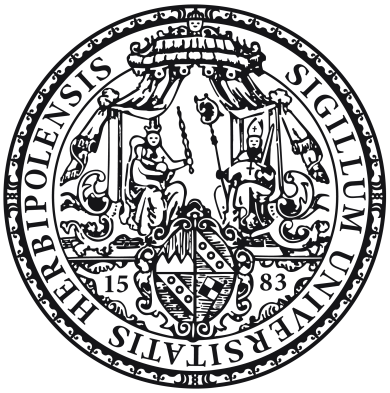
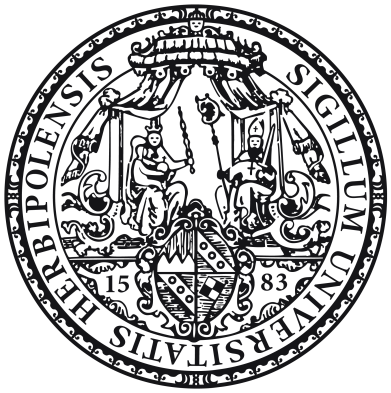



# Small proteins in *Salmonella*: an updated annotation and a global analysis to find new regulators of virulence

**Elisa Venturini**

Würzburg 2021




# Small proteins in *Salmonella*: an updated annotation and a global analysis to find new regulators of virulence

Kleine Proteine in *Salmonella*: Eine aktualisierte Annotation und eine globale Analyse, um neue Regulatoren der Virulenz zu finden

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Univsersität Würzburg,
Section Infection and Immunity

submitted by

**Elisa Venturini**

from Udine

Würzburg 2021

**Submitted on:**

## Members of the *Promotionskomitee*:

**Chairperson: Prof. Dr. Thomas Dandekar**

**Primary Supervisor: Prof. Dr. Jörg Vogel**

**Supervisor (Second): Prof. Dr. Kai Papenfort**

**Supervisor (Third): Prof. Dr. Cynthia Sharma**

**Supervisor (Fourth): Prof. Dr. Wilma Ziebuhr**

**Date of Public Defense:**

**Date of Receipt of Certificates:**

## Affidavit

I hereby confirm that my thesis entitled "Small proteins in *Salmonella*: an updated annotation and a global analysis to find new regulators of virulence", is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date Signature

## Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation "Kleine Proteine in *Salmonella*: Eine aktualisierte Annotation und eine globale Analyse, um neue Regulatoren der Virulenz zu finden", eigenständig, d.h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Ort, Datum Unterschrift

# Summary

Small proteins, often defined as shorter than 50 amino acids, have been implicated in fundamental cellular processes. Despite this, they have been largely understudied throughout all domains of life, since their size often makes their identification and characterization challenging.

This work addressed the knowledge gap surrounding small proteins with a focus on the model bacterial pathogen *Salmonella* Typhimurium. In a first step, new small proteins were identified with a combination of computational and experimental approaches. Infection-relevant datasets were then investigated with the updated *Salmonella* annotation to prioritize promising candidates involved in virulence.

To implement the annotation of new small proteins, predictions from the algorithm sPepFinder were merged with those derived from Ribo-seq. These were added to the *Salmonella* annotation and used to (re)analyse different datasets. Information regarding expression during infection (dual RNA-seq) and requirement for virulence (TraDIS) was collected for each given coding sequence. In parallel, Grad-seq data were mined to identify small proteins engaged in intermolecular interactions.

The combination of dual RNA-seq and TraDIS lead to the identification of small proteins with features of virulence factors, namely high intracellular induction and a virulence phenotype upon transposon insertion. As a proof of principle of the power of this approach in highlighting high confidence candidates, two small proteins were characterized in the context of *Salmonella* infection.

MgrB, a known regulator of the PhoPQ two-component system, was shown to be essential for the infection of epithelial cells and macrophages, possibly via its stabilizing effect on flagella or by interacting with other sensor kinases of two-component systems. YjiS, so far uncharacterized in *Salmonella*, had an opposite role in infection, with its deletion rendering *Salmonella* hypervirulent. The mech-

anism underlying this, though still obscure, likely relies on the interaction with inner-membrane proteins.

Overall, this work provides a global description of *Salmonella* small proteins in the context of infection with a combinatorial approach that expedites the identification of interesting candidates. Different high-throughput datasets available for a broad range of organisms can be analysed in a similar manner with a focus on small proteins. This will lead to the identification of key factors in the regulation of various processes, thus for example providing targets for the treatment of bacterial infections or, in the case of commensal bacteria, for the modulation of the microbiota composition.

# Zusammenfassung

Kleine Proteine, oft definiert als kürzer als 50 Aminosäuren, sind in fundamentale zelluläre Prozesse involviert. Trotzdem sind sie in allen Domänen des Lebens noch weitgehend unerforscht, da ihre Größe ihre Identifizierung und Charakterisierung oft schwierig macht.

Diese Arbeit adressiert die Wissenslücke um kleine Proteine mit einem Fokus auf das bakterielle Modellpathogen *Salmonella* Typhimurium. In einem ersten Schritt wurden neue kleine Proteine mit einer Kombination aus bioinformatischen und experimentellen Ansätzen identifiziert. Anschließend wurden infektionsrelevante Datensätze mit der aktualisierten *Salmonella*-Annotation untersucht, um vielversprechende Kandidaten zu priorisieren, die an der Virulenz beteiligt sind.

Um die Annotation neuer kleiner Proteine zu implementieren, wurden die Vorhersagen aus dem Algorithmus sPepFinder mit denen aus Ribo-seq kombiniert. Diese wurden der *Salmonella*-Annotation hinzugefügt und zur (Re-)Analyse verschiedener Datensätze verwendet. Für jede gegebene kodierende Sequenz wurden Informationen zur Expression während der Infektion (duale RNA-seq) und zum Beitrag zur Virulenz (TraDIS) gesammelt. Parallel dazu wurden Grad-seq-Daten ausgewertet, um kleine Proteine zu identifizieren, die an intermolekularen Interaktionen beteiligt sind.

Die Kombination von dualer RNA-seq und TraDIS führte zur Identifizierung von kleinen Proteinen mit Merkmalen von Virulenzfaktoren, nämlich einer hohen intrazellulären Induktion und einem Virulenz-Phänotyp nach Transposon-Insertion. Als Beweis für die Leistungsfähigkeit dieses Ansatzes Identifikation von vielversprechenden Kandidaten wurden zwei kleine Proteine im Kontext einer *Salmonella*-Infektion charakterisiert.

