesting that C5 results in a general perturbation of iron homeostasis. Further research is required to show whether C5 might directly repress holo-Fur. For example, the conjugated double bonds of C5 could react with cysteine residues in the [2Fe-2S] cluster essential for Fur activity (Fontenot et al., 2020). Alternatively, these double bonds may make C5 act as a prodrug, on the condition that they could be reduced upon entry into the bacteria (Stella, 2004; Brewster et al., 2019; Nasibullin et al., 2022). A C5 derivative with a spacer composed solely of C-C single bonds may be sufficiently flexible to enable the chelation of iron by the two pyridyl groups. Further studies are needed to decipher the mechanism of action of C5.

C5 also reduced the expression of *icsA* gene implicated in cell-to-cell spread. The mechanism of this repression is not well established. Intriguingly, a poster presentation dating back to 2010 argued that *icsA* could be inhibited by RyhB, but these findings have not yet been reported in a peer-reviewed publication. By nucleotide alignment, we found putative binding sites of RyhB into the coding sequence of *icsA,* but we have not been able to test their role in the regulation of *icsA*.

We evaluated C5 and C6 medicinal properties with the software ADMETlab 2.0 (Xiong et al., 2021). This software predicts absorption, distribution, metabolism, excretion, and toxicity (ADMET) of compounds. The results indicated that C5 matched or exceeded all the essential

criteria endowing a molecule with drug-like properties in contrast to C6. The measure of drug likeness is based on the concept of desirability called Quantitative Estimate of Drug-likeness (QED) (Bickerton et al., 2012; Xiong et al., 2021). QED scores greater than 0.67 suggest drug likeness, whereas lower values indicate the presence of non-drug-like features (Xiong et al., 2021). The QED of C5 at 0.77 indicates an attractive drug likeness, whereas that of C6 at 0.43 highlights its poor drug potential, similarly to the score obtained with another software (Song et al., 2020) . C5 also satisfied the Lipinski, Pfizer, GSK, and Golden Triangle Rule, but C6 was rejected by the GSK Rules. These predictions indicated that one could improve the pharmacological features of C5 whereas this would be more challenging for C6. On this regard, we started a collaboration with Christopher Boddy to synthetize analogs of C5. We have already received the first few analogs, which should allow us to further the structural activity relationship reported in Chapter 4.

Taken together, the discovery of C5 highlights iron homeostasis as a relevant target to tackle bacterial pathogens and specifically to downregulate the expression of the T3SS in *Shigella*. The superficial analyses of the properties of C5 suggest it is a more potent dysregulator of iron homeostasis in *Shigella* than known iron(II) chelators or iron(III) chelators used in the clinic to treat iron overload. Again, this supports the notion that C5 has a unique mechanism of action to dysregulate iron homeostasis. Thus, C5 might be used to target other pathogens, but at the same time, this might come at the cost of unwanted effects on other microbes. The next steps for C5 are to test it against the T3SS of *Salmonella* and, in the medium term, to assess whether it would be reasonable to test its efficacy to curtail the infection in animal models. The main goal of antivirulence drugs is to disarm harmful bacteria without severely impacting the microbiota. We have also established a collaboration with Daniel Figeys to measure the effect of C5 on the human microbiota.

In the coming months, I hope to complete the experiments required for the publication of this study. First, we would like to complete the structural activity relationship of C5. Finally, we would also like to test whether *Shigella* strains lacking *ryhB* or a RyhB binding site in *virB* are resistant to C5 in host cells. This would allow us to demonstrate that C5 acts solely through the action of *ryhB* on *virB* or on additional targets as well, such as *icsA*, to dampen the virulence of *Shigella*.

Nonetheless, the data presented in Chapter 4 described a potent and novel approach to curbing the virulence of *Shigella*. In addition, this study contributed to the knowledge on the regulation of the T3SS. Future works may cement the drug targeting of iron homeostasis as an alternative to classical antibiotics for treating or preventing shigellosis and perhaps other infectious diseases.

References

Abdul-Tehrani, H., Hudson, A. J., Chang, Y.-S., Timms, A. R., Hawkins, C., Williams, J. M., et al. (1999). Ferritin Mutants of *Escherichia coli* Are Iron Deficient and Growth Impaired, and *fur* Mutants are Iron Deficient. *J Bacteriol* 181, 1415–1428. doi: 10.1128/JB.181.5.1415-1428.1999.

Abram, K., Udaondo, Z., Bleker, C., Wanchai, V., Wassenaar, T. M., Robeson, M. S., et al. (2021). Mash-based analyses of Escherichia coli genomes reveal 14 distinct phylogroups. *Communications Biology* 4. doi: 10.1038/s42003-020-01626-5.

Africa, L. A. A., Murphy, E. R., Egan, N. R., Wigley, A. F., and Wing, H. J. (2011). The iron-responsive Fur/RyhB regulatory cascade modulates the Shigella outer membrane protease icsp. *Infection and Immunity* 79, 4543–4549. doi: 10.1128/IAI.05340-11.

Alberta Health Services (2023). Shigella outbreak in Edmonton Zone reopened. *Alberta Health Services*. Available at: https://www.albertahealthservices.ca/news/Page17209.aspx [Accessed April 25, 2023].

Ali, S. S., Xia, B., Liu, J., and Navarre, W. W. (2012). Silencing of foreign DNA in bacteria. *Current Opinion in Microbiology* 15, 175–181. doi: 10.1016/j.mib.2011.12.014.

Allaoui, A., Sansonetti, P. J., and Parsot, C. (1992). MxiJ, a lipoprotein involved in secretion of Shigella Ipa invasins, is homologous to YscJ, a secretion factor of the Yersinia Yop proteins. *Journal of Bacteriology* 174, 7661–7669. doi: 10.1128/jb.174.23.7661-7669.1992.

Allaoui, A., Sansonetti, P. J., and Parsot, C. (1993). MxiD, an outer membrane protein necessary for the secretion of the Shigella flexneri Ipa invasins. *Mol Microbiol* 7, 59–68. doi: 10.1111/j.1365-2958.1993.tb01097.x.

Altae-Tran, H., Kannan, S., Demircioglu, F. E., Oshiro, R., Nety, S. P., McKay, L. J., et al. (2021). The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* 374, 57–65. doi: 10.1126/science.abj6856.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2.

Andrews, S. C., Robinson, A. K., and Rodríguez-Quiñones, F. (2003). Bacterial iron homeostasis. *FEMS Microbiol Rev* 27, 215–237. doi: 10.1016/S0168-6445(03)00055-X.

Ashida, H., Toyotome, T., Nagai, T., and Sasakawa, C. (2007). Shigella chromosomal IpaH proteins are secreted via the type III secretion system and act as effectors. *Molecular Microbiology* 63, 680–693. doi: 10.1111/j.1365-2958.2006.05547.x.

Auld, D. S., and Inglese, J. (2004). "Interferences with Luciferase Reporter Enzymes," in *Assay Guidance Manual [Internet]* (Eli Lilly & Company and the National Center for

Advancing Translational Sciences). Available at: https://www.ncbi.nlm.nih.gov/books/NBK374281/ [Accessed May 11, 2023].

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373, 871–876. doi: 10.1126/science.abj8754.

Bahrani, F. K., Sansonetti, P. J., and Parsot, C. (1997). Secretion of Ipa proteins by Shigella flexneri: inducer molecules and kinetics of activation. *Infect Immun* 65, 4005–4010. doi: 10.1128/iai.65.10.4005-4010.1997.

Bajunaid, W., Haidar-Ahmad, N., Kottarampatel, A. H., Manigat, F. O., Silué, N., Tchagang, C. F., et al. (2020). The T3SS of Shigella: Expression, structure, function, and role in vacuole escape. *Microorganisms* 8, 1–28. doi: 10.3390/microorganisms8121933.

Barry, E. M., Pasetti, M. F., Sztein, M. B., Fasano, A., Kotloff, K. L., and Levine, M. M. (2013). Progress and pitfalls in Shigella vaccine research. *Nat Rev Gastroenterol Hepatol* 10, 245–255. doi: 10.1038/nrgastro.2013.12.

Basta, D. W., Pew, K. L., Immak, J. A., Park, H. S., Picker, M. A., Wigley, A. F., et al. (2013). Characterization of the ospZ promoter in Shigella flexneri and its regulation by VirB and H-NS. *J Bacteriol* 195, 2562–2572. doi: 10.1128/JB.00212-13.

Beloin, C., Deighan, P., Doyle, M., and Dorman, C. J. (2003). Shigella flexneri 2a strain 2457T expresses three members of the H-NS-like protein family: Characterization of the Sfh protein. *Molecular Genetics and Genomics* 270, 66–77. doi: 10.1007/s00438-003-0897-0.

Beloin, C., and Dorman, C. J. (2003). An extended role for the nucleoid structuring protein H-NS in the virulence gene regulatory cascade of Shigella flexneri. *Molecular Microbiology* 47, 825–838. doi: 10.1046/j.1365-2958.2003.03347.x.

Bernardini, M. L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M., and Sansonetti, P. J. (1989). Identification of icsA, a plasmid locus of Shigella flexneri that governs bacterial intra- and intercellular spread through interaction with F-actin. *Proceedings of the National Academy of Sciences* 86, 3867–3871. doi: 10.1073/pnas.86.10.3867.

Bernhofer, M., Dallago, C., Karl, T., Satagopam, V., Heinzinger, M., Littmann, M., et al. (2021). PredictProtein - Predicting protein structure and function for 29 years. *Nucleic Acids Research* 49, W535–W540. doi: 10.1093/nar/gkab354.

Beyerle, E. R., Dinpajooh, M., Ji, H., von Hippel, P. H., Marcus, A. H., and Guenza, M. G. (2021). Dinucleotides as simple models of the base stacking-unstacking component of DNA 'breathing' mechanisms. *Nucleic Acids Research* 49, 1872–1885. doi: 10.1093/nar/gkab015.

Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., and Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature Chem* 4, 90–98. doi: 10.1038/nchem.1243.

Bongrand, C., Sansonetti, P. J., and Parsot, C. (2012). Characterization of the promoter, MxiE box and 5′ UTR of genes controlled by the activity of the type iii secretion apparatus in Shigella flexneri. *PLoS ONE* 7, 19–21. doi: 10.1371/journal.pone.0032862.

Breuer, W., Epsztejn, S., and Cabantchik, Z. I. (1995). Iron Acquired from Transferrin by K562 Cells Is Delivered into a Cytoplasmic Pool of Chelatable Iron(II). *Journal of Biological Chemistry* 270, 24209–24215. doi: 10.1074/jbc.270.41.24209.

Brewster, R. C., Suitor, J. T., Bennett, A. W., and Wallace, S. (2019). Transition Metal-Free Reduction of Activated Alkenes Using a Living Microorganism. *Angewandte Chemie International Edition* 58, 12409–12414. doi: 10.1002/anie.201903973.

Broach, W. H., Egan, N., Wing, H. J., Payne, S. M., and Murphy, E. R. (2012). Virf-independent regulation of shigella virB transcription is mediated by the small RNA RyhB. *PLoS ONE* 7. doi: 10.1371/journal.pone.0038592.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., et al. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of Shigella flexneri. *Molecular Microbiology* 38, 760–771. doi: 10.1046/j.1365-2958.2000.02179.x.

Calderwood, S. B., and Mekalanos, J. J. (1987). Iron regulation of Shiga-like toxin expression in Escherichia coli is mediated by the fur locus. *J Bacteriol* 169, 4759–4764. doi: 10.1128/jb.169.10.4759-4764.1987.

Campbell-Valois, F. X., Schnupf, P., Nigro, G., Sachse, M., Sansonetti, P. J., and Parsot, C. (2014). A fluorescent reporter reveals on/off regulation of the shigella type III secretion apparatus during entry and cell-to-cell spread. *Cell Host and Microbe* 15, 177–189. doi: 10.1016/j.chom.2014.01.005.

Campbell-Valois, F.-X., and Pontier, S. M. (2016). Implications of Spatiotemporal Regulation of Shigella flexneri Type Three Secretion Activity on Effector Functions: Think Globally, Act Locally. *Frontiers in Cellular and Infection Microbiology* 6, 1–13. doi: 10.3389/fcimb.2016.00028.

Campbell-Valois, F.-X., Sachse, M., Sansonetti, P. J., and Parsot, C. (2015). Escape of Actively Secreting Shigella flexneri from ATG8/LC3-Positive Vacuoles Formed during Cell-To-Cell Spread Is Facilitated by IcsB and VirA. *mBio* 6, e02567-14. doi: 10.1128/mBio.02567-14.

Campilongo, R., Di Martino, M. L., Marcocci, L., Pietrangeli, P., Leuzzi, A., Grossi, M., et al. (2014). Molecular and functional profiling of the polyamine content in enteroinvasive E. coli: Looking into the gap between commensal E. coli and harmful Shigella. *PLoS ONE* 9. doi: 10.1371/journal.pone.0106589.

Canada Public Health Agency (2020). Surveillance of shigellosis (Shigella). Available at: https://www.canada.ca/en/public-health/services/diseases/shigella/surveillance.html [Accessed April 25, 2023].

Cao, B., Wang, Y., Ding, K., Neamati, N., and Long, Y.-Q. (2012). Synthesis of the pyridinyl analogues of dibenzylideneacetone (pyr-dba) via an improved Claisen-Schmidt condensation, displaying diverse biological activities as curcumin analogues. *Org Biomol Chem* 10, 1239–1245. doi: 10.1039/c1ob06773g.

Case, H. B., Mattock, D. S., and Dickenson, N. E. (2018). Shutting Down Shigella Secretion: Characterizing Small Molecule Type Three Secretion System ATPase Inhibitors. *Biochemistry* 57, 6906–6916. doi: 10.1021/acs.biochem.8b01077.

Case, H. B., Mattock, D. S., Miller, B. R., and Dickenson, N. E. (2020). Novel Noncompetitive Type Three Secretion System ATPase Inhibitors Shut Down Shigella Effector Secretion. *Biochemistry* 59, 2667–2678. doi: 10.1021/acs.biochem.0c00431.

CDC (2023). Increase in Extensively Drug-Resistant Shigellosis in the United States. *CDC Health Alert Network*. Available at: https://emergency.cdc.gov/han/2023/pdf/CDC_HAN_486.pdf [Accessed April 24, 2023].

Cervantes-Rivera, R., Tronnet, S., and Puhar, A. (2020). Complete genome sequence and annotation of the laboratory reference strain Shigella flexneri serotype 5a M90T and genome-wide transcriptional start site determination. *BMC Genomics* 21, 285. doi: 10.1186/s12864-020-6565-5.

Charles, H., Prochazka, M., Thorley, K., Crewdson, A., Greig, D. R., Jenkins, C., et al. (2022). Outbreak of sexually transmitted, extensively drug-resistant Shigella sonnei in the UK, 2021–22: a descriptive epidemiological study. *The Lancet Infectious Diseases* 22, 1503–1510. doi: 10.1016/S1473-3099(22)00370-X.

Chen, Y.-J., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T., et al. (2013). Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Methods* 10, 659–664. doi: 10.1038/nmeth.2515.

Cosby, R. L., Judd, J., Zhang, R., Zhong, A., Garry, N., Pritham, E. J., et al. (2021). Recurrent evolution of vertebrate transcription factors by transposase capture. *Science* 371. doi: 10.1126/science.abc6405.

Daswani, P., Muthuraman, V., Macaden, R., Dias, M., and Birdi, T. (2020). Effect of Psidium guajava (guava) L. Leaf Decoction on Antibiotic-resistant Clinical Diarrhoeagenic Isolates of Shigella spp. *Int J Enteric Pathog* 8, 122–129. doi: 10.34172/ijep.2020.26.

De Jong, M. F., Liu, Z., Chen, D., and Alto, N. M. (2016). Shigella flexneri suppresses NF-κB activation by inhibiting linear ubiquitin chain ligation. *Nature Microbiology* 1. doi: 10.1038/nmicrobiol.2016.84.

Deighan, P., Beloin, C., and Dorman, C. J. (2003). Three-way interactions among the Sfh, StpA and H-NS nucleoid-structuring proteins of Shigella flexneri 2a strain 2457T. *Molecular Microbiology* 48, 1401–1416. doi: 10.1046/j.1365-2958.2003.03515.x.

Demers, B., Sansonetti, P. J., and Parsot, C. (1998). Induction of type III secretion in Shigella flexneri is associated with differential control of transcription of genes encoding secreted proteins. *The EMBO Journal* 17, 2894–2903. doi: 10.1093/emboj/17.10.2894.

Deng, W., Marshall, N. C., Rowland, J. L., McCoy, J. M., Worrall, L. J., Santos, A. S., et al. (2017). Assembly, structure, function and regulation of type III secretion systems. *Nature Reviews Microbiology* 15, 323–337. doi: 10.1038/nrmicro.2017.20.

Dey, S., Anbanandam, A., Mumford, B. E., Guzman, R. N. D., and De Guzman, R. N. (2017). Characterization of Small-Molecule Scaffolds That Bind to the Shigella Type III Secretion System Protein IpaD. *ChemMedChem* 12, 1534–1541. doi: 10.1002/cmdc.201700348.

Di Martino, M. L., Romilly, C., Wagner, E. G. H., Colonna, B., and Prosseda, G. (2016). One gene and two proteins: A leaderless mRNA supports the translation of a shorter form of the Shigella VirF regulator. *mBio* 7. doi: 10.1128/mBio.01860-16.