MgrB, ein bekannter Regulator des PhoPQ-Zweikomponentensystems, erwies sich als ein für die Infektion von Epithelzellen und Makrophagen essentielles Protein, möglicherweise über seine stabilisierende Wirkung von Flagellen oder durch

Interaktion mit Sensorkinasen von Zweikomponentensystemen. YjiS, das in *Salmonella* bisher nicht charakterisiert wurde, hatte eine entgegengesetzte Rolle bei der Infektion, wobei seine Deletion *Salmonella* hypervirulent macht. Der Mechanismus, der dem zugrunde liegt, ist zwar noch unklar, beruht aber wahrscheinlich auf der Interaktion mit inneneren Membranproteinen.

Insgesamt liefert diese Arbeit eine globale Beschreibung der kleinen *Salmonella*-Proteine im Kontext der Infektion mit einem kombinatorischen Ansatz, der die Identifizierung interessanter Kandidaten beschleunigt. Verschiedene Hochdurchsatz-Datensätze, die für ein breites Spektrum von Organismen verfügbar sind, können auf ähnliche Weise mit einem Fokus auf kleine Proteine analysiert werden. Dies wird zur Identifizierung von Schlüsselfaktoren in der Regulation verschiedener Prozesse führen und damit z. B. Targets für die Behandlung bakterieller Infektionen oder, im Falle kommensaler Bakterien, für die Modulation der Mikrobiota-Zusammensetzung liefern.

# Contents

# List of Figures

# List of Tables

# 1 Introduction

## 1.1 A brief history of small protein annotation

*Escherichia coli* will soon mark its hundredth anniversary as the model organism of microbiology par excellence (Blount, 2015). This might suggest that its genome is now fairly well annotated, and that its pathways are described almost to completion. However, reality does not fully reflect this, as exemplified by the case of small proteins. Often defined as the unprocessed product of coding sequences (CDSs) shorter than 150 nucleotides (i.e. $<$ 50 aa), small proteins have recently begun to be appreciated as a largely unannotated and poorly characterized subset of the proteome (Storz et al., 2014) whose impact is quite large. A timeline of important advances in genome annotation and investigation of small proteins is shown in Fig. 1.1.

The genome assembly of *E. coli* K-12 strain MG1655, the workhorse of bacteriology, dates back to 1997 ((Blattner et al., 1997); Fig. 1.1). At the time, annotation of open reading frames (ORFs) relied on homology to known genes, presence of Shine-Dalgarno (SD) motifs, patterns of codon usage resembling those of known protein domains, or cross-referencing with databases featuring N-terminus protein sequences. Each of these approaches had, and often still has, limitations for the annotation of small proteins. This resulted in an under-representation, in an early work, of proteins shorter than 100 aa: 381 out of 4'288 (Blattner et al., 1997).

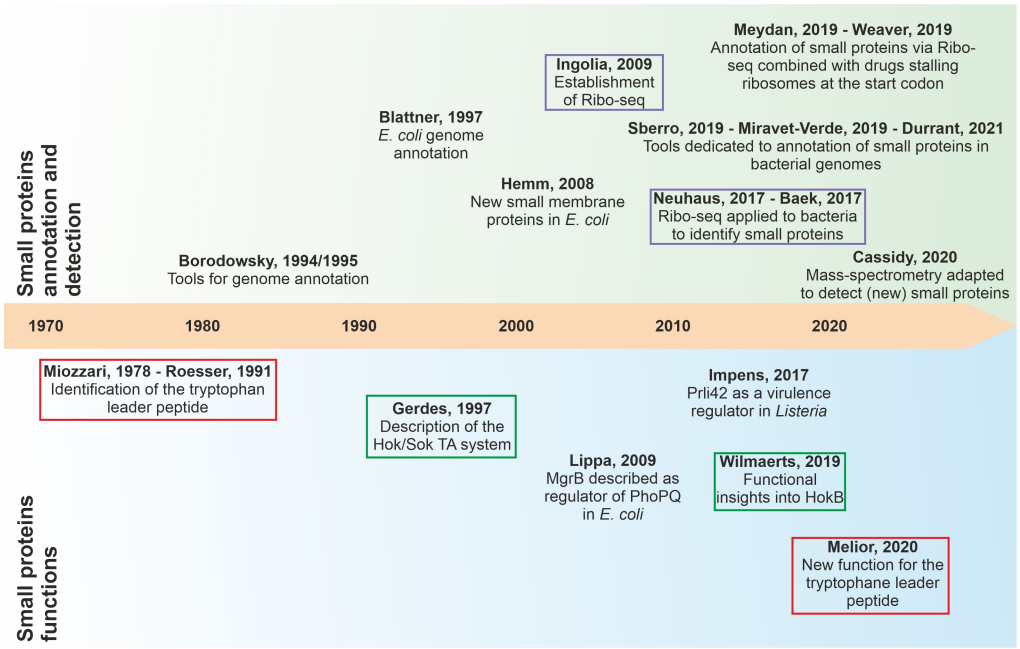



**Figure 1.1: Timeline of the evolution of the field of small proteins.** Overview of hallmarks for small proteins annotation and characterization, mostly in bacteria. The references are mentioned throughout the main text. Works on a similar topic are enclosed in rectangles of the same color.

In bacterial genomes, considering all in-frame start and stop codons as potential CDSs would result in an enormous amount of proteins, and a high rate of false-positives (Poptsova and Gogarten, 2010). For this reason, in early annotations, the largest possible ORF was arbitrarily chosen in cases with multiple in-frame start codons or overlapping genes (Blattner et al., 1997; Poptsova and Gogarten, 2010). Only recent evidence showed that bacterial genomes might be more densely packed than previously thought, opening doors to the annotation of overlapping (short) CDSs (Adams et al., 2021).