Dickey, S. W., Cheung, G. Y. C., and Otto, M. (2017). Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance. *Nat Rev Drug Discov* 16, 457–471. doi: 10.1038/nrd.2017.23.

Dinis, T. C. P., Madeira, V. M. C., and Almeida, L. M. (1994). Action of Phenolic Derivatives (Acetaminophen, Salicylate, and 5-aminosalicylate) as Inhibitors of Membrane Lipid Peroxidation and Peroxyl Radical Scavengers. *Archives of Biochemistry and Biophysics* 315, 161–169. doi: 10.1006/abbi.1994.1485.

Dorman, C. J. (2004). Virulence Gene Regulation in Shigella. *EcoSal Plus* 1. doi: 10.1128/ecosalplus.8.9.3.

Dorman, M. J., and Dorman, C. J. (2018). Regulatory hierarchies controlling virulence gene expression in shigella flexneriand vibrio cholera. *Frontiers in Microbiology* 9, 1–21. doi: 10.3389/fmicb.2018.02686.

Dreo, T., Pirc, M., Ramšak, Ž., Pavšič, J., Milavec, M., Žel, J., et al. (2014). Optimising droplet digital PCR analysis approaches for detection and quantification of bacteria: A case study of fire blight and potato brown rot. *Analytical and Bioanalytical Chemistry* 406, 6513–6528. doi: 10.1007/s00216-014-8084-1.

Du, J., Reeves, A. Z., Klein, J. A., Twedt, D. J., Knodler, L. A., and Lesser, C. F. (2016). The type III secretion system apparatus determines the intracellular niche of bacterial pathogens. *Proceedings of the National Academy of Sciences* 113, 4794–4799. doi: 10.1073/pnas.1520699113.

Duncan, M. C., Linington, R. G., and Auerbuch, V. (2012). Chemical Inhibitors of the Type Three Secretion System: Disarming Bacterial Pathogens. *Antimicrobial agents and chemotherapy* 56, 5433–5441.

Dunne, K. A., Chaudhuri, R. R., Rossiter, A. E., Beriotto, I., Browning, D. F., Squire, D., et al. (2017). Sequencing a piece of history: complete genome sequence of the original Escherichia coli strain. doi: 10.6084/m9.figshare.4235762.

Durand, J. M. B., Björk, G. R., Kuwae, A., Yoshikawa, M., and Sasakawa, C. (1997). The modified nucleoside 2-methylthio-N6-isopentenyladenosine in tRNA of Shigella flexneri is required for expression of virulence genes. *Journal of Bacteriology* 179, 5777–5782. doi: 10.1128/jb.179.18.5777-5782.1997.

Durand, J. M. B., Dagberg, B., Uhlin, B. E., and Björk, G. R. (2000). Transfer RNA modification, temperature and DNA superhelicity have a common target in the regulatory network of the virulence of Shigella flexneri: The expression of the virF gene. *Molecular Microbiology* 35, 924–935. doi: 10.1046/j.1365-2958.2000.01767.x.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797. doi: 10.1093/nar/gkh340.

Ernst, N. H., Reeves, A. Z., Ramseyer, J. E., Lesser, F., and Lesser, C. F. (2018). High-throughput screening of type III secretion determinants reveals a major chaperone-independent pathway. *mBio* 9, 1–13. doi: 10.1128/mBio.01050-18.

Escobar-Páramo, P., Giudicelli, C., Parsot, C., and Denamur, E. (2003). The evolutionary history of Shigella and enteroinvasive Escherichia coli revised. *Journal of Molecular Evolution* 57, 140–148. doi: 10.1007/s00239-003-2460-3.

Escobar-Páramo, P., Grenet, K., Le Menac'h, A., Rode, L., Salgado, E., Amorin, C., et al. (2004). Large-scale population structure of human commensal Escherichia coli isolates. *Applied and Environmental Microbiology* 70, 5698–5700. doi: 10.1128/AEM.70.9.5698-5700.2004.

Falconi, M., Colonna, B., Prosseda, G., Micheli, G., and Gualerzi, C. O. (1998). Thermoregulation of Shigella and Escherichia coli EIEC pathogenicity. A temperature-dependent structural transition of DNA modulates accessibility of virF promoter to transcriptional repressor H-NS. *EMBO Journal* 17, 7033–7043. doi: 10.1093/emboj/17.23.7033.

Felise, H. B., Nguyen, H. V., Pfuetzner, R. A., Barry, K. C., Jackson, S. R., Blanc, M.-P., et al. (2008). An inhibitor of gram-negative bacterial virulence protein secretion. *Cell Host Microbe* 4, 325–336. doi: 10.1016/j.chom.2008.08.001.

Fontenot, C. R., Tasnim, H., Valdes, K. A., Popescu, C. V., and Ding, H. (2020). Ferric uptake regulator (Fur) reversibly binds a [2Fe-2S] cluster to sense intracellular iron homeostasis in Escherichia coli. *J Biol Chem* 295, 15454–15463. doi: 10.1074/jbc.RA120.014814.

Garcia-Mazcorro, J. F., and Barcenas-Walls, J. R. (2016). Thinking beside the box: Should we care about the non-coding strand of the 16S rRNA gene? *FEMS Microbiology Letters* 363, 1–8. doi: 10.1093/femsle/fnw171.

Giangrossi, M., Giuliodori, A. M., Tran, C. N., Amici, A., Marchini, C., and Falconi, M. (2017). VirF relieves the transcriptional attenuation of the virulence gene icsA of Shigella flexneri affecting the icsA mRNA-RnaG complex formation. *Frontiers in Microbiology* 8, 1–12. doi: 10.3389/fmicb.2017.00650.

Giannoukos, G., Ciulla, D. M., Huang, K., Haas, B. J., Izard, J., Levin, J. Z., et al. (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol* 13, r23. doi: 10.1186/gb-2012-13-3-r23.

Gonzalez-Alba, J. M., Baquero, F., Cantón, R., and Galán, J. C. (2019). Stratified reconstruction of ancestral Escherichia coli diversification. *BMC Genomics* 20, 1–15. doi: 10.1186/s12864-019-6346-1.

Gore, A. L., and Payne, S. M. (2010). CsrA and Cra influence Shigella flexneri pathogenesis. *Infection and Immunity* 78, 4674–4682. doi: 10.1128/IAI.00589-10.

Hall, C. P., Jadeja, N. B., Sebeck, N., and Agaisse, H. (2022). Characterization of MxiE- and H-NS-Dependent Expression of ipaH7.8 , ospC1 , yccE , and yfdF in Shigella flexneri. *mSphere*. doi: 10.1128/msphere.00485-22.

Headley, V., Hong, M., Galko, M., and Payne, S. M. (1997). Expression of aerobactin genes by Shigella flexneri during extracellular and intracellular growth. *Infect Immun* 65, 818–821. doi: 10.1128/iai.65.2.818-821.1997.

Hotinger, J. A., Pendergrass, H. A., and May, A. E. (2021). Molecular Targets and Strategies for Inhibition of the Bacterial Type III Secretion System (T3SS); Inhibitors Directly Binding to T3SS Components. *Biomolecules* 11, 316. doi: 10.3390/biom11020316.

Hu, B., Lara-Tejero, M., Kong, Q., Galán, J. E., and Liu, J. (2017). In Situ Molecular Architecture of the Salmonella Type III Secretion Machine. *Cell* 168, 1065-1074.e10. doi: 10.1016/j.cell.2017.02.022.

Hu, B., Morado, D. R., Margolin, W., Rohde, J. R., Arizmendi, O., Picking, W. L., et al. (2015). Visualization of the type III secretion sorting platform of *Shigella flexneri*. *Proceedings of the National Academy of Sciences* 112, 1047–1052. doi: 10.1073/pnas.1411610112.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., et al. (2002). Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157. *Nucleic acids research* 30, 4432–4441. doi: 10.1093/nar/gkf566.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2.

Kane, C. D., Schuch, R., Day, W. A. J., and Maurelli, A. T. (2002). MxiE regulates intracellular expression of factors secreted by the Shigella flexneri 2a type III secretion system. *Journal of Bacteriology* 184, 4409–4419. doi: 10.1128/JB.184.16.4409–4419.2002.

Kane, K. A., and Dorman, C. J. (2012). VirB-mediated positive feedback control of the virulence gene regulatory cascade of Shigella flexneri. *J Bacteriol* 194, 5264–5273. doi: 10.1128/JB.00800-12.

Kang, E., Crouse, A., Chevallier, L., Pontier, S. M., Alzahrani, A., Silué, N., et al. (2018). Enterobacteria and host resistance to infection. *Mammalian Genome*. doi: 10.1007/s00335-018-9749-4.

Karney, M. M. A., McKenna, J. A., Weatherspoon-Griffin, N., Karabachev, A. D., Millar, M. E., Potocek, E. A., et al. (2019). Investigating the DNA-binding site for VirB, a key transcriptional regulator of Shigella virulence genes, using an in vivo binding tool. *Genes* 10, 1–12. doi: 10.3390/genes10020149.

Katsowich, N., Elbaz, N., Pal, R. R., Mills, E., Kobi, S., Kahan, T., et al. (2017). Host cell attachment elicits posttranscriptional regulation in infecting enteropathogenic bacteria. *Science* 355, 735–739. doi: 10.1126/science.aah4886.

Khalil, I. A., Troeger, C., Blacker, B. F., Rao, P. C., Brown, A., Atherly, D. E., et al. (2018). Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *The Lancet Infectious Diseases* 18, 1229–1240. doi: 10.1016/S1473-3099(18)30475-4.

Koppolu, V., Osaka, I., Skredenske, J. M., Kettle, B., Hefty, P. S., Li, J., et al. (2013). Small-molecule inhibitor of the Shigella flexneri master virulence regulator VirF. *Infect Immun* 81, 4220–4231. doi: 10.1128/IAI.00919-13.

Kotlajich, M. V., Hron, D. R., Boudreau, B. A., Sun, Z., Lyubchenko, Y. L., and Landick, R. (2015). Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *eLife* 2015, 1–24. doi: 10.7554/eLife.04970.

Kotloff, K. L., Platts-Mills, J. A., Nasrin, D., Roose, A., Blackwelder, W. C., and Levine, M. M. (2017). Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine* 35, 6783–6789. doi: 10.1016/j.vaccine.2017.07.036.

Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P., and Zaidi, A. K. M. (2018). Shigellosis. *The Lancet* 391, 801–812. doi: 10.1016/S0140-6736(17)33296-8.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., et al. (1999). Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bulletin of the World Health Organization*.

Krzywinski, M., Schein, J., Birol, I., Jones, S., and Marra, M. (2008). CIRCOS - an information aesthetic for comparative genomics.pdf.

Kurihara, S., Suzuki, H., Oshida, M., and Benno, Y. (2011). A novel putrescine importer required for type 1 pili-driven surface motility induced by extracellular putrescine in

Escherichia coli K-12. *Journal of Biological Chemistry* 286, 10185–10192. doi: 10.1074/jbc.M110.176032.

Lalaouna, D., Prévost, K., Park, S., Chénard, T., Bouchard, M. P., Caron, M. P., et al. (2021). Binding of the rna chaperone hfq on target mrnas promotes the small rna ryhb-induced degradation in escherichia coli. *Non-coding RNA* 7. doi: 10.3390/ncrna7040064.

Lamberte, L. E., Baniulyte, G., Singh, S. S., Stringer, A. M., Bonocora, R. P., Stracy, M., et al. (2017). Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. *Nat Microbiol* 2, 1–9. doi: 10.1038/nmicrobiol.2016.249.

Lampel, K. A., Formal, S. B., and Maurelli, A. T. (2018). A Brief History of Shigella. *EcoSal Plus* 8. doi: 10.1128/ecosalplus.ESP-0006-2017.

Lan, R., Lumb, B., Ryan, D., and Reeves, P. R. (2001). Molecular evolution of large virulence plasmid in Shigella clones and enteroinvasive Escherichia coli. *Infection and Immunity* 69, 6303–6309. doi: 10.1128/IAI.69.10.6303-6309.2001.

Lan, R., and Reeves, P. R. (2002). Escherichia coli in disguise: molecular origins of Shigella. *Microbes and Infection* 4, 1125–1132. doi: 10.1016/S1286-4579(02)01637-4.

Lawrence, J. G., and Ochman, H. (1997). Amelioration of Bacterial Genomes: Rates of Change and Exchange. *Journal of Molecular Evolution* 44, 383–397. doi: 10.1007/pl00006158.

Layton, A. N., Hudson, D. L., Thompson, A., Hinton, J. C. D., Stevens, J. M., Galyov, E. E., et al. (2010). Salicylidene acylhydrazide-mediated inhibition of type III secretion system-1 in Salmonella enterica serovar Typhimurium is associated with iron restriction and can be reversed by free iron. *FEMS Microbiology Letters* 302, 114–122. doi: 10.1111/j.1574-6968.2009.01847.x.

Le Gall, T., Clermont, O., Gouriou, S., Picard, B., Nassif, X., Denamur, E., et al. (2007). Extraintestinal Virulence Is a Coincidental By-Product of Commensalism in B2 Phylogenetic Group Escherichia coli Strains. *Molecular Biology and Evolution* 24, 2373–2384. doi: 10.1093/molbev/msm172.

Le Gall, T., Mavris, M., Martino, M. C., Bernardini, M. L., Denamur, E., and Parsot, C. (2005). Analysis of virulence plasmid gene expression defines three classes of effectors in the type III secretion system of Shigella flexneri. *Microbiology* 151, 951–962. doi: 10.1099/mic.0.27639-0.

Lefèvre, S., Njamkepo, E., Feldman, S., Ruckly, C., Carle, I., Lejay-Collin, M., et al. (2023). Rapid emergence of extensively drug-resistant Shigella sonnei in France. *Nat Commun* 14, 462. doi: 10.1038/s41467-023-36222-8.

Lill, R. (2020). Do FeS clusters rule bacterial iron regulation? *J Biol Chem* 295, 15464–15465. doi: 10.1074/jbc.H120.016190.

Liu, S., Feng, J., Pu, J., Xu, X., Lu, S., Yang, J., et al. (2019). Genomic and molecular characterisation of Escherichia marmotae from wild rodents in Qinghai-Tibet plateau as a potential pathogen. *Scientific Reports* 9, 1–9. doi: 10.1038/s41598-019-46831-3.

Liu, Y., Chen, H., and Kenney, L. J. (2010). A divalent switch drives conformations between stiffening and bridging modes. *GENES & DEVELOPMENT*, 339–344. doi: 10.1101/gad.1883510.H-NS.

Lloyd, R. V., Hanna, P. M., and Mason, R. P. (1997). The Origin of the Hydroxyl Radical Oxygen in the Fenton Reaction. *Free Radical Biology and Medicine* 22, 885–888. doi: 10.1016/S0891-5849(96)00432-7.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S., and Shafee, T. (2017). Transcriptomics technologies. *PLoS Computational Biology* 13, 1–23. doi: 10.1371/journal.pcbi.1005457.

Lucchini, S., Liu, H., Jin, Q., Hinton, J. C. D., and Yu, J. (2005). Transcriptional Adaptation of Shigella flexneri during Infection of Macrophages and Epithelial Cells: Insights into the Strategies of a Cytosolic Bacterial Pathogen. *Infection and Immunity* 73, 88–102. doi: 10.1128/IAI.73.1.88-102.2005.

Marman, H. E., Mey, A. R., and Payne, S. M. (2014). Elongation factor P and modifying enzyme PoxA are necessary for virulence of Shigella flexneri. *Infection and Immunity* 82, 3612–3621. doi: 10.1128/IAI.01532-13.

Marteyn, B., West, N. P., Browning, D. F., Cole, J. A., Shaw, J. G., Palm, F., et al. (2010). Modulation of Shigella virulence in response to available oxygen in vivo. *Nature* 465, 355–358. doi: 10.1038/nature08970.

Martino, M. L. D., Falconi, M., Micheli, G., Colonna, B., Prosseda, G., Di Martino, M. L., et al. (2016). The multifaceted activity of the VirF regulatory protein in the Shigella Lifestyle. *Frontiers in Molecular Biosciences* 3, 1–11. doi: 10.3389/fmolb.2016.00061.

Massé, E., Escorcia, F. E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in Escherichia coli. *Genes and Development* 17, 2374–2383. doi: 10.1101/gad.1127103.

Massé, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* 99, 4620–4625. doi: 10.1073/pnas.032066599.

Maurelli, A. T., Fernández, R. E., Bloch, C. A., Rode, C. K., and Fasano, A. (1998). "Black holes" and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3943–3948. doi: 10.1073/pnas.95.7.3943.

Mavris, M., Page, A. L., Tournebize, R., Demers, B., Sansonetti, P., and Parsot, C. (2002a). Regulation of transcription by the activity of the Shigella flexneri type III secretion

apparatus. *Molecular Microbiology* 43, 1543–1553. doi: 10.1046/j.1365-2958.2002.02836.x.

Mavris, M., Sansonetti, P. J., and Parsot, C. (2002b). Identification of the cis-acting site involved in activation of promoters regulated by activity of the type III secretion apparatus in shigella flexneri. *Journal of Bacteriology* 184, 6751–6759. doi: 10.1128/JB.184.24.6751-6759.2002.

McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., et al. (2013). Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Research* 41, 1–16. doi: 10.1093/nar/gkt444.

McKenna, J. A., Karney, M. M. A., Chan, D. K., Weatherspoon-Griffin, N., Larios, B. B., Pilonieta, M. C., et al. (2022). The AraC/XylS Protein MxiE and Its Coregulator IpgC Control a Negative Feedback Loop in the Transcriptional Cascade That Regulates Type III Secretion in Shigella flexneri. *Journal of Bacteriology* 204. doi: 10.1128/jb.00137-22.