At the same time, early programs designed for the annotation of new CDSs searching for sequence conservation, such as BLAST or training models to call new CDSs based on codon usage patterns or known domains (Borodovsky et al., 1994, 1995), were not designed with the knowledge that the most highly conserved CDSs are usually long (Brocchieri and Karlin, 2005; Wolf et al., 2009). This resulted in a failure to annotate shorter, more variable genes. Indeed, sORFs are often recently diverged genes that hold the potential to be rapidly evolved and provide adaptive

potential to different environments (Carvunis et al., 2012; Ekman and Elofsson, 2010). The lack of recognition of this

Overall, whether early genome annotations were based on sequence similarity to known genes or predictions based on statistical measurements of, for example, codon usage, small proteins have always been challenging to confidently map. Considering all these limitations that lead to an underrepresentation of small proteins identification in bacteria, it is unsurprising that recent works have successfully annotated a considerable number of new short CDSs in *E. coli* (Hemm et al., 2008; Weaver et al., 2019), as well as in less studied bacteria including microbiota species (Miravet-Verde et al., 2019; Sberro et al., 2019).

## 1.2 New approaches to annotate small proteins

After recognizing the gap in small protein annotations, in the past decade different approaches have been developed to specifically identify small ORFs (sORFs) in bacterial genomes. These are divided between those based on either *in silico* predictions or experimental data. The first class includes refined algorithms aimed at searching features such as ribosome binding sites (RBSs) and/or known protein motifs (Hemm et al., 2008; Miravet-Verde et al., 2019), as well as comparative genomics algorithms (Samayoa et al., 2011; Sberro et al., 2019), with a specific focus on small proteins (Fig. 1.1). Given that new small proteins are still being discovered, the progressive addition of these to training sets will further improve the discovery rate of these tools.

The most important technique for experimental data-driven annotation probably is ribosome profiling by sequencing (Ribo-seq, (Ingolia et al., 2009); Fig. 1.1). Ribo-seq relies on the protection of mRNAs from nuclease digestion by translating ribosomes: sequencing these RNA footprints indicates which genes are being translated at a given time and in a given condition. Analysis of the resulting data can be carried out not only to estimate the translation rate of annotated mRNAs, but can also show which unannotated portions of the genome are being translated, indicating new candidate CDSs. This approach was successful in the annotation of new short genes in bacteria, alone (Neuhaus et al., 2017; Baek et al., 2017) or

in combination with *in silico* predictions (Miravet-Verde et al., 2019).

One limitation of Ribo-seq in bacteria is that ribosome footprints are not as "sharp" as the eukaryotic ones, therefore lacking a clear codon periodicity (Mohammad et al., 2019). This makes prediction of sORFs in prokaryotes more challenging, especially when multiple start codons occur close to each other. This has been addressed by the use of drugs that stall ribosomes at the initiation site, making it easier to confidently identify start codons, as footprints will accumulate at the SD sequence ((Meydan et al., 2019; Weaver et al., 2019); Fig. 1.1). This step helped not only in the annotation of independent sORFs, but also in the annotation of internal sORFs overlapping out of frame with larger CDSs, a novel class of small bacterial genes that is being recognized as more widespread than previously believed (Orr et al., 2020). Annotation of internal sORFs is indeed unachievable with classic Ribo-seq data, in which read coverage from the two overlapping CDSs would be indistinguishable. However, despite some limitations, Ribo-seq is so far among the best tools for the annotation of new sORFs.

Unlike Ribo-seq, mass spectrometry can provide direct evidence of the existence of a protein. In the context of proteome analysis, it has been used to detect protein levels and thus test how different factors affect protein abundance on a global scale. Furthermore, it can be used to detect new proteins if analysed with a reference proteome generated with all six possible reading frames. Classic mass spectrometry approaches, relying for example on digestion of the proteins with trypsin, run into intrinsic limitations of small proteins that i) give rise to few peptides, making them more prone to being discarded as false-positives, ii) are more likely to have their signal overruled by the more abundant peptides generated by larger proteins. Bottom-up approaches, in which the sample is not fragmented, offer an effective solution to this problem (Cassidy et al., 2020), even more so when coupled to a prior enrichment of small proteins. Similarly, mass spectrometry can be coupled to fractionation methods for the enrichment of peptides that contain methionines, representing the N-terminus of a protein (Gevaert et al., 2003), a protocol already established in bacteria. This can lead not only to the identification of new small proteins (Impens et al., 2017), but also adds another level of confidence in precisely mapping the boundaries of a CDS unambiguously placing a start codon.

It goes without saying that the highest confidence in annotating new genes derives from the combination of complementary techniques. For example, mass spectrometry or Ribo-seq data can be analysed with an annotation generated from *in silico* predictions. Further confirmation of expression of candidate loci can be derived from increasing number of RNA-seq datasets, now available for several bacterial species and often from several experimental conditions.

A list of strengths and weaknesses of each approach discussed so far can be found in Table 1.1.

**Table 1.1:** Advantages and limitations of different *in silico* and experimental techniques that can be used for prediction of small protein candidates.

| Method | Advantages | Limitations |
|---|---|---|
| *in silico* predictions | • No restriction to growth condition<br>• Applicable to any available genome | • Lacks direct experimental validation<br>• Relies on known small protein features, missing unknown ones |
| Ribo-seq | • Not limited by protein size<br>• Annotation-independent | • Lowly abundant proteins are easily overlooked<br>• Only few growth conditions can be tested at a time |
| Mass spectrometry | • Provides direct proof of protein existence<br>• Provides indication of protein abundance | • The classic fragmentation step limits its detection of small proteins<br>• Bottom-up approaches are still under development |

The annotation process, as discussed so far, is prone to false positives and thus requires a further layer of validation in which a subset of candidate genes is chosen for independent detection. Most often, this is carried out via epitope tagging and western blotting. Successful detection of the candidates will indicate the rate of true positives of the predictions. It should however be considered that the presence of a large tag might interfere with the half-life of a small protein, by blocking stabilizing interactions with its partners. Nevertheless, the use of a large epitope is in some cases necessary, as the protein candidate requires a significant increase in its size for it not to diffuse out of a gel during classical electrophoresis, or run through a membrane during standard western blotting. Hence, these techniques require adjustments when applied to small proteins.

## 1.3 The major functional classes of small proteins

Small proteins are often too small to contain a catalytic domain and carry out an independent enzymatic function. Despite this, they are proving themselves as important and versatile players in bacterial cellular processes. The following paragraphs paint with broad strokes an overview of the functions most often executed by small proteins in bacteria. The works cited below frequently include examples of elegant and creative ways in which classical biochemical approaches have been adapted to overcome limitations in handling small proteins.