McKenna, J. A., and Wing, H. J. (2020). The anti-activator of type III secretion, OspD1, is transcriptionally regulated by VirB and H-NS from remote sequences in Shigella flexneri . *Journal of Bacteriology*. doi: 10.1128/jb.00072-20.

McShan, A. C., and De Guzman, R. N. (2015). The Bacterial Type III Secretion System as a Target for Developing New Antibiotics. *Chemical Biology and Drug Design* 85, 30–42. doi: 10.1111/cbdd.12422.

McVicker, G., and Tang, C. M. (2016). Deletion of toxin–antitoxin systems in the evolution of Shigella sonnei as a host-adapted pathogen. *Nature Microbiology* 2, 16204. doi: 10.1038/nmicrobiol.2016.204.

Ménard, R., Sansonetti, P., and Parsot, C. (1994). The secretion of the Shigella flexneri Ipa invasins is activated by epithelial cells and controlled by IpaB and IpaD. *The EMBO journal* 13, 5293–302. doi: 10.1002/j.1460-2075.1994.tb06863.x.

Mettert, E. L., and Kiley, P. J. (2018). Reassessing the Structure and Function Relationship of the O2 Sensing Transcription Factor FNR. *Antioxidants and Redox Signaling* 29, 1830–1840. doi: 10.1089/ars.2017.7365.

Mey, A. R., Gómez-Garzón, C., and Payne, S. M. (2021). Iron Transport and Metabolism in Escherichia, Shigella, and Salmonella. *EcoSal Plus* 9, eESP-0034-2020. doi: 10.1128/ecosalplus.ESP-0034-2020.

Murphy, E. R., and Payne, S. M. (2007). RyhB, an iron-responsive small RNA molecule, regulates Shigella dysenteriae virulence. *Infection and Immunity* 75, 3470–3477. doi: 10.1128/IAI.00112-07.

Muschiol, S., Bailey, L., Gylfe, Å., Sundin, C., Hultenby, K., Bergström, S., et al. (2006). A small-molecule inhibitor of type III secretion inhibits different stages of the infectious

cycle of *Chlamydia trachomatis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14566–14571. doi: 10.1073/pnas.0606412103.

Nakayama, S., and Watanabe, H. (1995). Involvement of cpxA, a sensor of a two-component regulatory system, in the pH-dependent regulation of expression of Shigella sonnei virF gene. *J Bacteriol* 177, 5062–5069. doi: 10.1128/jb.177.17.5062-5069.1995.

Nasibullin, I., Smirnov, I., Ahmadi, P., Vong, K., Kurbangalieva, A., and Tanaka, K. (2022). Synthetic prodrug design enables biocatalytic activation in mice to elicit tumor growth suppression. *Nat Commun* 13, 39. doi: 10.1038/s41467-021-27804-5.

NCCID (2022). Shigellosis. *National Collaborating Centre for Infectious Diseases*. Available at: https://nccid.ca/debrief/shigellosis/ [Accessed April 30, 2023].

Ni, Z., Jiang, L., Feng, L., Wang, L., and Liu, B. (2015). Transcriptional adaptation of Shigella flexneri during adherence to epithelial cells. *Journal of Basic Microbiology* 55, 186–194. doi: 10.1002/jobm.201400414.

Nigro, G., Arena, E. T., Sachse, M., Moya-Nilges, M., Marteyn, B. S., Sansonetti, P. J., et al. (2019). Mapping of Shigella flexneri's tissue distribution and type III secretion apparatus activity during infection of the large intestine of guinea pigs. *Pathogens and Disease* 77, 1–9. doi: 10.1093/femspd/ftz054.

Onodera, N. T., Ryu, J., Durbic, T., Nislow, C., Archibald, J. M., and Rohde, J. R. (2012). Genome sequence of shigella flexneri serotype 5a strain M90T Sm. *Journal of Bacteriology* 194, 3022. doi: 10.1128/JB.00393-12.

Panagiotidis, C. A., Iwang, S.-C., and Canellakis, E. S. (1995). Relationship of the expression of the S20 and L34 ribosomal proteins to polyamine biosynthesis in Escherichia coli. *Int J Biochem Cell Biol* . 27, 157–68. doi: 10.1016/1357-2725(94)00068-m.

Parsot, C. (2009). Shigella type III secretion effectors: how, where, when, for what purposes? *Current Opinion in Microbiology* 12, 110–116. doi: 10.1016/j.mib.2008.12.002.

Parsot, C., Ageron, E., Penno, C., Mavris, M., Jamoussi, K., D'Hauteville, H., et al. (2005). A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in Shigella flexneri. *Molecular Microbiology* 56, 1627–1635. doi: 10.1111/j.1365-2958.2005.04645.x.

Pasqua, M., Michelacci, V., Di Martino, M. L., Tozzoli, R., Grossi, M., Colonna, B., et al. (2017). The Intriguing Evolutionary Journey of Enteroinvasive E. coli (EIEC) toward Pathogenicity. *Front. Microbiol.* 8, 2390. doi: 10.3389/fmicb.2017.02390.

Payne, S. M., Wyckoff, E. E., Murphy, E. R., Oglesby, A. G., Boulette, M. L., and Davies, N. M. L. (2006). Iron and Pathogenesis of Shigella: Iron Acquisition in the Intracellular Environment. *Biometals* 19, 173–180. doi: 10.1007/s10534-005-4577-x.

Penno, C., and Parsot, C. (2006). Transcriptional slippage in mxiE controls transcription and translation of the downstream mxiD gene, which encodes a component of the Shigella flexneri type III secretion apparatus. *Journal of Bacteriology* 188, 1196–1198. doi: 10.1128/JB.188.3.1196-1198.2006.

Penno, C., Sansonetti, P., and Parsot, C. (2005). Frameshifting by transcriptional slippage is involved in production of MxiE, the transcription activator regulated by the activity of the type III secretion apparatus in Shigella flexneri. *Molecular Microbiology* 56, 204–214. doi: 10.1111/j.1365-2958.2005.04530.x.

Phalipon, A., Arondel, J., Nato, F., Rouyre, S., Mazie, J. C., and Sansonetti, P. J. (1992). Identification and characterization of B-cell epitopes of IpaC, an invasion-associated protein of Shigella flexneri. *Infection and Immunity* 60, 1919–1926. doi: 10.1128/iai.60.5.1919-1926.1992.

Picker, M. A., and Wing, H. J. (2016). H-NS, its family members and their regulation of virulence genes in Shigella species. *Genes* 7. doi: 10.3390/genes7120112.

Pilla, G., Arcari, G., Tang, C. M., and Carattoli, A. (2022). Virulence plasmid pINV as a genetic signature for Shigella flexneri phylogeny. *Microbial Genomics* 8. doi: 10.1099/mgen.0.000846.

Pilonieta, M. C., and Munson, G. P. (2008). The chaperone IpgC copurifies with the virulence regulator MxiE. *Journal of Bacteriology* 190, 2249–2251. doi: 10.1128/JB.01824-07.

Pinaud, L., Ferrari, M. L., Friedman, R., Jehmlich, N., von Bergen, M., Phalipon, A., et al. (2017a). Identification of novel substrates of Shigella T3SA through analysis of its virulence plasmid-encoded secretome. *PLoS ONE* 12, e0186920.

Pinaud, L., Samassa, F., Porat, Z., Ferrari, M. L., Belotserkovsky, I., Parsot, C., et al. (2017b). Injection of T3SS effectors not resulting in invasion is the main targeting mechanism of Shigella toward human lymphocytes. *Proc Natl Acad Sci U S A* 13, 201707098. doi: 10.1073/pnas.1707098114.

Pinheiro, L. B., Coleman, V. A., Hindson, C. M., Herrmann, J., Hindson, B. J., Bhat, S., et al. (2012). Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Analytical Chemistry* 84, 1003–1011. doi: 10.1021/ac202578x.

Porcheron, G., and Dozois, C. M. (2015). Interplay between iron homeostasis and virulence: Fur and RyhB as major regulators of bacterial pathogenicity. *Veterinary Microbiology* 179, 2–14. doi: 10.1016/j.vetmic.2015.03.024.

Porcheron, G., Habib, R., Houle, S., Caza, M., Lépine, F., Daigle, F., et al. (2014). The small RNA RyhB contributes to siderophore production and virulence of uropathogenic Escherichia coli. *Infection and Immunity* 82, 5056–5068. doi: 10.1128/IAI.02287-14.

Poulsen, L. D., and Vinther, J. (2018). RNA-Seq for Bacterial Gene Expression. *Current Protocols in Nucleic Acid Chemistry* 73, 1–12. doi: 10.1002/cpnc.55.

Prosseda, G., Di Martino, M. L., Campilongo, R., Fioravanti, R., Micheli, G., Casalino, M., et al. (2012). Shedding of genes that interfere with the pathogenic lifestyle: the Shigella model. *Research in Microbiology* 163, 399–406. doi: 10.1016/j.resmic.2012.07.004.

Puccio, S., Grillo, G., Licciulli, F., Severgnini, M., Liuni, S., Bicciato, S., et al. (2017). WoPPER: Web server for Position Related data analysis of gene Expression in Prokaryotes. *Nucleic Acids Research* 45, W109–W115. doi: 10.1093/nar/gkx329.

Pupo, G. M., Lan, R., and Reeves, P. R. (2000). Multiple independent origins of Shigella clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10567–10572. doi: 10.1073/pnas.180094797.

Rahman, K. M., Arifeen, S. E., Zaman, K., Rahman, M., Raqib, R., Yunus, M., et al. (2011). Safety, dose, immunogenicity, and transmissibility of an oral live attenuated Shigella flexneri 2a vaccine candidate (SC602) among healthy adults and school children in Matlab, Bangladesh. *Vaccine* 29, 1347–1354. doi: 10.1016/j.vaccine.2010.10.035.

Rangwala, S. H., Kuznetsov, A., Ananiev, V., Asztalos, A., Borodin, E., Evgeniev, V., et al. (2021). Accessing NCBI data using the NCBI sequence viewer and genome data viewer (GDV). *Genome Research* 31, 159–169. doi: 10.1101/gr.266932.120.

Ranjbar, R., and Farahani, A. (2019). Shigella: Antibiotic-Resistance Mechanisms And New Horizons For Treatment. *IDR* Volume 12, 3137–3167. doi: 10.2147/IDR.S219755.

Raso, M. M., Arato, V., Gasperini, G., and Micoli, F. (2023). Toward a Shigella Vaccine: Opportunities and Challenges to Fight an Antimicrobial-Resistant Pathogen. *IJMS* 24, 4649. doi: 10.3390/ijms24054649.

Ratledge, C., and Dover, L. G. (2000). Iron Metabolism in Pathogenic Bacteria. *Annu. Rev. Microbiol.* 54, 881–941. doi: 10.1146/annurev.micro.54.1.881.

Raymond, K. N., Dertz, E. A., and Kim, S. S. (2003). Enterobactin: An archetype for microbial iron transport. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3584–3588. doi: 10.1073/pnas.0630018100.

Ren, C. P., Chaudhuri, R. R., Fivian, A., Bailey, C. M., Antonio, M., Barnes, W. M., et al. (2004). The ETT2 gene cluster, encoding a second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition. *Journal of Bacteriology* 186, 3547–3560. doi: 10.1128/JB.186.11.3547-3560.2004.

Retchless, A. C., and Lawrence, J. G. (2010). Phylogenetic incongruence arising from fragmented speciation in enteric bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 107, 11453–11458. doi: 10.1073/pnas.1001291107.

Riccardi, E., Mastbergen, E. C. V., Navarre, W. W., and Vreede, J. (2019). Predicting the mechanism and rate of H-NS binding to at-rich DNA. *PLoS Computational Biology* 15, 1–21. doi: 10.1371/journal.pcbi.1006845.

Rice, P., Longden, L., and Bleasby, A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2.

Richardson, D., Devlin, J., Fitzpatrick, C., and Pinto-Sander, N. (2021). Sexually transmitted *Shigella flexneri* and *Shigella sonnei* in men who have sex with men. *Sex Transm Infect* 97, 244–244. doi: 10.1136/sextrans-2020-054589.

Sansonetti, P. J. (2006). Shigellosis: An old disease in new clothes? *PLoS Medicine* 3, 1465–1466. doi: 10.1371/journal.pmed.0030354.

Sansonetti, P. J., and Arondel, J. (1989). Construction and evaluation of a double mutant of Shigella flexneri as a candidate for oral vaccination against shigellosis. *Vaccine* 7, 443–450. doi: 10.1016/0264-410X(89)90160-6.

Sansonetti, P. J., and Mounier, J. (1987). Metabolic events mediating early killing of host cells infected by Shigella flexneri. *Microbial Pathogenesis* 3, 53–61. doi: 10.1016/0882-4010(87)90037-4.

Sansonnetti, P. J., Kopecko, D. J., and Formal, S. B. (1982). Involvement of a plasmid in the invasive ability of Shigella flexneri. *Infection and Immunity* 35, 852–860. doi: doi: 10.1128/iai.35.3.852-860.1982.

Santapaola, D., Casalino, M., Petrucca, A., Presutti, C., Zagaglia, C., Berlutti, F., et al. (2002). Enteroinvasive Escherichia coli virulence-plasmid-carried apyrase (apy) and ospB genes are organized as a bicistronic operon and are subject to differential expression. *Microbiology* 148, 2519–2529. doi: 10.1016/S0026-0576(08)80167-5.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676–682. doi: 10.1038/nmeth.2019.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* 9, 671–675. doi: 10.1038/nmeth.2089.

Seo, S. W., Kim, D., Latif, H., O'Brien, E. J., Szubin, R., and Palsson, B. O. (2014). Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. *Nat Commun* 5, 4910. doi: 10.1038/ncomms5910.

Shishkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature Methods* 12, 323–325. doi: 10.1038/nmeth.3313.

Siguier, P., Gourbeyre, E., Varani, A., Ton-Hoang, B., and Chandler, M. (2015). Everyman's guide to bacterial insertion sequences. *Mobile DNA III*, 555–590. doi: 10.1128/9781555819217.ch26.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic acids research* 34, 32–36. doi: 10.1093/nar/gkj014.

Silué, N., and Campbell-Valois, F.-X. (2022). icaR and icaT are Ancient Chromosome Genes Encoding Substrates of the Type III Secretion Apparatus in Shigella flexneri. *mSphere* 7. doi: 10.1128/msphere.00115-22.

Silué, N., Marcantonio, E., and Campbell-Valois, F. X. (2020). RNA-Seq analysis of the T3SA regulon in Shigella flexneri reveals two new chromosomal genes upregulated in the on-state. *Methods* 176, 71–81. doi: 10.1016/j.ymeth.2019.03.017.

Sims, G. E., and Kim, S. H. (2011). Whole-genome phylogeny of Escherichia coli/Shigella group by feature frequency profiles (FFPs). *Proceedings of the National Academy of Sciences of the United States of America* 108, 8329–8334. doi: 10.1073/pnas.1105168108.

Skredenske, J. M., Koppolu, V., Kolin, A., Deng, J., Kettle, B., Taylor, B., et al. (2013). Identification of a small-molecule inhibitor of bacterial AraC family activators. *J Biomol Screen* 18, 588–598. doi: 10.1177/1087057112474690.

Slater, S., Sågfors, A., Pollard, D., Ruano-Gallego, D., and Frankel, G. (2018). The Type III Secretion System of Pathogenic Escherichia coli. *Curr Top Microbiol Immunol.* 416, 51–72. doi: 10.1007/82_2018_116.

Solaiyappan Mani, S., and Reinhold-Hurek, B. (2021). RNA-Seq Provides New Insights into the Gene Expression Changes in Azoarcus olearius BH72 under Nitrogen-Deficient and Replete Conditions beyond the Nitrogen Fixation Process. *Microorganisms* 9, 1888. doi: 10.3390/microorganisms9091888.

Song, J., Malwal, S. R., Baig, N., Schurig-Briccio, L. A., Gao, Z., Vaidya, G. S., et al. (2020). Discovery of Prenyltransferase Inhibitors with in Vitro and in Vivo Antibacterial Activity. *ACS Infectious Diseases* 6, 2979–2993. doi: 10.1021/acsinfecdis.0c00472.

Standish, A. J., Teh, M. Y., Tran, E. N. H., Doyle, M. T., Baker, P. J., and Morona, R. (2016). Unprecedented Abundance of Protein Tyrosine Phosphorylation Modulates Shigella flexneri Virulence. *Journal of Molecular Biology* 428, 4197–4208. doi: 10.1016/j.jmb.2016.06.016.

Stella, V. J. (2004). Prodrugs as therapeutics. *Expert Opinion on Therapeutic Patents* 14, 277–280. doi: 10.1517/13543776.14.3.277.

Subtil, A., Parsot, C., and Dautry-Varsat, A. (2001). Secretion of predicted Inc proteins of Chlamydia pneumoniae by a heterologous type III machinery. *Molecular Microbiology* 39, 792–800. doi: 10.1046/j.1365-2958.2001.02272.x.

Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., et al. (2017). The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research* 45, D362–D368. doi: 10.1093/nar/gkw937.