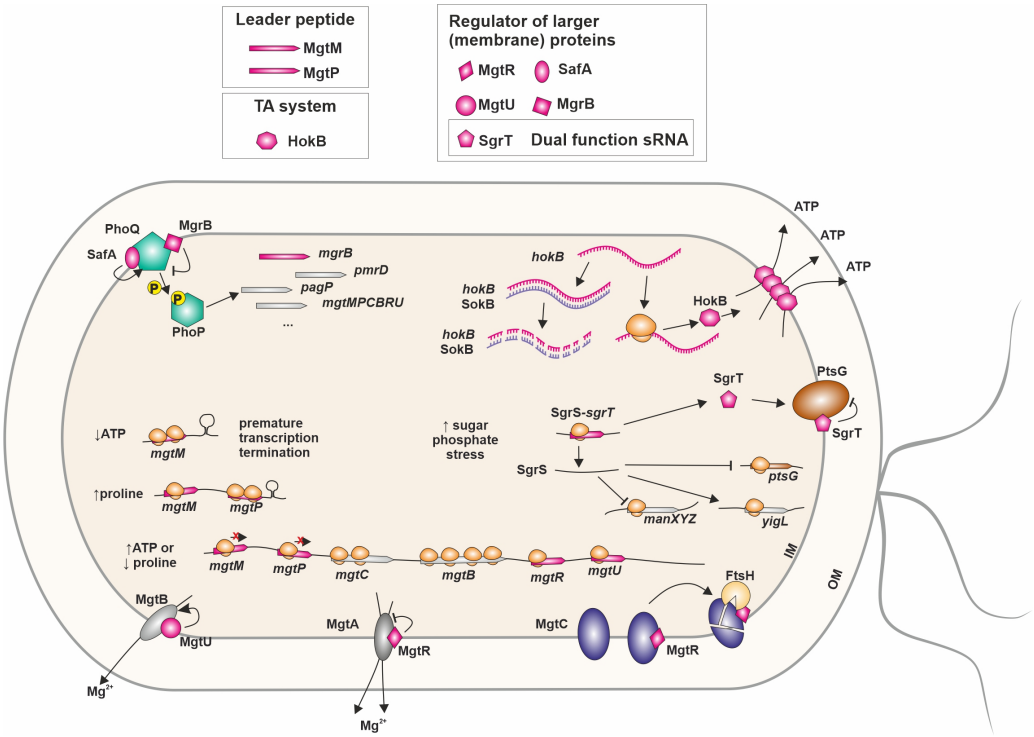



**Figure 1.2: Overview of small proteins functions.** Small proteins play role in various cellular processes, some of which are depicted here. A detailed description of these examples can be found in the main text of Section 1.3.

### 1.3.1 Leader peptides

One of the first described functional classes of small proteins, along with small ribosomal proteins (Wada, 1986), are leader peptides (Das et al., 1983). Despite the name "peptide", which suggests post-translational processing, leader peptides are the uncleaved product of sORFs encoded upstream of larger genes. Their translation results in conformational changes in the downstream mRNA that affect transcription or translation of the following gene. The translation of the leader peptide is usually dependent on the abundance of a compound, or metabolite, that the operon is involved in producing (Hemm et al., 2020).

A classic example of leader peptides is the tryptophan operon leader **TrpL** (14 aa), conserved in different bacteria (Das et al., 1983). Its CDS contains multiple tryptophan codons and low level of intracellular tryptophan results in ribosome stalling, allowing the folding of the downstream mRNA in an antiterminator stem-

loop. When tryptophan is not limiting, translation proceeds uninterrupted. The first antiterminator cannot form, but instead there is folding of a terminator further downstream. This halts transcription of the operon. This mechanism ensures that the enzymes for the production of tryptophan, encoded in the operon, are only expressed when necessary (Miozzari and Yanofsky, 1978; Roesser and Yanofsky, 1991).

Since their discovery, leader peptides were not believed to carry out any function other than being byproducts of the regulation of downstream genes. Surprisingly, a recent work described the tryptophan leader peptide in *Sinorhizobium meliloti* as a post-transcriptional regulator acting *in trans* (Melior et al., 2019). **peTrpL** (14 aa) was shown to bind an antisense RNA that inhibits expression of an efflux pump involved in antibiotic resistance. This binding results in degradation of the non-coding RNA (ncRNA) and thus expression of the efflux pump. Overexpression of peTrpL alone resulted in increased resistance of *Sinorhizobium meliloti* to tetracyclin (Melior et al., 2020). This work opens the doors to the possibility that leader peptides might have independent roles and are not merely a product of the regulation.

### 1.3.2 Toxin-antitoxin systems

Small proteins are frequent members of toxin-antitoxin (TA) systems, as toxin (counteracted by an sRNA) in type I and III TA systems, or as both toxin and antitoxin in the type II, IV, V, and VI TA systems (Gerdes and Maisonneuve, 2012). These systems are involved in the maintenance of plasmids via a post-segregational killing mechanism and, when localized on the genome, can regulate dormancy and awakening of a bacterial pathogen, processes that determine onset of persisters often triggered by exposure to antibiotics (Fisher et al., 2017). Persisters, being non-growing albeit metabolically active cells (Stapels et al., 2018), are the causative agents of recurrent infections post antibiotic treatment. The plasmid-encoded Hok-Sox system was one of the first described as a prototype for type I TA systems (Gerdes et al., 1997, 1985) (Fig. 1.2), exemplified by *E. coli* **HokB** (49 aa). HokB localizes to the inner membrane (Wilmaerts et al., 2018) where oxidoreductase DsbC can form or reduce disulphide bridges between oligomers of

HokB, leading to pore formation or disassembly. HokB can then be degraded by the protease DegQ (Wilmaerts et al., 2019). This mechanism regulates ATP efflux from the cytosol and controls the onset of persisters through loss of membrane potential, or awakening through membrane re-polarization and production of ATP (Wilmaerts et al., 2018).

The *E. coli* **TisB** toxin (29 aa), whose action is antagonized by the IstR-1 sRNA (Vogel et al., 2004; Darfeuille et al., 2007), is induced upon activation of the SOS-response and, by forming pores into the inner membrane, causes cell depolarization and loss of ATP, leading to cell death (Unoson and Wagner, 2008). A recent study suggested that the TisB might itself be involved in the activation of the SOS-response, and that the abundance of TisB and radical oxygen species might determine the successful generation of persisters upon antibiotic treatment, as opposed to cell death (Edelmann and Berghoff, 2019). TisB and HokB depict the most frequent mechanism of action of toxins, where small hydrophobic proteins localize to the membrane.

### 1.3.3 Regulators of larger proteins

Two-component systems (TCSs) are widespread mechanisms that feature a histidine kinase localized in the membrane and a response regulator. The histidine kinase is responsible for sensing a specific signal that leads to its autophosphorylation and subsequent phosphorylation of the corresponding transcription regulator. The regulator is then in its active state, which leads to transcription of its target genes that are usually required for adaptation to the environment (Laub and Goulian, 2007). Due to their size, small proteins are ideal regulators that bacteria can quickly evolve for fine-tuning the activity of larger proteins such as sensor kinases or transcriptional regulators (Orr et al., 2020).

The PhoPQ TCS, in which PhoQ is the histidine kinase and PhoP the transcription regulator, is of particular importance for *Salmonella* enterica serovar Typhimurium (from here on, *Salmonella*), as it regulates a plethora of genes important for intracellular survival. In particular, it is essential for responding to low magnesium (Garcia Vescovi, 1996), low pH (Aranda et al., 1992), and antimicrobial peptides (Miller et al., 1990). Among its targets there is *ssrB*, encoding

a master regulator of virulence genes (Bijlsma and Groisman, 2005). Considering this, it is not surprising that *Salmonella* has evolved several layers of regulation to fine-tune the timing and level of expression of PhoPQ, including regulatory small proteins.

**MgrB** (47 aa) was first described in *E. coli*, where the PhoPQ system is conserved with that of *Salmonella* (Salazar et al., 2016), and was shown to repress PhoQ auto-phosphorylation by direct protein-protein interaction in the membrane ((Lippa and Goulian, 2009); Fig. 1.2). The small protein is part of a negative feedback loop, as *mgrB* is itself positively regulated by the active form of PhoP (Kato et al., 1999; Lippa and Goulian, 2009). Interestingly, *E. coli* has a second small protein that regulates PhoPQ. **SafA** (65 aa), absent from *Salmonella*, positively regulates PhoQ (Ishii et al., 2013) independently from the regulation by MgrB (Yoshitani et al., 2019). Given that SafA is induced by the EvgS/EvgA TCS, it represents a link between two TCSs (Eguchi et al., 2012), exemplifying how small proteins might be more widespread and necessary regulators than currently known.

### 1.3.4 Dual function sRNAs

Although not homogeneous in terms of function, transcripts which act both as interfering RNAs as well as code for a small protein are usually put in the same category defined as "dual function sRNAs". In bacteria, only a handful of such genes have been described (Gimpel and Brantl, 2017). RNAIII in *Staphylococcus aureus* was the first ncRNA which was shown to encode the small protein **δ-hemolysin** (26 aa; (Janzon et al., 1989)). Both molecules act to promote infection. RNAIII regulates the transition from early to late infection, inhibiting the genes involved in the first phase while promoting those for the second one (Novick et al., 1993; Huntzinger et al., 2005; Gupta et al., 2015). The small protein δ-hemolysin, on the other hand, is secreted and by inserting in the host cell membrane leads to lysis (Verdon et al., 2009).

An extensively studied, and so far the only characterized, dual function sRNA from gamma-proteobacteria is SgrS. In *E. coli*, SgrS is involved in the response to sugar phosphate stress by interfering with the translation of sugar transporters (PtsG, ManX, and ManY) while promoting the expression of YigL, a sugar phos-

phatase (Vanderpool and Gottesman, 2004; Rice et al., 2012; Papenfort et al., 2013). Upon further inspection, a CDS within the sRNA was identified: **SgrT** (43 aa) is inserted in the inner membrane and inhibits the activity of PtsG (Vanderpool and Gottesman, 2004; Lloyd et al., 2017), one of the sugar transporters whose translation is repressed by SgrS (Rice et al., 2012). Overall, this elegant system allows *E. coli* to quickly respond to elevated intracellular concentrations of sugar phosphate by blocking both the import of sugar and the production of new transporters.

### 1.3.5 Small proteins involved in infection

Direct involvement of small proteins in regulating infection has been described in only a handful of cases. One such example is **Prli42** (34 aa) in *Listera monocytogenes*. Previously unannotated, it was identified via a mass spectrometry adaptation that includes a step to enrich N-terminal peptides (Impens et al., 2017). Its functional characterization showed it is essential for activation of the stressosome (Impens et al., 2017; Williams et al., 2019), on which *Listeria* depends for survival inside the host cell (Guerreiro et al., 2020). Although the exact mechanism by which Prli42 regulates the activation of the stressosome is unknown, evidence suggests that the small protein is required at the early stages of stress sensing and response (Williams et al., 2019).

In *Salmonella*, the small protein **MgtR** (30 aa) regulates the virulence factor MgtC by binding it and inducing its protease-mediated degradation (Alix and Blanc-Potard, 2008). Unsurprisingly, overexpression of MgtR decreased *Salmonella* fitness during infection (Alix and Blanc-Potard, 2008). Interestingly, *mgtC* and *mgtR* are encoded on the same operon and separated by the gene coding for the magnesium transporter MgtB, which is unaffected by MgtR (Alix and Blanc-Potard, 2008). Another magnesium transporter, MgtA, is instead repressed *in trans* by the small protein MgtR (Choi et al., 2012). Three other small proteins have been added to the *mgtCBR* operon. **MgtU** (28 aa), located at the 3' end of the operon, was recently annotated in *Salmonella* (Baek et al., 2017) and shown to protect the MgtB magnesium transporter from protease degradation at late stages of *Salmonella* infection (Yeom et al., 2020). Mutants lacking *mgtU* cannot

withstand magnesium starvation during macrophage infection (Yeom et al., 2020). Lastly, two small proteins located at the 5' end regulate the transcription of the entire operon in response to high levels of ATP (**MgtM**, 90 aa) or low levels of proline (**MgtP**, 17 aa) with an attenuation mechanism typical of leader peptides (Park et al., 2010; Lee and Groisman, 2012a,b). This operon is one of the most small-protein-dense that has been annotated to date (Fig. 1.2).

## 1.4 Global datasets to describe bacterial genes and proteins: The case of *Salmonella*

High-throughput data that describe different aspects of *Salmonella* growth in a broad spectrum of conditions are available. Their analysis with a specific focus on small proteins can indicate which ones have an interesting behaviour. When their induction correlates with, for example, infection conditions, this can suggest that the protein of interest is involved in pathogenesis. Three of these *Salmonella*-related datasets are described below.