Taniya, T., Mitobe, J., Nakayama, S. I., Mingshan, Q., Okuda, K., and Watanabe, H. (2003). Determination of the InvE binding site required for expression of IpaB of the Shigella sonnei virulence plasmid: Involvement of a ParB BoxA-like sequence. *Journal of Bacteriology* 185, 5158–5165. doi: 10.1128/JB.185.17.5158-5165.2003.

The, H. C., Thanh, D. P., Holt, K. E., Thomson, N. R., and Baker, S. (2016). The genomic signatures of Shigella evolution, adaptation and geographical spread. *Nature Reviews Microbiology* 14, 235–250. doi: 10.1038/nrmicro.2016.10.

Tjaden, B. (2015). De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biology* 16, 1–10. doi: 10.1186/s13059-014-0572-2.

Tobe, T., Yoshikawa, M., Mizuno, T., and Sasakawa, C. (1993). Transcriptional control of the invasion regulatory gene virB of Shigella flexneri: Activation by VirF and repression by H-NS. *Journal of Bacteriology* 175, 6142–6149. doi: 10.1128/jb.175.19.6142-6149.1993.

Touchon, M., Perrin, A., De Sousa, J. A. M., Vangchhia, B., Burn, S., O'Brien, C. L., et al. (2020). *Phylogenetic background and habitat drive the genetic diversification of Escherichia coli*. doi: 10.1371/journal.pgen.1008866.

Tran, C. N., Giangrossi, M., Prosseda, G., Brandi, A., Di Martino, M. L., Colonna, B., et al. (2011). A multifactor regulatory circuit involving H-NS, VirF and an antisense RNA modulates transcription of the virulence gene icsA of Shigella flexneri. *Nucleic Acids Research* 39, 8122–8134. doi: 10.1093/nar/gkr521.

Trivett, H. (2022). Increase in extensively drug resistant Shigella sonnei in Europe. *The Lancet Microbe* 3, e481. doi: 10.1016/S2666-5247(22)00160-4.

Tsou, L. K., Lara-Tejero, M., Rosefigura, J., Zhang, Z. J., Wang, Y. C., Yount, J. S., et al. (2016). Antibacterial Flavonoids from Medicinal Plants Covalently Inactivate Type III Protein Secretion Substrates. *Journal of the American Chemical Society* 138, 2209–2218. doi: 10.1021/jacs.5b11575.

Turner, E. C., and Dorman, C. J. (2007). H-NS antagonism in Shigella flexneri by VirB, a virulence gene transcription regulator that is closely related to plasmid partition factors. *Journal of Bacteriology* 189, 3403–3413. doi: 10.1128/JB.01813-06.

Vakulskas, C. A., Potts, A. H., Babitzke, P., Ahmer, B. M. M., and Romeo, T. (2015). Regulation of Bacterial Virulence by Csr (Rsm) Systems. *Microbiology and Molecular Biology Reviews* 79, 193–224. doi: 10.1128/mmbr.00052-14.

Venkatesan, M. M., Goldberg, M. B., Rose, D. J., Grotbeck, E. J., Burland, V., and Blattner, F. R. (2001). Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri. *Infection and Immunity*. doi: 10.1128/IAI.69.5.3271-3285.2001.

Vergara-Irigaray, M., Fookes, M. C., Thomson, N. R., and Tang, C. M. (2014). RNA-seq analysis of the influence of anaerobiosis and FNR on Shigella flexneri. *BMC Genomics* 15, 1–22. doi: 10.1186/1471-2164-15-438.

Vilà, N., and Walcarius, A. (2020). Bis(terpyridine) Iron(II) Functionalized Vertically-Oriented Nanostructured Silica Films: Toward Electrochromic Materials. *Frontiers in Chemistry* 8. doi: 10.3389/fchem.2020.00830.

Viollier, E., Inglett, P. W., Hunter, K., Roychoudhury, A. N., and Van Cappellen, P. (2000). The ferrozine method revisited: Fe(II)/Fe(III) determination in natural waters. *Applied Geochemistry* 15, 785–790. doi: 10.1016/S0883-2927(99)00097-9.

von Seidlein, L., Kim, D. R., Ali, M., Lee, H., Wang, X., Thiem, V. D., et al. (2006). A Multicentre Study of Shigella Diarrhoea in Six Asian Countries: Disease Burden, Clinical Manifestations, and Microbiology. *PLoS Med* 3, e353. doi: 10.1371/journal.pmed.0030353.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033.

Watkins, L. K. F., and Appiah, G. D. (n.d.). Shigellosis - Chapter 4 - 2020 Yellow Book | Travelers' Health | CDC. Available at: https://wwwnc.cdc.gov/travel/yellowbook/2020/travel-related-infectious-diseases/shigellosis [Accessed August 12, 2022].

Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research* 42, 581–591. doi: 10.1093/nar/gkt1099.

Weatherspoon-Griffin, N., Picker, M. A., Pew, K. L., Park, H. S., Ginete, D. R., Karney, M. M. A., et al. (2018). Insights into transcriptional silencing and anti-silencing in Shigella flexneri: a detailed molecular analysis of the icsP virulence locus. *Molecular Microbiology* 108, 505–518. doi: 10.1111/mmi.13932.

Wei, Y., and Murphy, E. R. (2016). Shigella Iron Acquisition Systems and their Regulation. *Front Cell Infect Microbiol* 6, 18. doi: 10.3389/fcimb.2016.00018.

WHO (2017). WHO publishes list of bacteria for which new antibiotics are urgently needed. Available at: https://www.who.int/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed [Accessed May 13, 2023].

WHO (2022). Shigella: Immunization, Vaccines and Biologicals. Available at: https://www.who.int/teams/immunization-vaccines-and-biologicals/diseases/shigella [Accessed May 2, 2023].

Wing, H. J., Yan, A. W., Goldman, S. R., and Goldberg, M. B. (2004). Regulation of IcsP, the Outer Membrane Protease of the Shigella Actin Tail Assembly Protein IcsA, by Virulence Plasmid Regulators VirF and VirB. *Journal of Bacteriology* 186, 699–705. doi: 10.1128/JB.186.3.699-705.2004.

Xia, X. (2017). ARSDA: A New Approach for Storing, Transmitting and Analyzing Transcriptomic Data. *G3: Genes|Genomes|Genetics* 7, 3839–3848. doi: 10.1534/g3.117.300271.

Xiong, G., Wu, Z., Yi, J., Fu, L., Yang, Z., Hsieh, C., et al. (2021). ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research* 49, W5–W14. doi: 10.1093/nar/gkab255.

Yang, Y., Harris, D. P., Luo, F., Xiong, W., Joachimiak, M., Wu, L., et al. (2009). Snapshot of iron response in Shewanella oneidensis by gene network reconstruction. *BMC Genomics* 10, 131. doi: 10.1186/1471-2164-10-131.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC bioinformatics* 13, 134. doi: 10.1186/1471-2105-13-134.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., and Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE* 9, e78644. doi: 10.1371/journal.pone.0078644.

Zhou, M., Guo, Z., Duan, Q., Hardwidge, P. R., and Zhu, G. (2014). Escherichia coli type III secretion system 2: A new kind of T3SS? *Veterinary Research* 45, 1–5. doi: 10.1186/1297-9716-45-32.

Appendices

Appendix 1: Supplementary Figures and Tables for chapter 2

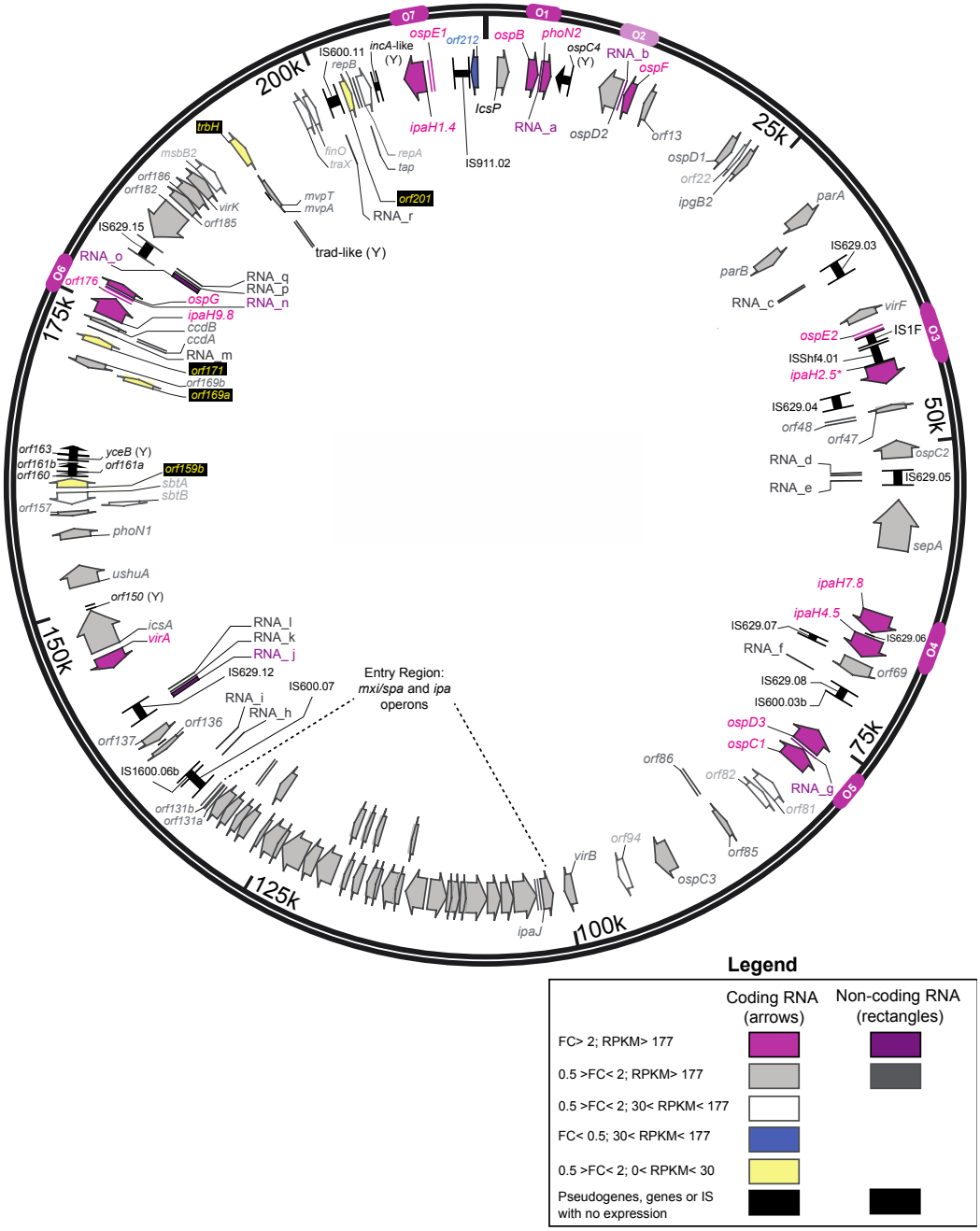


Figure 2.S. 1. Mapping of the RNA-Seq data on the virulence plasmid pWR100. The RNA-Seq data is embedded in the map using the color code indicated in the legend. The coding transcripts in magenta and light gray, as well as non-coding transcripts in plum and charcoal correspond to those reported in the wheel chart in Figure 2.2. The other ones are lower expressed gene, whose raw data are also available in Table

2.S.2. For the sake of clarity, only IS most relevant to this study are shown on the map. Empty spaces between defined genetic elements are corresponding mostly to the remaining IS. Operons discussed in Figure 3 (O1-O7) are also indicated on the outside circle as magenta or lavender rounded rectangles. It is noteworthy that *ipaH2.5* (marked by a *) has many characteristics of a pseudogene. The genetic elements marked by the symbol Ψ were annotated as pseudogenes in the Genbank file and were not detected in the analysis. Expression of transcripts corresponding to operons composed of *orf160-163* was not detected (black arrows). On this map, we have used the original annotations of pWR100 **(Buchrieser et al., 2000; Venkatesan et al., 2001)**. Note that since the release of the virulence plasmid sequence, some hypothetical protein-coding genes (*orfs*) have been given a systematic name such as *gmvT* (*orf47*), *gmvA* (*orf48*), *ospZ* (*orf 212*), and *ospI* (*orf169b*) (Table A.2).

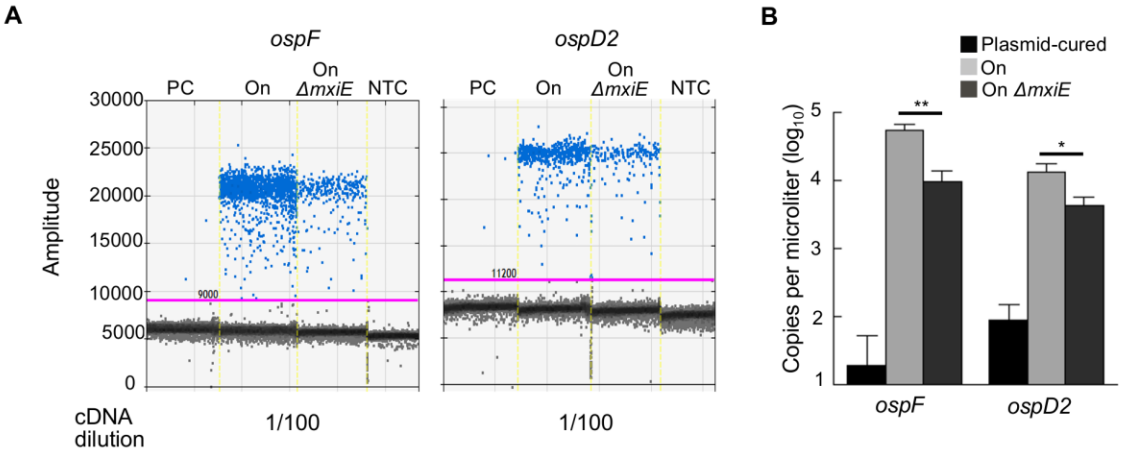
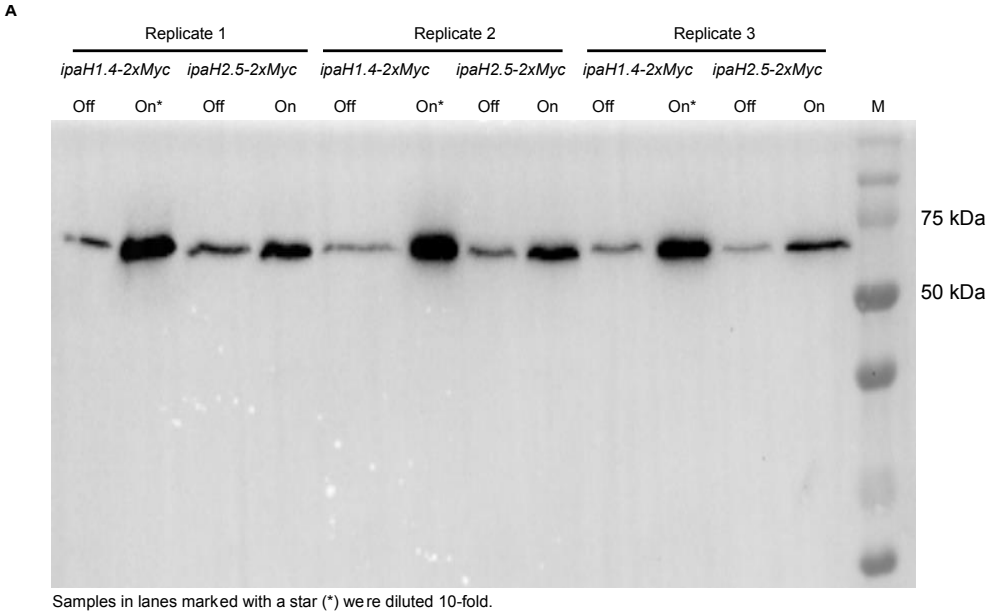


Figure 2.S. 2. *ospD2* gene is a member of the T3SA regulon. (**A**) Representative droplet amplitude graph obtained for the *ospF* and *ospD2* genes tested in the virulence plasmid-cured (BS176), on-state (*ipaB4*) and on-state Δ*mxiE* (*ipaB4 mxiE*) as well as the corresponding non-template control (NTC). The NTC of each primer pair was used to fix the threshold (pink line) that allowed discriminating positive (blue dots) from negative (grey dots) droplets. (**B**) Histogram representation of the number of copies of each transcript per microliter of cDNA, as calculated from the droplet intensities graph shown in A. The mean value and standard deviation of biological triplicates, as well as the result of Student's t-tests for unpaired data are shown; *p< 0.05, ** p< 0.01, *** p< 0.001, NS: non-statistically significant (p> 0.05).



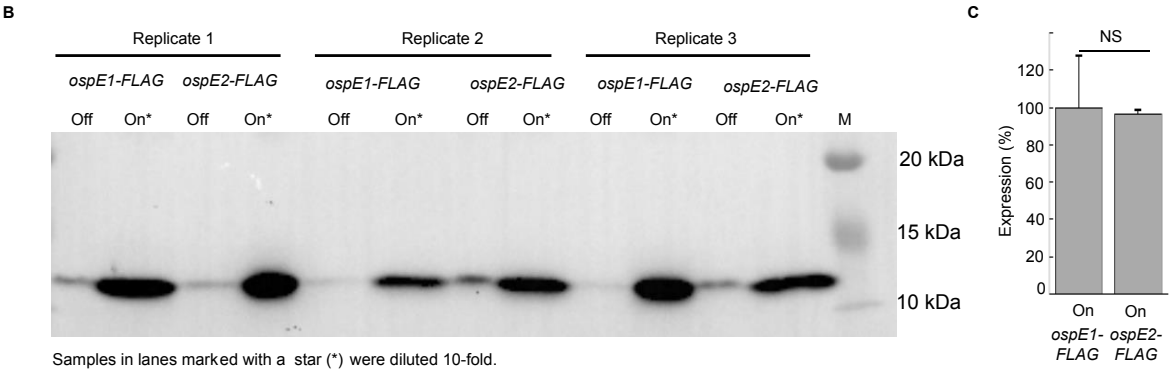


Figure 2.S. 3. Western blot supporting the operon validation data ((Figure 2.3). (**A**) Detection of the expression of IpaH1.4-2xMyc and IpaH2.5-Myc by WB in the off-state and on-state using cell lysate of WT and *ipaD* cells harboring plasmids pUC18Δ *ospE1-ipaH1.4-2xMyc*, or pUC18Δ *ospE2-ipaH2.5-2xMyc.* This membrane represents the three biological triplicates that were used for the quantification shown in Figure 3D. Note that the lysate of the *ipaD* (on-state) pUC18Δ *ospE1-ipaH1.4-2xMyc* cells were diluted 10-fold prior to electrophoresis. (**B**) Detection of the expression of OspE1-FLAG and OspE2-FLAG by WB in the off- and on-state using cell lysate of WT and *ipaD* cells harboring plasmids pUC18Δ *ospE1-FLAG-ipaH1.4-2xMyc* and pUC18Δ *ospE2-FLAG-ipaH2.5-2xMyc* in biological triplicates. M: Protein marker. C. Histogram plot of the relative expression of OspE1-FLAG and OspE2-FLAG in the on-state was estimated from the density of bands in the panel B. The mean value and standard deviation, as well as the result of Student's t-tests for unpaired data are shown; *p< 0.05, ** p< 0.01, *** p< 0.001, ****p<0.0001, NS: non-statistically significant (p> 0.05).

Table 2.S. 1. Expression data of cistrons from Operons O1-O7 and of RNA mapped to non-coding intercistronic sequence (adapted from Table 2.S.3).

Start	End	Strand	Operon names	Genes or RNA name	T3SA off-state (RPKM)	T3SA on-state (RPKM)	FC on/off
3477	4505	+	O1	*ospB*	72	685	9.51
4663	4679	+		RNA_a	120	564	4.70
4680	5796	+		*phoN2*	837	4216	5.04
11475	7804	-	O2	*ospD2*	72	106	1.47
11539	11476	-		RNA_b	382	1314	3.44
12361	11540	-		*ospF*	745	2116	2.84
40297	41471	+	O3	*ospE2*	127	1727	13.6
43257	46531	+		*ipaH2.5*	76	1020	13.4
64054	65767	+	O4	*ipaH7.8*	39	1294	33.2
65768	70154	+		*ipaH4.5*	43	1248	29.0
76847	72690	-	O5	*ospD3*	63	666	10.6
76881	76848	-		RNA_g	32	636	19.9
78483	76882	-		*ospC1*	17	327	19.2
174324	175980	+	O6	*ipaH9.8*	33	1154	35.0
175981	176551	+		*orf176*	223	1251	5.61
176552	176621	+		RNA_n	145	1155	7.97
176622	177599	+		*ospG*	145	661	4.56
208592	205072	-	O7	*ipaH1.4*	70	1342	19.2
209666	208593	-		*ospE1*	107	2639	24.7

Table 2.S. 2. Primers used for ddPCR and RT-PCR.

Labels	Primer names	Primer sequence (5' -3')	Usage
qPCRO01	IpaH7.8_FWD_s1	CTGAGAATCCTGACTGAATGG	ddPCR analysis
qPCRO02	IpaH7.8_REV_s1	GCTTTGGTAATTCGGGAAGA	ddPCR analysis
qPCRO51	OspD2_FWD_s3	CCATAAACAACCCATGCTTTAC	ddPCR analysis
qPCRO52	OspD2_REV_s3	GCGAAAGACAGCCATGAT	ddPCR analysis
qPCRO61	RS13510_FWD_s3 (gem1)	CAGCTTACCCAGAATCCATC	ddPCR analysis
qPCRO62	RS13510_REV_s3 (gem1)	GGATCGTCGTTTACTCGTATG	ddPCR analysis
qPCRO67	RS15810_FWD_s1 (gem2)	ACGTACCGCTCTGGTAAT	ddPCR analysis
qPCRO68	RS15810_REV_s1 (gem2)	CTAAAGCCATGCGGGATAC	ddPCR analysis
qPCRO57	RS23500_FWD_s3 (gem3)	CACGCCAGAACCTTGTTAG	ddPCR analysis
qPCRO58	RS23500_REV_s3 (gem3)	ACTTGCCTCAGCTTCTACTA	ddPCR analysis
qPCR87	RecA_FWD_s2	ATCTACGGACCGGAATCTT	ddPCR analysis
qPCR88	RecA_REV_s2	GATATCGACGCCCAGTTTAC	ddPCR analysis
qPCR95	ospF_FWD_s3	TTCTCAGCAATCTCGTGTGG	ddPCR analysis
qPCR96	ospF_REV_s3	TAACATCTGACGCCGGATAC	ddPCR analysis
HMIO158	OspB_50_ncRNA_S	AAAGAAGCTCTCAGAAAAAGACTTTTATTATTAATAAAGAACTGG	RT-PCR analysis
HMIO159	PhoN2_50_ncRNA_R	AAAATCATATTTGTAGCAATACAAAAAAGAAGAAAGTTTTTGG	RT-PCR analysis
HMIO151	OspF_50_ncRNA_S	AAGAGCAAATGTTACGAGAGGAACCATTTTATCG	RT-PCR analysis
HMIO152	OspD2_50_ncRNA_R	GAATTTTTTGTTGAAAATATACTAGATGAAAATGTTTTATTTAACGG	RT-PCR analysis
HMIO147	OspE2_ncRNAd_S	GTAATGCTGCCAACTTATTG	RT-PCR analysis
HMIO148	IpaH2.5_ncRNAd_R	CCCCGTTACTTTATTCGTAC	RT-PCR analysis
HMIO149	IpaH7.8_50_ncRNA_S	TACTGGCCCTGCGATTGTC	RT-PCR analysis

HMIO150	IpaH4.5_50_ncRNA_R	GATAAGCCACAAAGGGAACGA	RT-PCR analysis
HMIO156	OspC1_50_ncRNA_S	TAAAACTATTATTGAAATATGGTGCAACATCTGAC	RT-PCR analysis
HMIO157	OspD3_50_ncRNA_R	ATATTTTGCAAACATATTTTCCTTGATGGGA	RT-PCR analysis
HMIO154	IpaH9.8_50_ncRNA_S	TACTGGCCCTGCGATTGTTTG	RT-PCR analysis
HMIO155	OspG_50_ncRNA_R	TTATTCTCAAATGGAAAAGGTGTTTGAATAATGG	RT-PCR analysis
HMIO145	OspE1_ncRNAd_S	CGATACGTACTGAACGTTATC	RT-PCR analysis
HMIO146	IpaH1.4_ncRNAd_R	GTACCCTTTCTCCGGGTA	RT-PCR analysis

Table 2.S. 3. Expression data of virulence plasmid genes according to RNA-Seq data analyzed with Rockhopper.

Start	End	Strand	Genes name	T3SA off-state (RPKM)	T3SA on-state (RPKM)	FC on/off	q-value
173396	173643	+	*ccdA*	130	151	1.16	0.35
173624	174171	+	*ccdB*	112	140	1.25	0.43
146225	150119	+	*icsA*	221	169	0.76	0.58
111459	109900	-	*icsB*	568	695	1.22	0.64
265	3343	+	*icsP*	863	480	0.56	0.16
104415	102514	-	*ipaA*	655	661	1.01	0.75
108326	106565	-	*ipaB*	2026	2732	1.35	0.78
106564	105473	-	*ipaC*	3541	4933	1.39	0.7
105472	104416	-	*ipaD*	1035	526	0.51	0.1
208592	205072	-	*ipaH1.4*	70	1342	19.2	4.94E-36
43257	46531	+	*ipaH2.5*	76	1020	13.4	1.44E-18
65768	70154	+	*ipaH4.5*	43	1248	29.0	3.49E-33
64054	65767	+	*ipaH7.8*	39	1294	33.2	9.84E-34
174324	175980	+	*ipaH9.8*	33	1154	35.0	6.98E-22
98160	100669	+	*ipaJ*	433	488	1.13	1
109899	109483	-	*ipgA*	125	128	1.02	0.27
109482	108800	-	*ipgB1*	1465	1621	1.11	0.9
22749	26131	+	*ipgB2*	1085	716	0.66	0.23
108799	108327	-	*ipgC*	1724	1999	1.16	0.95
111285	113323	+	*ipgD*	851	1110	1.30	0.19
113324	113686	+	*ipgE*	887	939	1.06	1
113686	114152	+	*ipgF*	1047	1248	1.19	1
192175	189440	-	*mvpA*	105	133	1.27	0.45
192439	192175	-	*mvpT*	201	249	1.24	0.4
122073	124157	+	*mxiA*	480	481	1	1
121005	122072	+	*mxiC*	861	851	0.99	1
119286	121004	+	*mxiD*	541	512	0.95	1
118667	119299	+	*mxiE*	440	423	0.96	0.41
114153	115275	+	*mxiG*	441	486	1.1	1
115276	115539	+	*mxiH*	1826	1498	0.82	0.34
115540	115833	+	*mxiI*	2336	2495	1.07	1
115834	116564	+	*mxiJ*	1290	1427	1.11	0.86
116561	117088	+	*mxiK*	1242	1269	1.02	1
117718	118125	+	*mxiL*	134	128	0.96	0.22

118109	118666	+	*mxiM*	356	354	0.99	0.27
117030	117725	+	*mxiN*	893	894	1.00	1
17282	12362	-	*orf13*	262	202	0.77	0.08
130774	131132	+	*orf131a*	186	154	0.83	0.15
131133	132280	+	*orf131b*	274	190	0.69	0.1
137574	135578	-	*orf136*	279	247	0.89	0.16
138375	137578	-	*orf137*	132	115	0.87	0.15
158004	156516	-	*orf157*	372	340	0.91	0.17
170681	166839	-	*orf169b*	143	141	0.99	0.19
175981	176551	+	*orf176*	223	1251	5.61	1
185229	180937	-	*orf182*	299	304	1.02	1
185270	186211	+	*orf185*	191	140	0.73	0.06
186212	187302	+	*orf186*	175	126	0.72	0.15
47220	45746	-	*orf47*	242	263	1.09	0.24
47586	47211	-	*orf48*	214	239	1.12	0.23
86071	82695	-	*orf85*	491	574	1.17	1
89861	86059	-	*orf86*	402	481	1.20	0.37
3477	4505	+	*ospB*	72	685	9.51	0.15
78483	76882	-	*ospC1*	17	327	19.2	9.74E-4
51400	49039	-	*ospC2*	876	544	0.62	0.36
89920	93158	+	*ospC3*	193	166	0.86	0.23
18104	22099	+	*ospD1*	470	481	1.02	0.83
11475	7804	-	*ospD2*	72	106	1.47	0.63
76847	72690	-	*ospD3*	63	666	10.6	1.11E-5
209666	208593	-	*ospE1*	107	2639	24.7	2.59E-07
40297	41471	+	*ospE2*	127	1727	13.6	0.06
12361	11540	-	*ospF*	745	2116	2.84	0.02
176622	177599	+	*ospG*	145	661	4.56	1
28937	30219	+	*parA*	313	348	1.11	1
30219	32223	+	*parB*	307	327	1.07	0.88
155779	156933	+	*phoN1*	352	317	0.90	0.27
4680	5796	+	*phoN2*	837	4216	5.04	1.01E-12
202528	202851	+	*repB*	270	250	0.93	0.18
58786	53926	-	*sepA*	677	904	1.34	0.76
125856	126286	+	*spa13*	554	538	0.97	0.55
124158	124559	+	*spa15*	749	573	0.77	0.17
128027	128681	+	*spa24*	369	310	0.84	0.27
126273	127151	+	*spa32*	614	526	0.86	0.54
127145	128026	+	*spa33*	306	276	0.9	0.27

129733	130773	+	*spa40*	162	151	0.93	0.27
124560	125855	+	*spa47*	2005	2066	1.03	1
128682	128956	+	*spa9*	430	390	0.91	0.17
202854	203113	+	*tap*	167	193	1.16	0.27
150121	154159	+	*ushA*	296	234	0.79	0.32
145866	141286	-	*virA*	320	2133	6.67	4.76E-14
102064	100813	-	*virB*	1748	1219	0.7	1
39342	37446	-	*virF*	158	136	0.86	0.15
187303	188257	+	*virK*	104	79	0.76	0.06
4663	4679	+	RNA_a	120	564	4.7	1
11539	11476	-	RNA_b	382	1314	3.44	1
34497	34319	-	RNA_c[1]	523	507	0.97	0.19
52252	52448	+	RNA_d	512	873	1.71	0.81
52975	53120	+	RNA_e[1]	513	507	0.99	0.18
70899	70755	-	RNA_f[1]	513	507	0.99	0.18
76881	76848	-	RNA_g	32	636	19.9	7.64E-5
133039	132999	-	RNA_h[2]	474	477	1.01	0.18
133854	133770	-	RNA_i	412	471	1.14	0.28
139851	140187	+	RNA_j[3]	425	867	2.04	1
140376	140439	+	RNA_k[4]	438	449	1.03	0.27
140677	140822	+	RNA_l[1]	516	510	0.99	0.18
173200	173076	-	RNA_m	488	552	1.13	0.25
176552	176621	+	RNA_n	145	1155	7.97	1.12E-45
180113	180454	+	RNA_o[3]	426	865	2.03	1
180643	180706	+	RNA_p[4]	438	449	1.03	0.27
180944	181073	+	RNA_q[1]	516	510	0.99	0.2
200671	200631	-	RNA_r[2]	474	477	1.01	0.18

Legend: the exponent numbers to the right of the RNA names identify transcripts that are identical but that mapped to different area of the virulence plasmid (100% identical with full or partial coverage).

Table 2.S. 4. Expression data of chromosomal non-coding genes according to RNA-Seq data analysed with Rockhopper

Start	End	Transcript Length	Strand	Genes name	Product	T3SA off-state (RPKM)	T3SA on-state (RPKM	FC on/off	q-value
3737868	3737949	82	+	*rrlA*	AS_rrlA_1b	502	214	0.43	1.46E-07
3737216	3737506	291	+	*rrlA*	AS_rrlA_2	555	235	0.42	5.43E-06
3735927	3736794	868	+	*rrlA*	AS_rrlA_3	607	253	0.42	2.82E-05
4166778	4166647	132	-	*rrlB*	AS_rrlB_1	541	229	0.42	8.22E-05
4167383	4167093	291	-	*rrlB*	AS_rrlB_2	556	236	0.42	5.41E-06
4168672	4167805	868	-	*rrlB*	AS_rrlB_3	607	253	0.42	2.82E-05
3834632	3834713	82	+	*rrlC*	AS_rrlC_1b	502	214	0.43	1.46E-07
3833980	3834270	291	+	*rrlC*	AS_rrlC_2	555	235	0.42	5.43E-06
3832691	3833558	868	+	*rrlC*	AS_rrlC_3	607	253	0.42	2.82E-05
3375705	3375786	82	+	*rrlD*	AS_rrlD_1b	502	214	0.43	1.46E-07
3375053	3375343	291	+	*rrlD*	AS_rrlD_2	555	235	0.42	5.43E-06
3373764	3374631	868	+	*rrlD*	AS_rrlD_3	607	253	0.42	2.82E-05
4207214	4207133	82	-	*rrlE*	AS_rrlE_1b	502	214	0.43	1.46E-07
4207866	4207576	291	-	*rrlE*	AS_rrlE_2	555	235	0.42	5.43E-06
4209155	4208288	868	-	*rrlE*	AS_rrlE_3	607	253	0.42	2.82E-05
2726009	2726299	291	+	*rrlG*	AS_rrlG_2	555	235	0.42	5.43E-06
2724814	2725587	774	+	*rrlG*	AS_rrlG_3b	608	255	0.42	3.21E-05
217114	217033	82	-	*rrlH*	AS_rrlH_1b	502	214	0.43	1.46E-07
217766	217476	291	-	*rrlH*	AS_rrlH_2	555	235	0.42	5.43E-06
219055	218188	868	-	*rrlH*	AS_rrlH_3	607	253	0.42	2.82E-05
3739434	3739582	149	+	*rrsA*	AS_rrsA_2	543	230	0.42	3.21E-05
3739138	3739318	181	+	*rrsA*	AS_rrsA_3	534	238	0.45	9.65E-05
3739835	3740024	190	+	*rrsA*	AS_rrsA_4b	498	211	0.42	5.43E-06
4165165	4165016	150	-	*rrsB*	AS_rrsB_2	546	231	0.42	2.84E-05
4165461	4165281	181	-	*rrsB*	AS_rrsB_3	535	238	0.44	9.65E-05
4164716	4164635	82	-	*rrsB*	AS_rrsB_4c	472	203	0.43	3.12E-07
3836410	3836425	16	+	*rrsC*	AS_rrsC_1	466	196	0.42	3.42E-08
3836198	3836350	153	+	*rrsC*	AS_rrsC_2	552	233	0.42	2.13E-05
3835902	3836082	181	+	*rrsC*	AS_rrsC_3	534	238	0.45	9.65E-05
3836486	3836788	303	+	*rrsC*	AS_rrsC_4	545	230	0.42	3.81E-06
3377271	3377419	149	+	*rrsD*	AS_rrsD_2	543	230	0.42	3.21E-05
3376975	3377155	181	+	*rrsD*	AS_rrsD_3	534	238	0.45	9.65E-05
3377672	3377861	190	+	*rrsD*	AS_rrsD_4b	498	211	0.42	5.43E-06
4205436	4205421	16	-	*rrsE*	AS_rrsE_1	466	196	0.42	3.42E-08

4205648	4205496	153	-	*rrsE*	AS_rrsE_2	552	233	0.42	2.13E-05
4205944	4205764	181	-	*rrsE*	AS_rrsE_3	535	238	0.44	9.65E-05
4205360	4205058	303	-	*rrsE*	AS_rrsE_4	545	230	0.42	3.81E-06
2728439	2728454	16	+	*rrsG*	AS_rrsG_1	466	196	0.42	3.42E-08
2728227	2728379	153	+	*rrsG*	AS_rrsG_2	552	233	0.42	2.13E-05
2727931	2728111	181	+	*rrsG*	AS_rrsG_3	534	238	0.45	9.65E-05
2728515	2728817	303	+	*rrsG*	AS_rrsG_4	545	230	0.42	3.81E-06
2726661	2726742	82	+	*rrsG*	AS_rrsG_4c	502	214	0.43	1.46E-07
215336	215321	16	-	*rrsH*	AS_rrsH_1	466	196	0.42	3.42E-08
215548	215396	153	-	*rrsH*	AS_rrsH_2	552	233	0.42	2.13E-05
215844	215664	181	-	*rrsH*	AS_rrsH_3	535	238	0.44	9.65E-05
215260	214958	303	-	*rrsH*	AS_rrsH_4	545	230	0.42	3.81E-06
3724587	3724373	215	-	ND	RNA_up30_yihI	4653	2182	0.47	0.042

Table 2.S. 5. Expression data of chromosomal coding genes according to RNA-Seq data analysed with Rockhopper.