### 1.4.1 Dual RNA-seq

With dual RNA-seq, the transcriptomes of both host and pathogen can be mapped throughout an infection time course (Westermann et al., 2012; Westermann and Vogel, 2021). Briefly, fluorescently-labelled bacteria allow for the enrichment of infected host cells via fluorescence-activated cell sorting (FACS) at different time points during infection. The cells are then subjected to total RNA-seq and each transcript is mapped to either the pathogen or the host. The abundance of each bacterial transcript at the desired time point is compared to its level in the inoculum, showing how gene expression changes in the course of an infection.

Dual RNA-seq of *Salmonella* infecting epithelial cells showed that genes required for survival inside the host are strongly induced intracellularly (Westermann et al., 2016), validating previous work that focused on a limited number of genes (Valdivia and Falkow, 1997). This approach helped to identify the previously uncharacterized sRNA PinT as a regulator of virulence gene expression during

infection (Westermann et al., 2016; Santos et al., 2021). Considering this, it is reasonable to interpret a strong intracellular induction as a proxy for involvement in infection. Therefore, this type of data can be mined to rank uncharacterized genes based on their expression pattern, and at the same time support annotations of ORFs specifically induced in infection conditions that had been previously overlooked.

### 1.4.2 TraDIS

The use of transposons, coupled with DNA sequencing, has made possible the global assessment of gene essentiality (Cain et al., 2020). Different types of transposons have been used to generate libraries of mutants with random transposon insertions throughout the genome. Knowing the sequence of the transposon allows to map the site of insertion in each clone. The first type of analysis that can be carried out with this type of data is to identify genomic regions in which transposon insertion did not occur, suggesting them to be essential loci (Langridge et al., 2009). The same library can be tested for growth in the condition of interest: comparing the abundance of each mutant before and after growth indicates which genes impact, positively or negatively, bacterial fitness (Barquist et al., 2013).

One variant of the transposon insertion sequencing is TraDIS, an acronym for transposon-directed insertion sequencing (Langridge et al., 2009). This approach has been successfully applied to *Salmonella* and, notably, to the characterization of short genes such as sRNAs (Barquist et al., 2013). Given that one of the parallels between sRNAs and sORFs is the challenge in targeting short genes, which are less likely than larger genes to have a transposon insertion, this approach can help in ranking short CDSs in terms of requirement for cell survival in a given condition.

### 1.4.3 Grad-seq

Analysis of soluble cytoplasmic complexes by density-gradient separation has a long-standing history as the key experiment in understanding major biological events such as DNA replication or uncovering the existence of RNA (Meselson and Stahl, 1958; Brenner et al., 1961). Coupling gradient separation with high-

throughput techniques such as RNA-seq is at the center of Ribo-seq (Ingolia et al., 2009), as ribosome-bound footprints are isolated from the sucrose gradient fractions where assembled ribosomes localize after ultracentrifugation. This approach was taken a step further, coupling the separation of bacterial cell extracts on a linear glycerol gradient which, upon fractionation, was analysed by both RNA-seq and mass spectrometry (Grad-seq, (Smirnov et al., 2016)). In this way, a sedimentation pattern can be described for RNAs and proteins. Downstream analysis indicates which molecules possibly interact with each other, having a similar in-gradient behaviour. This approach has lead to the identification of ProQ as a novel global RNA-binding protein in *Salmonella* (Smirnov et al., 2016), later characterized as a regulator of infection-relevant genes (Westermann et al., 2019).

The sedimentation pattern of a protein on a glycerol gradient is largely governed by its molecular weight (Erickson, 2009). In other words, a protein with a high molecular weight (or a complex of multiple proteins) will localize deep within the gradient, while smaller proteins will localize at the top of the gradient. Therefore, detection of a small molecule beyond the first fractions is a strong indication that it is interacting with other molecules. This reasoning is behind the identification of a new toxic small protein (RyeG, a.k.a. YodE) and a new ribosome-associated small protein (YggL) in *E. coli* (Hör et al., 2020a). At the same time, analysis of mass spectrometry data derived from Grad-seq can validate the existence of a predicted small protein, although for this the limitations already listed still hold true.

### 1.4.4 Data integration

The datasets described so far have great potential in the characterization of new small proteins, giving information on gene expression and essentiality, as well as protein existence and engagement in intra-molecular interactions. Their integration can provide a global overview of the behaviour of small proteins and help prioritize them for follow-up studies. Furthermore, analysis of these global datasets with an updated annotation featuring sORF candidates can support their validation. A similar approach was already successfully applied to the identification of infection-related sRNAs in *Streptococcus pneumoniae*, combining RNA-seq

and transposon-insertion sequencing (Mann et al., 2012).

## 1.5 *Salmonella* Typhimurium as a model organism

*Salmonella* enterica serovar Typhimurium belongs to the subspecies enterica of the genus *Salmonella*, and is representative of non-typhoidal *Salmonella* that causes gastroenteritis in humans. The ease of *Salmonella* cultivation in laboratory conditions and of manipulation of its genome made it an ideal model organism, resulting early on in its genome assembly and annotation (Parkhill et al., 2001). It has been extensively used to study bacterial virulence, from how it can adapt to adverse environments encountered on its way to the intestine (Álvarez-Ordóñez et al., 2011), to the machinery it has acquired for invading and hijacking host cells (Fàbrega and Vila, 2013), to its intracellular lifestyle that ends as a self-limiting infection (Ibarra and Steele-Mortimer, 2009).

*Salmonella* enters the host via ingestion of contaminated food or water. At the intestine, it invades the epithelial layer and is taken up by phagocytic cells (LaRock et al., 2015). Once in the cytosol or in the *Salmonella*-containing vacuole, it faces challenges such as low magnesium, host-derived antimicrobial peptides, or low pH (LaRock et al., 2015). In response, *Salmonella* reprograms its gene expression thanks to a broad range of transcription regulators and TCSs (LaRock et al., 2015). The most important and well characterized tools *Salmonella* has acquired horizontally to this end are the *Salmonella* Pathogenicity Islands 1 and 2 (SPI-1 and SPI-2). These are essential for cell invasion and intracellular survival, respectively. Each of these islands encodes a type 3 secretion system (T3SS) and the effectors it secretes, which hijack different host cell pathways in order to create an environment favourable to its replication (Fàbrega and Vila, 2013).

Over the past decade, global datasets that cover several aspects of the *Salmonella* lifestyle have been made available. These span from transcriptomic data of *Salmonella* grown in different conditions (Kröger et al., 2013; Srikumar et al., 2015) to RNA-seq of both host and pathogen (dual RNA-seq) performed during an infection time course (Westermann et al., 2016), to mass spectrometry analysis to

quantify changes in the proteome in response to different conditions (Adkins et al., 2006), among others. This wealth of pre-existing knowledge made *Salmonella* the ideal candidate for studying heterogeneity of a population at a single cell level, particularly important for a pathogen, which has recently been made possible by refinements of the protocols for cell and RNA isolation (Imdahl et al., 2020).

Altogether, this makes *Salmonella* an ideal model organism to approach the annotation and characterization of infection-relevant small proteins.

## 1.6 Aims of this study

It can now be appreciated that small proteins are fundamental for bacterial physiology. Nevertheless, this class of macromolecules is underannotated and understudied. Therefore, with a focus on the model bacterial pathogen *Salmonella* Typhimurium, this work aims at answering two questions. Are there still unannotated small proteins? How are they involved in regulating infection? These points were addressed in the following steps:

- The expansion of the annotation of small proteins using computational and experimental approaches;
- The analysis of infection-relevant datasets with the updated annotation to prioritize uncharacterized candidates based on expression levels (dual RNA-seq) and requirement for virulence (TraDIS);
- A preliminary functional characterization of interesting candidates as a proof-of-principle of the power of datasets integration.

While the outcome of this work will shed light on new small proteins in *Salmonella* and how they might regulate pathogenesis, it will also provide a blueprint for similar works focusing on less studied bacteria, for which a growing body of work is producing global, unexplored datasets at an unprecedented pace.

# 2 Global datasets analysis

The work presented in this chapter has been published in:

**Venturini E.**, Svensson S.L., Maaß S., Gelhausen R., Eggenhofer F., Li L., Cain A.K., Parkhill J., Becher D., Backofen R., Barquist L., Sharma C.M., Westermann A.J., and Vogel J., 2020. A global data-driven census of *Salmonella* small proteins and their potential functions in bacterial virulence. *microLife* 1, no. 1, uqaa002.

As recently recognized, the annotation of small proteins, here defined as shorter than 100 aa, is far from being complete (Section 1). To address this knowledge gap, with a focus on the bacterial pathogen *Salmonella* Typhimurium, several datasets will be described with the purpose of annotating new sORFs and drawing a global picture of small proteins in relation to infection.

## 2.1 An updated annotation of the *Salmonella* small proteome

Two types of data will be analysed to compile a list of novel small protein candidates in *Salmonella*. The first was derived from sPepFinder (Li and Chao, 2020), a computational tool designed to predict small proteins in bacterial genomes. The second was generated with Ribo-seq (Fig. 2.1).

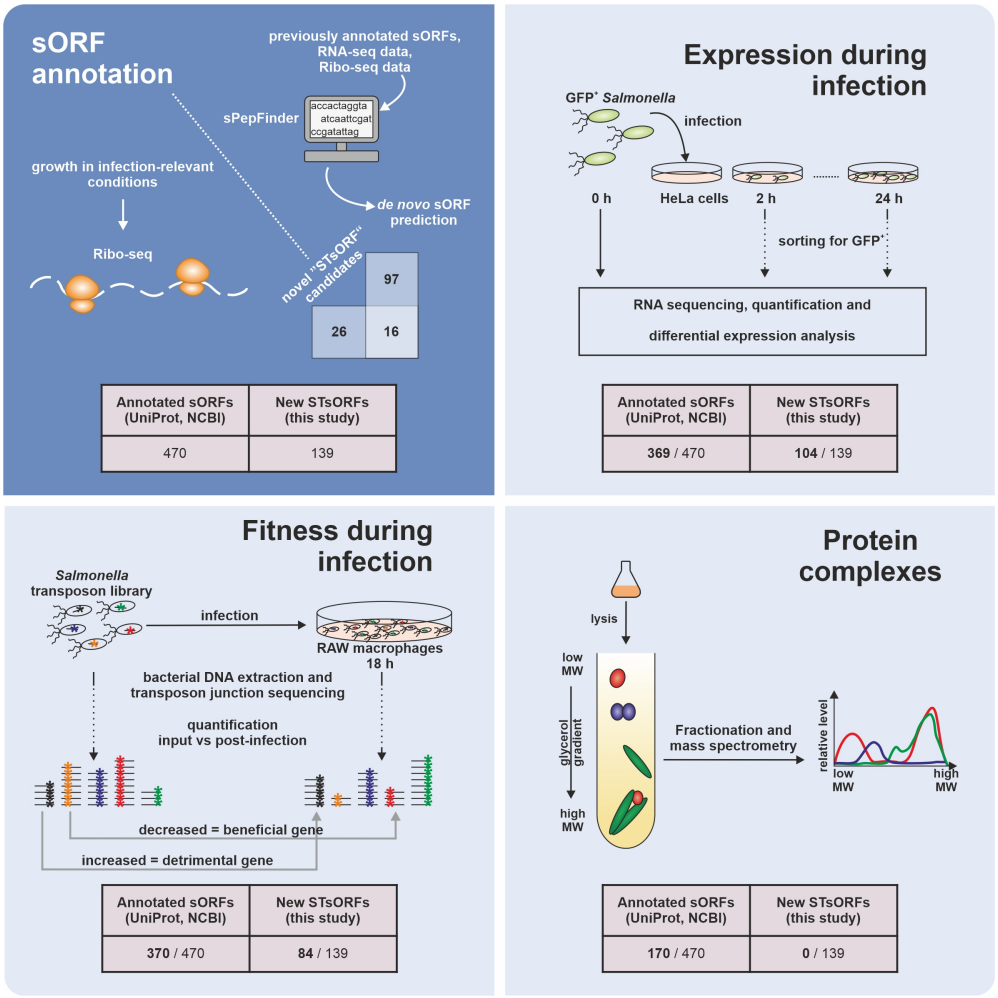



**Figure 2.1: Overview of the techniques applied for the annotation and preliminary functional characterization of small proteins in *Salmonella*.** For the annotation of new *Salmonella* Typhimurium sORFs (STsORFs; top left), a list of candidates was generated based on predictions from sPepFinder and Ribo-seq and merged with the current *Salmonella* annotation. Dual RNA-seq data (top right) were analysed with the updated sORF annotation to highlight differentially expressed transcripts. TraDIS data (bottom left) were used to rank sORFs based on their influence on infection outcome. Grad-seq (bottom right) informed on the engagement of small proteins in larger soluble complexes. Below each cartoon, the number of known (annotated sORFs) or newly annotated small proteins (new STsORFs) that were detected is shown.