Start	End	Strand	Gene name	T3SA off-state (RPKM)	T3SA on-state (RPKM)	FC on/off	q-value
4280519	4282483	+	*acs*	48	18	0.38	4.02E-07
465659	466284	+	*apt*	172	84	0.49	0.003
28611	29844	+	*carA*	43	22	0.51	0.003
29845	33073	+	*carB*	38	21	0.55	0.004
1239688	1241014	+	*dadA*	109	38	0.35	4.59E-07
1241015	1242265	+	*dadB*	79	28	0.35	7.08E-07
3642846	3644195	+	*dctA*	207	121	0.58	0.032
2467659	2469093	+	*fadL*	96	43	0.45	2.71E-04
3366481	3366869	+	*fis*	423	238	0.56	0.004
1666460	1664722	-	*fumA*	129	76	0.59	6.70E-04
609341	610376	+	*galE*	73	41	0.56	0.011
610377	611426	+	*galT*	65	38	0.58	0.011
912627	912342	-	*infA*	642	382	0.60	0.002
4033330	4031991	-	*lldD*	241	132	0.55	0.017
4035868	4034106	-	*lldP*	390	213	0.55	0.099
4034118	4033330	-	*lldR*	196	110	0.56	2.79E-04
2259425	2257840	-	*mglA*	99	45	0.45	6.88E-05
2260638	2259428	-	*mglB*	360	178	0.49	0.003
2257839	2256684	-	*mglC*	51	23	0.45	1.09E-04
2609730	2610791	+	*purM*	54	31	0.57	0.022
1130062	1128992	-	*pyrC*	53	31	0.58	0.024
1000834	1001864	+	*pyrD*	46	20	0.43	7.08E-07
858745	857338	-	*rimO*	61	29	0.48	1.22E-04
20285	19836	-	*rpsT*	1360	815	0.60	0.029
644533	642770	-	*sdhA*	375	200	0.53	0.090
642769	642038	-	*sdhB*	571	286	0.50	0.002
645497	644877	-	*sdhC*	420	192	0.46	1.41E-05
644880	644534	-	*sdhD*	234	119	0.51	7.35E-04
4158094	4156418	-	*sthA*	165	79	0.48	8.38E-07
638935	637608	-	*sucB*	692	339	0.49	0.079
637607	636444	-	*sucC*	670	344	0.51	0.099
636443	635445	-	*sucD*	675	351	0.52	0.030
2648609	2649465	+	*suhB*	172	88	0.51	5.15E-04

946709	944873	-	*ycaO*	50	30	0.60	0.047
2004518	2004771	+	*yecF*	63	35	0.56	3.12E-05
2098509	2096989	-	*yeeF*	38	14	0.37	4.19E-10
2286105	2286949	+	*yeiP*	48	29	0.60	0.024
3365442	3366480	+	*yhdG*	363	218	0.60	0.029
3428819	3430090	+	*yhfC*	135	75	0.56	1.61E-04
3068448	3067917	-	*yqgB*	59	28	0.47	6.12E-05

Appendix 2: Supplementary Figures and Tables for chapter 3

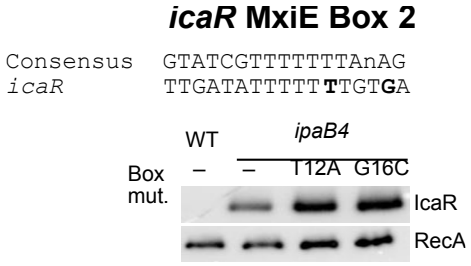


Figure 3.S. 1. MxiE-Box 2 of *icaR* is not active. Mutations of this box did not affect the production of IcaR

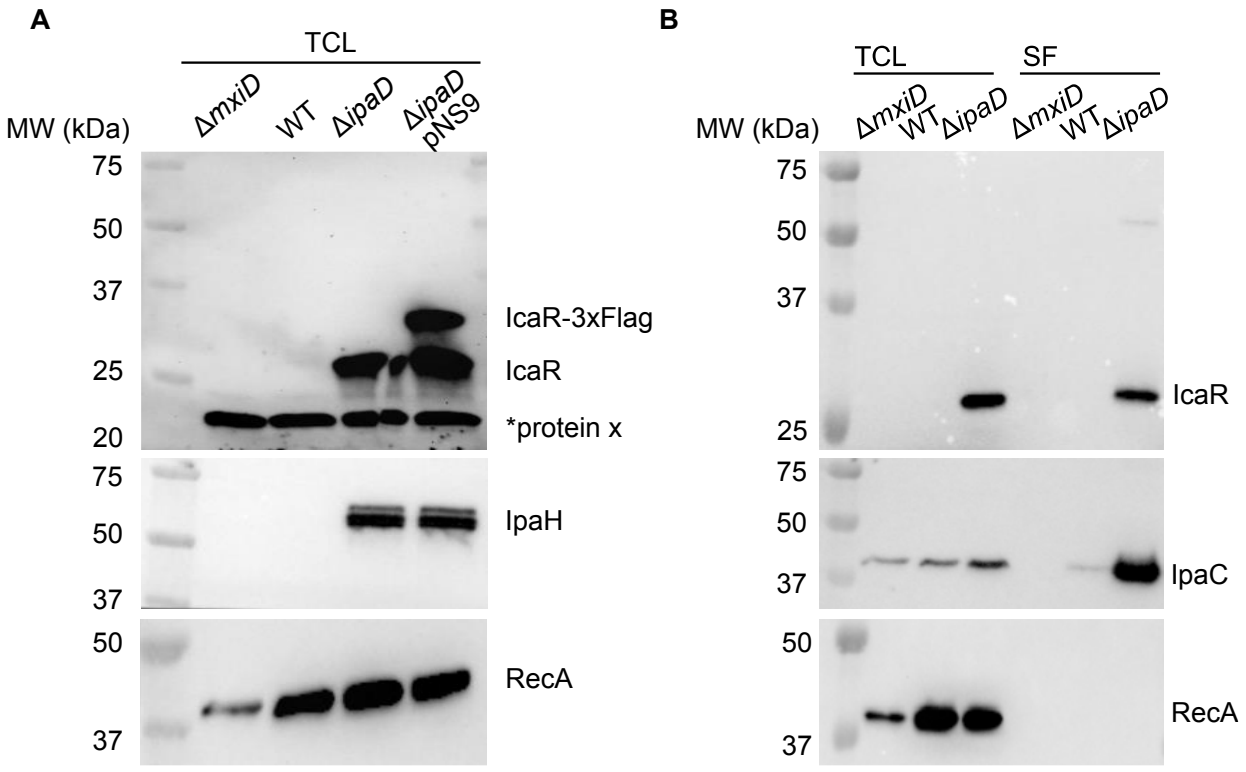


Figure 3.S. 2. IcaR produced from its endogenous locus is secreted in a T3SA-dependent fashion. (**A**) Detection of endogenous and recombinant IcaR by immunoblotting with a polyclonal antibody in the total cell lysate of the indicated *S. flexneri* 5a str. M90T derivatives grown in TSB at 37°C. pNS9 harbors *lacZ*p::*icaR*-3xFLAG. Hence, Δ*ipaD* pNS9 displayed two bands corresponding to the endogenous and recombinant IcaR at 27 kDa and 31 kDa, respectively. Protein X with an apparent MW of 20 kDa cross-reacted with the antibody, but it was easily distinguished by its mass and its constitutive production. (**B**) Detection of endogenous IcaR by immunoblotting in the total cell lysate and the secreted fraction of the indicated *S. flexneri* 5a str. M90T derivatives grown in TSB at 37°C. The endogenous IcaR was detected with a polyclonal antibody raised against its residues 143-158.

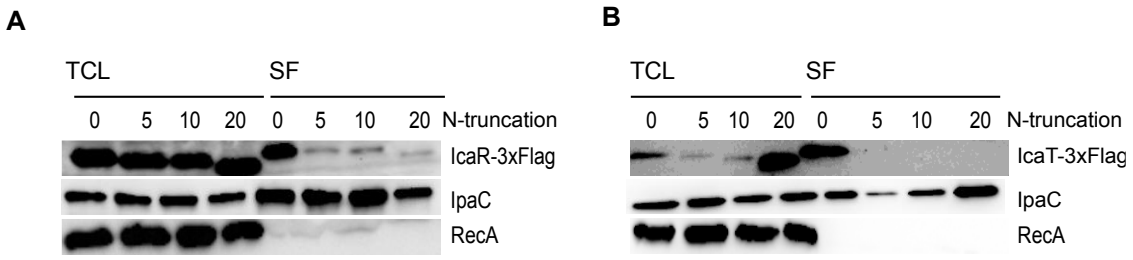


Figure 3.S. 3. The secretion of IcaR and IcaT depends on their amino-terminus in Δ*ipaD* strain. (**A** and **B**) Detection of IcaR and IcaT N-terminal truncation mutants (0, 5, 10 or 20 residues truncation) by immunoblotting with a FLAG-antibody in the total cell lysate (TCL) and the secreted fraction (SF) of the indicated *S. flexneri* Δ*ipaD* strains harboring plasmid-born *icaR* and *icaT* placed under the control of the lacZ promoter and grown in TSB at 37°C. IpaC, which is a T3SA substrate, and RecA, a housekeeping protein, were used as controls.

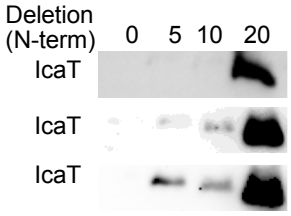


Figure 3.S. 4. N-terminal truncation of 20 residues of IcaT increases its production to a higher level.

Figure 3.S. 5. The locus of *icaR* and *icaT* are well conserved. Loci of selected strains representative of *Shigella* subgroups and *E. coli* phylogroups for (**A**) *icaR* (pink); (**B**) *icaT* (red). Functional genes are in black. Insertion sequences (IS) are dark grey. Genes and pseudo-genes encoding hypothetical protein (hP) are burgundy. The *argW* (green) encodes the $tRNA^{Arg}$ with the anticodon CCU. *shiE* (light grey) is a pseudogene.

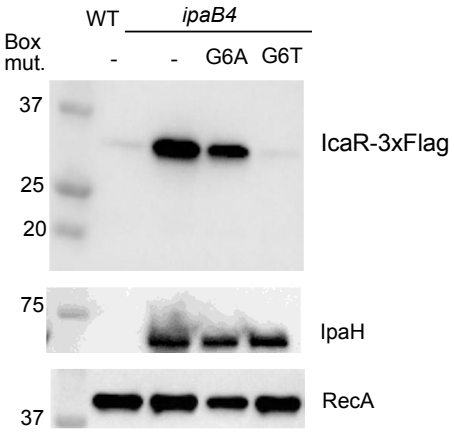


Figure 3.S. 6. T**he MxiE box of *icaR* in EIEC carries a G6T mutation.** This mutation decreases the activity of its promoter.

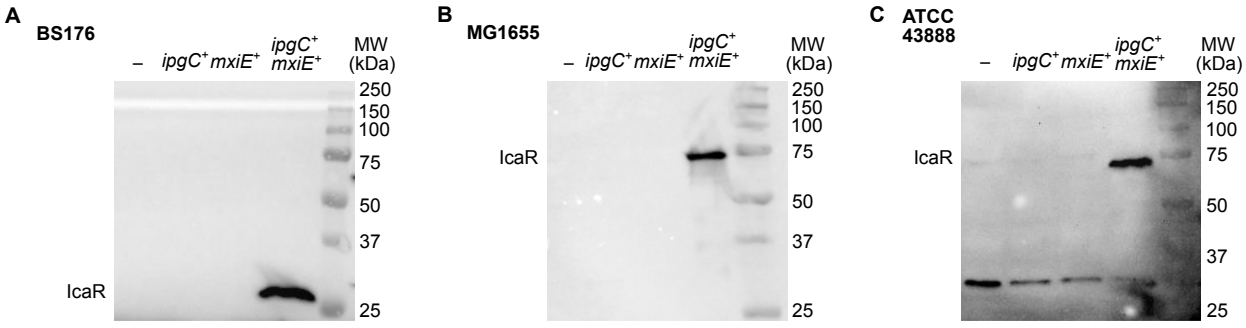


Figure 3.S. 7. Full membrane of immunoblotting corresponding to panels Figure 3.5D-F. Detection of IcaR produced from its endogenous locus by immunoblotting the TCL of strains described in Figure 3.5. (**A**) BS176, full membrane corresponding to Figure 3.5D. (**B**) MG1655, full membrane corresponding to Figure 3.5E. (**C**) ATCC43888, full membrane corresponding to Figure 3.5F.

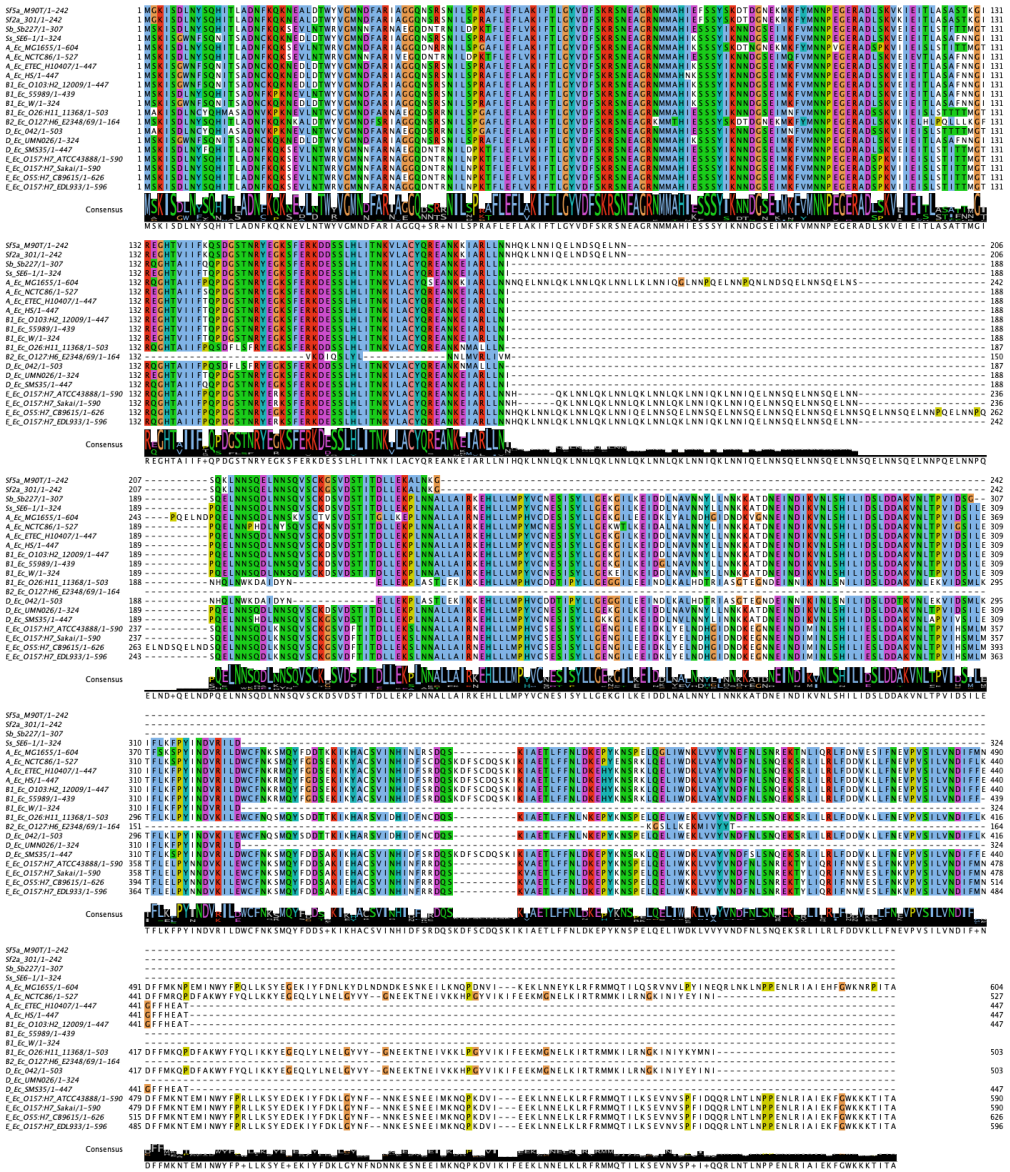
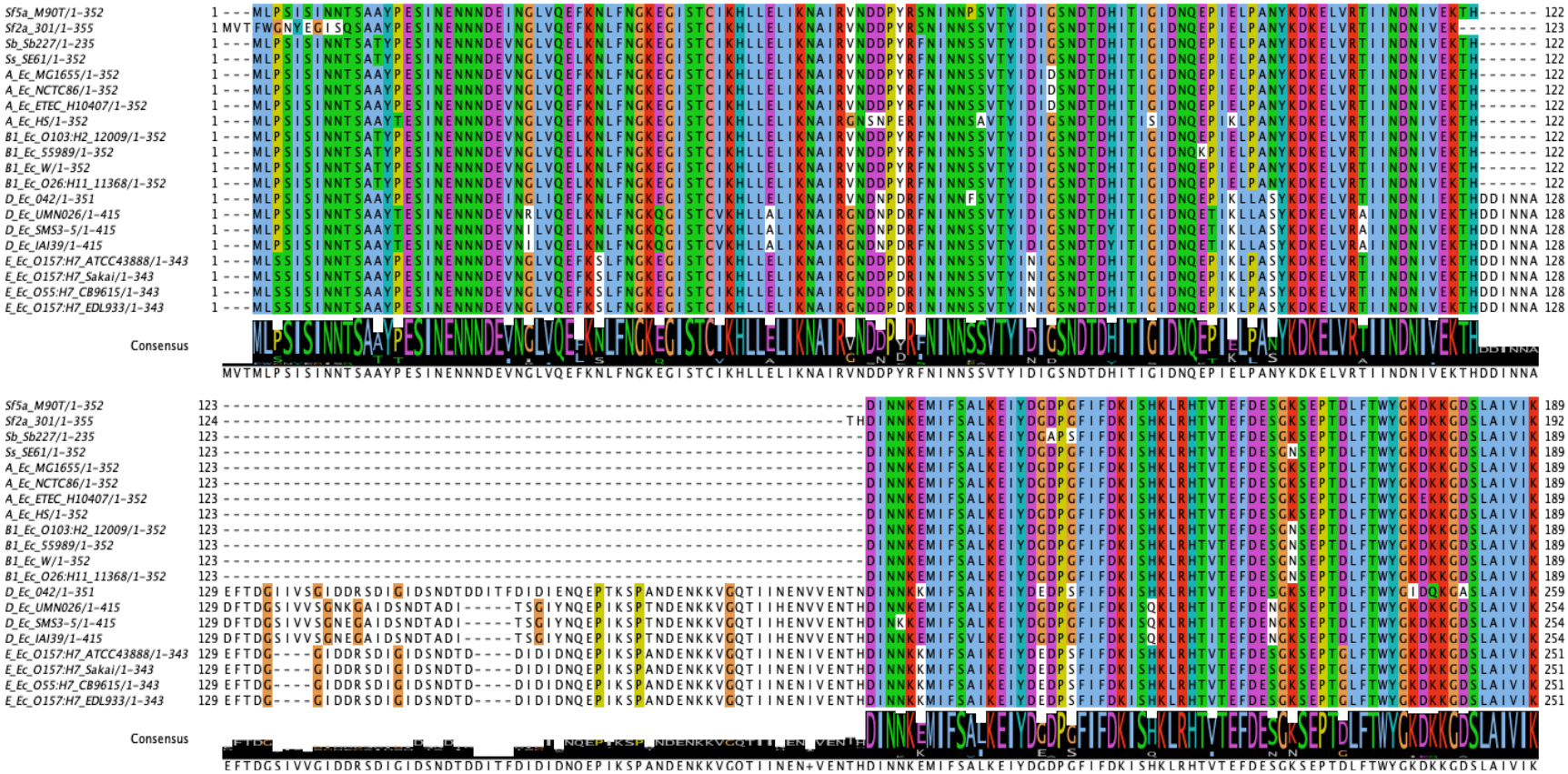


Figure 3.S. 8. The primary structure of IcaR is conserved in *E. coli*. Alignment of primary structure of IcaR from strain representative of *Shigella* subgroups and *E. coli* phylogroups. The prefix of each sequence indicates the *Shigella* subgroups or the *E. coli* phylogroups to which each corresponding strain belongs. Apart from various carboxy-terminal truncations, the sequence is relatively well conserved. See the text for more details


Sf5a_M90T/1-352
Sf2a_301/1-355
Sb_Sb227/1-235
Ss_SE61/1-352
A_Ec_MG1655/1-352
A_Ec_NCTC86/1-352
A_Ec_ETEC_H10407/1-352
A_Ec_HS/1-352
B1_Ec_O103:H2_12009/1-352
B1_Ec_55989/1-352
B1_Ec_W/1-352
B1_Ec_O26:H11_11368/1-352
D_Ec_042/1-351
D_Ec_UMN026/1-415
D_Ec_SMS3-5/1-415
D_Ec_IAI39/1-415
E_Ec_O157:H7_ATCC43888/1-343
E_Ec_O157:H7_Sakai/1-343
E_Ec_O55:H7_CB9615/1-343
E_Ec_O157:H7_EDL933/1-343
Consensus

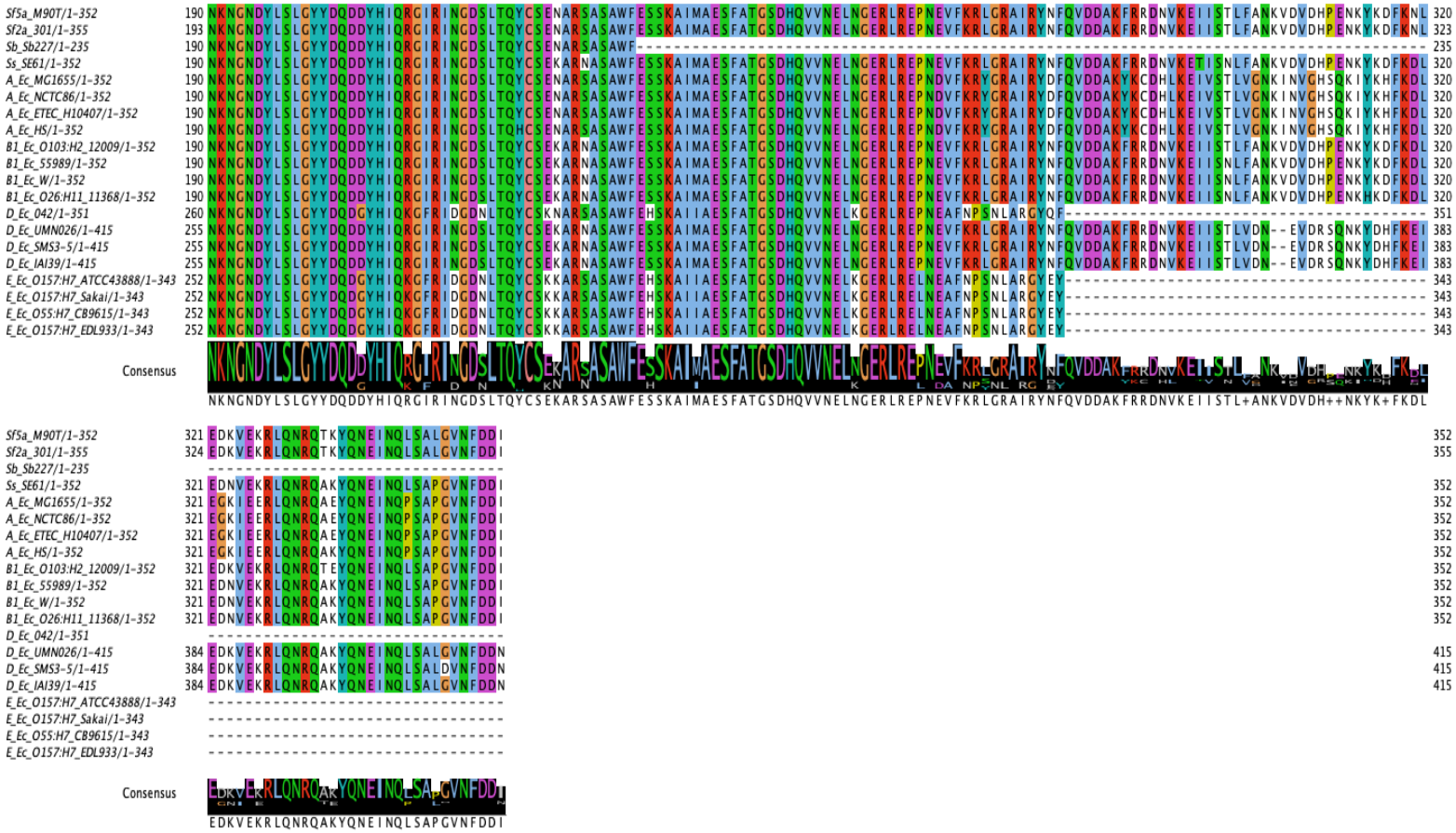


Figure 3.S. 9. The primary structure of IcaT is conserved in *E. coli*. Alignment of primary structure of IcaT from strain representative of *Shigella* subgroups and *E. coli* phylogroups. The prefix of each sequence indicates the *Shigella* subgroups or the *E. coli* phylogroups to which each corresponding strain belongs. Apart from various carboxy-terminal truncations, the sequence is relatively well conserved. See the text for more details

Table 3.S. 1. Occurrence of *icaR* and *icaT* in selected strains of *Shigella* and *Escherichia coli* representative of the main phylogroups

Phylogroups	Strains name	Accession number	*icaR*	*icaT*
Shigella	*S. flexneri* 5a str. M90T	NZ_CM001474.1	Yes	Yes
	S. flexneri 2a str. 301	NC_004337.2	Yes	Yes[9]
	S. boydii Sb227	NC_007613.1	Yes	Yes
	S. sonnei strain SE6-1	NZ_CP055292.1	Yes	Yes
	S. sonnei Ss046	CP000038.1	Yes	Pseudo.[10]
	S. dysenteriae Sd197	NC_007606.1	No[11]	Pseudo.[12]
A	*E. coli* str. K-12 substr. MG1655	U00096.3	Yes	Yes
	E. coli strain NCTC86	NZ_CP019778.1	Yes	Yes
	E. coli ETEC H10407	NC_017633.1	Yes	Yes
	E. coli HS	NC_009800.1	Yes	Yes
B1	*E. coli* O103:H2 str. 12009	NC_013353.1	Yes	Yes
	E. coli 55989	NC_011748.1	Yes	Yes
	E. coli W	NC_017635.1	Yes	Yes
	E. coli O26: H11 str. 11368	NC_013361.1	Yes	Yes
B2	*E. coli* O127:H6 str. E2348/69	NC_011601.1	Yes[13]	No
	E. coli 536	NC_008253.1	No	No
	E. coli SE15	NC_013654.1	No	No
	E. coli ABU 83972	NC_017631.1	No	No
	E. coli CFT073	NZ_CP051263.1	No	No
D	*E. coli* 042	NC_017626.1	Yes	Yes
	E. coli UMN026	NC_011751.1	Yes	Yes
	E. coli SMS-3-5	NC_010498.1	Yes	Yes
	E. coli IAI39	NC_011750.1	No	Yes
E	*E. coli* O157:H7 strain ATCC43888	NZ_CP041623.1	Yes	Yes
	E. coli O157:H7 str. Sakai	NC_002695.1	Yes	Yes
	E. coli O55:H7 str. CB9615	NC_013941.1	Yes	Yes
	E. coli O157:H7 str. EDL933	NZ_CP008957.1	Yes	Yes

[9] An IS sequence is inserted between the MxiE box-containing promoter and the coding sequence, which is integral. *icaT* if still functional may be co-transcribed with the IS.
[10] An IS is between the MxiE box and the coding sequence; the coding sequence is disrupted by 27 stops codons.
[11] Three IS are located in the locus. The coding sequence is almost completely removed, but the MxiE box is still present upstream of one of the IS.
[12] IS induced a large truncation at the 5' end of the coding sequence; the MxiE box is absent.
[13] This strain possesses the largest truncation at the 3' end of *icaR*.

Table 3.S. 2. Global occurrence of *icaR* and *icaT* in *Shigella* subgroups[14].

		Total hits (counts)	Integral (%)[15]	Moderate disruption (%)[16]	High disruption (%)[17]
icaR					
	boydii	26	3.8	96.2	0
	dysenteriae	23	0	60.9	39.1
	flexneri	56	60.7	35.7	3.6
	sonnei	43	0	100	0
icaT					
	boydii	27	11.1	81.5	7.4
	dysenteriae	21	0	23.8	76.2
	flexneri	63	66.7	28.6	4.8
	sonnei	43	2.3	97.7	0

[14] *icaR* and *icaT* from M90T and spanning from the MxiE box to the stop codon were used as queries.

[15] Integral genes were defined as hits ≥99% query coverage, ≥98% pairwise sequence identity, and no gap; Integral genes (counts)/total hits (counts)*100= percentage

[16] Genes with moderate disruption were defined as hits ≥80% query coverage subtracted by the integral genes; Moderately disrupted genes (counts)/total hits (counts)*100= percentage

[17]. Genes with high disruption were defined as hits <80% query coverage; highly disrupted genes (counts)/total hits (counts)*100= percentage.

Table 3.S. 3. The plasmids used in this work.

Plasmid ID	Plasmid name	Properties	Reference
pNS1	pUC18D *icaR*-3xFlag	*icaR*p, str. M90T	This study
pNS2	pUC18D *icaT*-3xFlag	*icaT*p, str. M90T	This study
pNS3	pUC18D G6C *icaR*-3xFlag	G6C MxiE-Box mutation	This study
pNS4	pUC18D G6C *icaT*-3xFlag	G6C MxiE-Box mutation	This study
pNS5	pUC18D T12A *icaR*-3xFlag	T12A MxiE-Box mutation	This study
pNS6	pUC18D T12A *icaT*-3xFlag	T12A MxiE-Box mutation	This study
pNS7	pUC18D A16C *icaR*-3xFlag	A16C MxiE-Box mutation	This study
pNS8	pUC18D A16C *icaT*-3xFlag	A16C MxiE-Box mutation	This study
pNS9	pUC18D *lacZ*p::*icaR*-3xFlag	*lacZ*p fusion with *icaR*'s CDS	This study
pNS10	pUC18D *lacZ*p::*icaT*- 3xFlag	*lacZ*p fusion with *icaT*'s CDS	This study
pNS11	pUC18D *lacZ*p::*icaR* D15-3xFlag	Residues 2-6 (D5) of IcaR are truncated	This study
pNS12	pUC18D *lacZ*p::*icaT* D15-3xFlag	Residues 2-6 (D5) of IcaR are truncated	This study
pNS13	pUC18D *lacZ*p::*icaR* D30-3xFlag	Residues 2-11 (D10) of IcaR are truncated	This study
pNS14	pUC18D *lacZ*p::*icaT* D30-3xFlag	Residues 2-11 (D10) of IcaT are truncated	This study
pNS15	pUC18D *lacZ*p::*icaR* D60-3xFlag	Residues 2-21 (D20) of IcaR are truncated	This study
pNS16	pUC18D *lacZ*p::*icaT* D60-3xFlag	Residues 2-21 (D20) of IcaT are truncated	This study
pNS17	pSU2.1tt *icaR*1-63-bla$_{TEM3}$ M182T	*icaR*p; produces IcaR residues 1-21 fused to bla	This study
pNS19	pSU2.1tt *icaT*1-63- bla$_{TEM3}$ M182T	*icaT*p; produces IcaT residues 1-21 fused to bla	This study
N/A	pSU2.1tt *ospD1*sh M31L bla$_{TEM3}$ M182T	*ospD1*p; produces OpsD1 residues 1-80 M31L fused to bla	(Pinaud et al., 2017b)

pNS21	pUC18.1 *lacZ*p::*ipgC*-3xFlag	str. M90T	Unpublished work by FXCV
pNS22	pSU2.1 *lacZ*p::*mxiE*-2xMyc	str. M90T	Unpublished work by FXCV
pNS23	pUC18.1 *lacZ*p::*ygeG*-3xFlag	O157:H7 str. ATCC43888	This study
pNS24	pSU2.1 *lacZ*p::*eivF*-2xMyc	O157:H7 str. ATCC43888	This study

Table 3.S. 4. The oligonucleotides used in this work.

Primer	Sequence (5'-3')	Description
HMIO117 HMIO118	AGAGAGGGTACCTAACGCGTTCGCCTGGATAAAG AGAGAGTCCGGAAATATCATCAAAATTAACACCTAGT GCAG	Amplification of *icaT* from M90T
HMIO119 HMIO120	AGAGAGGGTACCTGACTTTTCGCCGTAAATAACTCC AGAGAGTCCGGACCCTTTATTCAATGCTTTTTCTAATA A ATCC	Amplification of *icaR* from M90T
HMIO121 HMIO122	AAGTGTAAAGCCTGGGGTGCCTAAGGTACCGAGCTC GAATTCAAA TATGCTTCCGGCTCGTATGTTGTGTGGAAACAAATCAA T CAACATGGAATAAAATC	Insertion of lac-promoter upstream of *icaT*
HMIO121 HMIO123	AAGTGTAAAGCCTGGGGTGCCTAAGGTACCGAGCTC GAATTCAAA TATGCTTCCGGCTCGTATGTTGTGTGGAACGTT CTTATTAACTCAAGGAGTTCGT	Insertion of lac-promoter upstream of *icaR*
HMIO129 HMIO130	AGATAGTCTATTTCATTAGGTAATATATATTTG AAAAACGCTCATTTATTTAGA	*icaT* MxiE box mutation T12A
HMIO131 HMIO132	AGATATTTTTTTGTGAGTAAAATTTG AAAAACGCTCTTTTCACA	*icaR* MxiE box mutation T12A
HMIO133 HMIO134	CTTTTTTGATAGTCTATTTCATTAGG GCTCATTTATTTAGAACCTATC	*icaT* MxiE box mutation G6C
HMIO136 HMIO137	CTTTTTTGATATTTTTTTGTGAG GCTCTTTTCACAAACTCG	*icaR* MxiE box mutation G6C
HMIO139 HMIO141	CGTCTATTTCATTAGGTAATATATATTTGT ATCAAAAAACGCTCATTTATTTAG	*icaT* MxiE box mutation A16C
HMIO142 HMIO144	CTTTTTTTGTGAGTAAAATTTGTAA ATCAAAAAACGCTCTTTTC	*icaR* MxiE box mutation A16C
HMIO271 HMIO272	ATCAACAATACCAGCGCAGCTTAC CATGATTTTATTCCATGTTGATTGATTTGTTTCC	5' deletion of the first 15 bp of *icaT*
HMIO273 HMIO272	GCAGCTTACCCAGAATCCATCA CATGATTTTATTCCATGTTGATTGATTTGTTTCC	5' deletion of the first 30 bp of *icaT*
HMIO274 HMIO272	AACAATGATGAAATTAATGGATTAGTACAAGAGTTCA CATGATTTTATTCCATGTTGATTGATTTGTTTCC	5' deletion of the first 60 bp of *icaT*
HMIO276 HMIO275	TTGAATTATTCTCAACACATTACATTAGC CATGACGAACTCCTTGAGTTAATAAGAAC	5' deletion of the first 15 bp of *icaR*
HMIO277 HMIO275	CACATTACATTAGCCGACAATTTTAAACA CATGACGAACTCCTTGAGTTAATAAGAAC	5' deletion of the first 30 bp of *icaR*
HMIO278 HMIO275	AAAAATGAAGCCTTAGATACCTGGTATGTG CATGACGAACTCCTTGAGTTAATAAGAAC	5' deletion of the first 60 bp of *icaR*
ddO1 ddO2	ACTCAGACATACGGTAACGGAA ACAATAGCGAGAGAGTCG	ddPCR of *icaT*
ddO7 ddO8	CAACACATTACATTAGCCGACA CTTCGTTGGAGCGTTTGCTA	ddPCR of *icaR*
qPCR87 qPCR88	ATCTACGGACCGGAATCTT GATATCGACGCCCAGTTTAC	ddPCR of *recA*
HMIO546 HMIO547	CCCGGGGGTGGCGGATCC GAATTCGTAATCTTAGCTAGTTACTCGAGGTCATAG	Amplification of plasmid pSU2.1tt-TEM3-M182T (20)

HMIO548 HMIO549	CTAGCTAAGATTACGAATTCACGTTCTTATTAACTCAAGGAG TGGGATCCGCCACCCCCGGGCCCTTTATTCAATGCTTTTTCTAATAAATC	Amplification of *icaR* with his endogenous SD from pNS9
HMIO550 HMIO551	CTAGCTAAGATTACGAATTCAACAAATCAATCAACATGGAATAAAATC TGGGATCCGCCACCCCCGGGAATATCATCAAAATTAACACCTAGTG	Amplification of *icaT* with his endogenous SD from pNS10
HMIO574 HMIO578	CCCGGGGGTGGCGGATCC TTGTTTAAAATTGTCGGCTAATGTAATGTGTTGAG	Construction of *icaR* 1-60 bp
HMIO276 HMIO582	TTGAATTATTCTCAACACATTACATTAGC CATGACGAACTCCTTGAGTTAATAAG	5’ deletion of the first 15 bp of *icaR*
HMIO574 HMIO576	CCCGGGGGTGGCGGATCC ATTTTCATTGATGGATTCTGGGTAAGCTGC	Construction of *icaT* 1-60 bp
HMIO271 HMIO580	ATCAACAATACCAGCGCAGCTTAC CATGATTTTATTCCATGTTGATTGATTTG	5’ deletion of the first 15 bp of *icaT*
HMIO595 HMIO596	TTCTAGGAGGTGCTAGCGCTATGATTGAAGAAGGGCTGTTAC AGTTTCTGTTCCTCGAGCCCTAACTTTTTACATAGTTCTCTCGG	Amplification of *eivF* from ATCC43888 for Gibson cloning
HMIO605 HMIO606	TTCGCTAGGAGAAATTAACCATGGACACAGAAACAATTG AAATATTC TGATCTTTATAATCGGATCCGCCATTATCTTCTGAATTATCGG	Amplification of *ygeG* from ATCC43888 for Gibson cloning
HMIO593 HMIO594	GGGCTCGAGGAACAGAAAC AGCGCTAGCACCTCCTAG	Amplification of plasmid pSU2.1 for Gibson cloning
HMIO603 HMIO604	GGATCCGATTATAAAGATCATGAC GTTAATTTCTCCTAGCGAATTC	Amplification of plasmid pUC18.1 for Gibson cloning

Appendix 3: Supplementary Figures and Tables for chapter 4

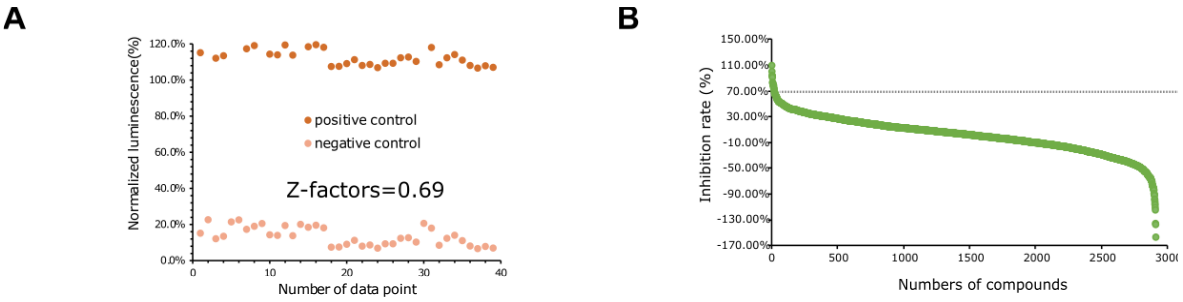


Figure 4.S. 1. Validation of the screening. (**A**) Discrimination power between negative and positive control strains with Z-factor of 0.69 (z = 0.69) (**B**) Selected compounds reducing light output of luciferase above 70%.

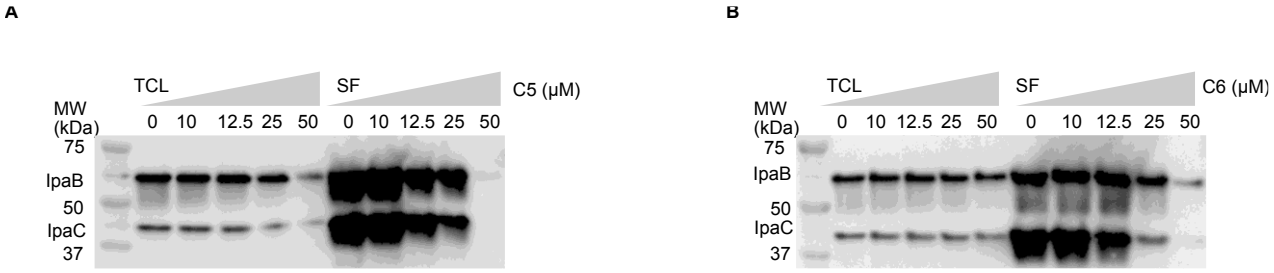


Figure 4.S. 2. Dose-dependent repression of IpaB and IpaC production and secretion in Δ*ipaD* by C5 and C6 respectively. (**A**) Dose response effect of C5 on TCL and SF of constitutive secreted strain Δ*ipaD*. (**B**) Dose response effect of C6 in the SF of Δ*ipaD* strain.

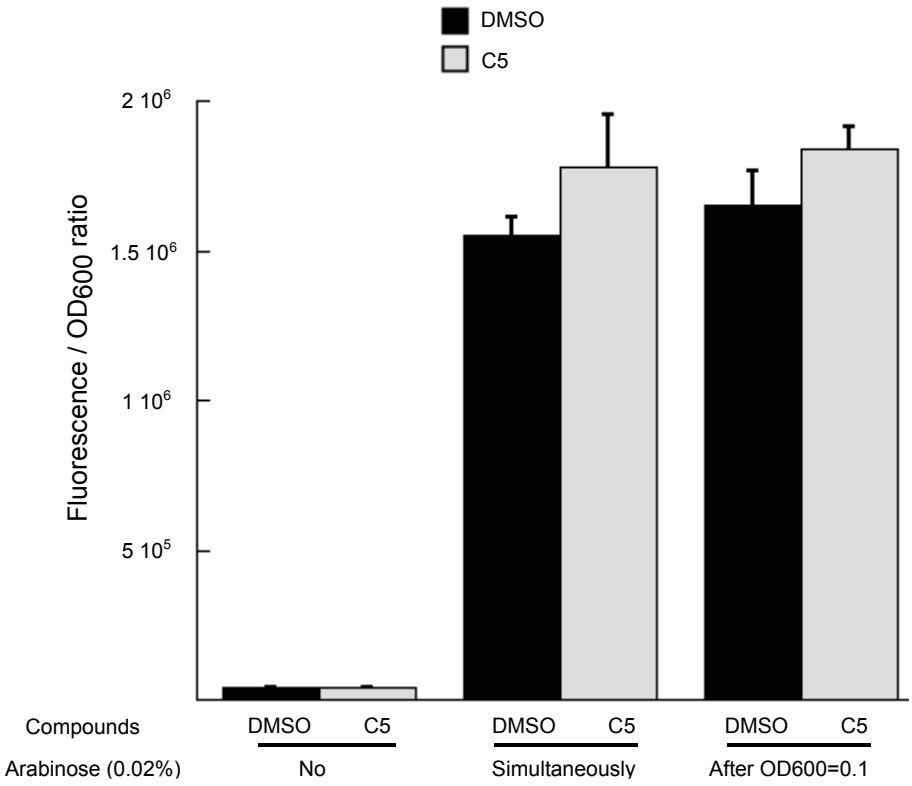


Figure 4.S. 3. C5 has no effect on the AraC promoter which induces GFP production in the presence of arabinose.

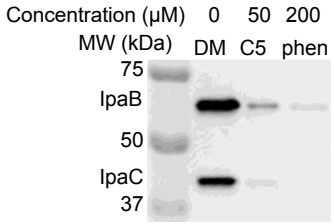


Figure 4.S. 4. Phenanthroline at 200 µM reduces IpaB and IpaC productions.

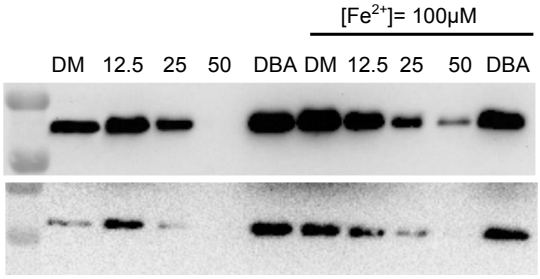


Figure 4.S. 5. Partial complementation by iron(II) (100 µM) of IpaBC productions in the presence of gradient concentration of C5.

Table 4.S. 1. The oligonucleotides used in this work

Labels	Sequence (5'-3')	Description
HMIO254 HMIO255	GTTTCTGGTCTCCCGCTAGCGAATTGCCCTCGTATTAAATGTGTATCG GAATTTGGTACCGAGTCAGTTCTAGATTATGCGGCCAGCGCATAATTTTC	pTSAR3.4t plasmid
HMIO257 HMIO258	CCATGGAGATATCCCAATTATTGAAGGCCTCCCTAACGGGGGGCCTTTTTTTGTTTCTGGTCTCCCGCTAGC GCTAGCGGGAGACCAGAAACAAAAAAAGGCCCCCCGTTAGGGAGGCCTTCAATAATTGGGATATCTCCATGG	L3S3P21 terminator
HMIO658 HMIO659	ACTAAAAAAATTTCATTCATTATTAACGG AATACGAGGGCAATTCAAAAAAAAG	Amplification of plasmid pTSAR-l
HMIO660 HMIO661	TTTTGAATTGCCCTCGTATTGCAGTTATTGCATTGGAGAG AATGAATGAAATTTTTTTAGTCTTTACCGTAGGTAATTC TCCG	Amplification of *icsA*p from M90T
HMIO662 HMIO663	TTTTGAATTGCCCTCGTATTTAATTGATAAGCATTTTTT CATCTATGG AATGAATGAAATTTTTTTAGTTTTATTACCAGAATGGAG TGTAAAC	Amplification of *virB*p from M90T
HMIO664 HMIO665	TTCTAGGAGGTGCTAGCGCTATATAGTTTGGTATATTCT GTTGAATTTATG AGTTTCTGTTCCTCGAGCCCAAATTTTTTATGATATAAG TAAAATTTCTTTGG	Amplification of plasmid *virF*
HMIO593 HMIO594	GGGCTCGAGGAACAGAAAC AGCGCTAGCACCTCCTAG	Amplification of plasmid pSU2.1
HMIO704 HMIO705	TCCGGATCTAGACCTGCAG TCCACACAACATACGAGC	Amplification of plasmid pUC18Δ-3xFlag
HMIO706 HMIO707	CGGCTCGTATGTTGTGTGGAAAATCAATATGAATAAACA GGGTG CCTGCAGGTCTAGATCCGGATGAAGACGATAGATGGC	Amplification of *virB* from M90T
HMIO730 HMIO731	TGGGCATCGGCCTTAATTTTCTGAAAGTATCAGG ACTCACGGATGCTATGCTCTTTTG	Mutagenesis to disrupt ryhB binding sites
qPCR87 qPCR88	ATCTACGGACCGGAATCTT GATATCGACGCCCAGTTTAC	ddPCR of *recA*
ddO31 ddO32	CACCGGCCTTGAAACTACAT GAACGGCTTCATTTGGTGAT	ddPCR of *mxiD*
ddO33 ddO34	GAAAGAAGCCGAGCTTGATG GTTACCTCCGGGAGAGATCC	ddPCR of *ipaD*
ddO39 ddO40	TTGATTGCGTTCGATGCTAC AGATTGAGCAGTGGCAAGGT	ddPCR of *ipaC*
ddO43 ddO44	TCCACCCATTTAAACCAAGC TGGCCGAATTTTTCTATTGC	ddPCR of *ipaB*
ddO47 ddO48	GTAAAACTGACCGGCTGGAA TATCTCCAGAGCCTCCCTCA	ddPCR of *icsA*

ddO51 ddO52	TCGAGAAGAGATGGCGTCTT CATGATGCATGCGAATATCA	ddPCR of *virF*
ddO53 ddO54	AGTCACTCGTGCCTTTCAGG GTTCTGACGCGATTGGAAAT	ddPCR of *virB*
ddO57 ddO58	TCGCGTACTGAACCAGTTTG TCAGGTGATCGTGGTGATGT	ddPCR of *fur*
ddO61 ddO62	CACCCGGCTGGCTAAGTAA GGAGAACCTGAAAGCACGAC	ddPCR of *ryhB*
ddO77 ddO78	TAGCCTTGCTGCCGTTAAAT TTGGCCAGTCCAGGTTTTAG	ddPCR of *hns*
ddO87 ddO88	GCTTCAAACCGCAACTTTTC GGCTTTGGGGATTAACATCA	ddPCR of *RnaG*
ddO91 ddO92	CAACTCTGCGAGCGTTTTTC AATTACCCAATCCCCACACA	ddPCR of *spa33*
ddO93 ddO94	CTTATTTGGGGGAGGCTGTT CGCCTTCTCGACTTTGTCTC	ddPCR of *spa40*
ddO101 ddO102	GCATATGCAGCGTCTGTTTG CAGCATGAGCCAGTTCGTTA	ddPCR of *ftnA*
ddO131 ddO132	AACAAAAACCCTGGCAACAG GCCAACCGTGAGAAACTGAT	ddPCR of *cirA*
ddO133 ddO134	CTACGGCAAGCATCATCAGA TGTTGAATACGCCACCTTCA	ddPCR of *sodB*
ddO137 ddO138	ATGAGGCTGCTTTTGCTGAT TCTTCCTTCAGTCCGGAGAA	ddPCR of *iucD*