### 2.1.1 Small proteins predicted by sPepFinder

sPepFinder is a machine-learning algorithm designed to annotate small proteins in bacterial genomes (Li and Chao, 2020). Its training is performed on known small proteins from ten bacterial species, among which is *Salmonella*, and is based on different features extrapolated from the training set.

Three hundred and forty small candidates shorter than 100 aa were predicted in *Salmonella*. These were filtered to exclude ones with low statistical confidence (support vector machine $< 0.9$) and ones which lacked evidence of transcription. For this, the candidate sORFs were cross-referenced with transcriptomic data and TSS annotation from *Salmonella* grown in 22 *in vitro* conditions (Srikumar et al., 2015). This step is particularly important as it adds a further layer of validation to sPepFinder predictions, which would otherwise be solely based on genomic features with no experimental indication of expression. Overall, these two filtering steps reduced the number of high-confidence candidates to 113.

### 2.1.2 Small proteins predicted by Ribo-seq

Ribo-seq was performed on *Salmonella* grown in three *in vitro* conditions to provide a list of protein candidates based on experimental data. In particular, wild-type *Salmonella* was grown in LB medium to an $OD_{600}$ of 0.4 (mid-exponential phase), and in two infection-mimicking conditions that are SPI-1- and SPI-2-inducing. The first condition is achieved by growing cells in LB medium to $OD_{600} = 2.0$, while the second by growing cells in minimal medium (called "SPI-2") to an $OD_{600}$ of 0.3. For each condition, total RNA was extracted in parallel to the preparation of 70S ribosome footprints.

The data were analysed with REPARATION (Ndah et al., 2017): the tool called 356 known sORFs and 282 new CDSs. The candidate sORFs were filtered to exclude lowly-abundant transcripts (see Methods Section 6.4.5). Coverage files of the remaining candidates were visually inspected to ensure that reads would accumulate in proximity to the putative start codon in the 70S-associated RNA samples compared to the total RNA.

Forty-two candidates remained after filtering (Fig. 2.2a). These overlapped

with 16 of the 113 genes predicted by sPepFinder (Fig. 2.2a). The limited number of shared genes can be explained by the fact that sPepFinder is not limited by experimental conditions, while Ribo-seq is. Therefore, the addition of other experimental conditions could improve this overlap. Nevertheless, the combination of the predictions resulted in 139 candidates with a prevalence of proteins shorter than 50 aa (Fig. 2.2a). The coordinates of each sORF were cross-referenced with a TSS annotation generated with *Salmonella* wild-type grown in different conditions (Srikumar et al., 2015) to classify the new genes based on their localization relative to known features. The sORFs were divided among independent, 5' untranslated region (UTR)-encoded, 3'UTR-encoded, overlapping with sRNAs, or encoded between other genes in an operon (Fig. 2.2b), showing a marked prevalence of 5'UTR-encoded sORFs. These candidates could account for new regulators of downstream genes expression, acting as leader peptides, or small proteins engaged in larger complexes with the products of the downstream genes. Another abundant category is that of independent sORFs. In few instances, these candidates attribute a product to TSSs currently annotated as orphans. Lastly, the identification of new sORFs overlapping partially or entirely with ncRNAs could expand the emerging class of small proteins encoded by dual function sRNAs.

The nomenclature proposed for these new genes is "STsORF", followed by sequential numbers.

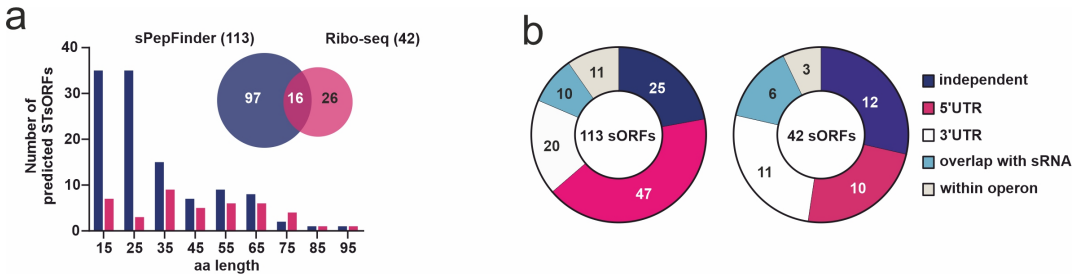



**Figure 2.2: Overview of the new STsORFs added to *Salmonella* annotation. a** Length distribution (histogram) of, and overlap (Venn diagram) between, the STsORFs called by sPepFinder and Ribo-seq. **b** Location of the new STsORFs predicted by sPepFinder (left) and Ribo-seq (right) relative to annotated genes.

### 2.1.3 An updated annotation of the *Salmonella* small proteome

The 16 candidates common to both approaches were chromosomally tagged wiht a SPA tag at the C-terminus and the resulting strains grown in the same *in vitro* conditions used for Ribo-seq, to experimentally validate protein existence via western blotting. This gave proof of translation of 15 out of 16 candidates (Fig. 2.3).

While lack of validation of one candidate, STsORF138, could be explained with low levels or low stability of the target protein, the remaining candidates provide interesting examples of novel small proteins. Overall, the majority of the proteins accumulated in SPI-2 conditions, with STsORF23, STsORF128, and STsORF139 being the most abundant ones. Of the two candidates STsORF43 and STsORF44, only STsORF44 was called by both Ribo-seq and sPepFinder, while STsORF43 only by Ribo-seq. However, translation of a STsORF43 homolog has been validated in the *Salmonella* strain 14028s (Baek et al., 2017). Both genes are encoded inside the sRNA STnc1480. The expression of this sRNA, induced in SPI-2 conditions, is regulated by PhoPQ and SlyA (Kröger et al., 2013); however, its role in *Salmonella* has not been characterized to date. This suggests that the sRNA might be a short mRNA encoding two small proteins.

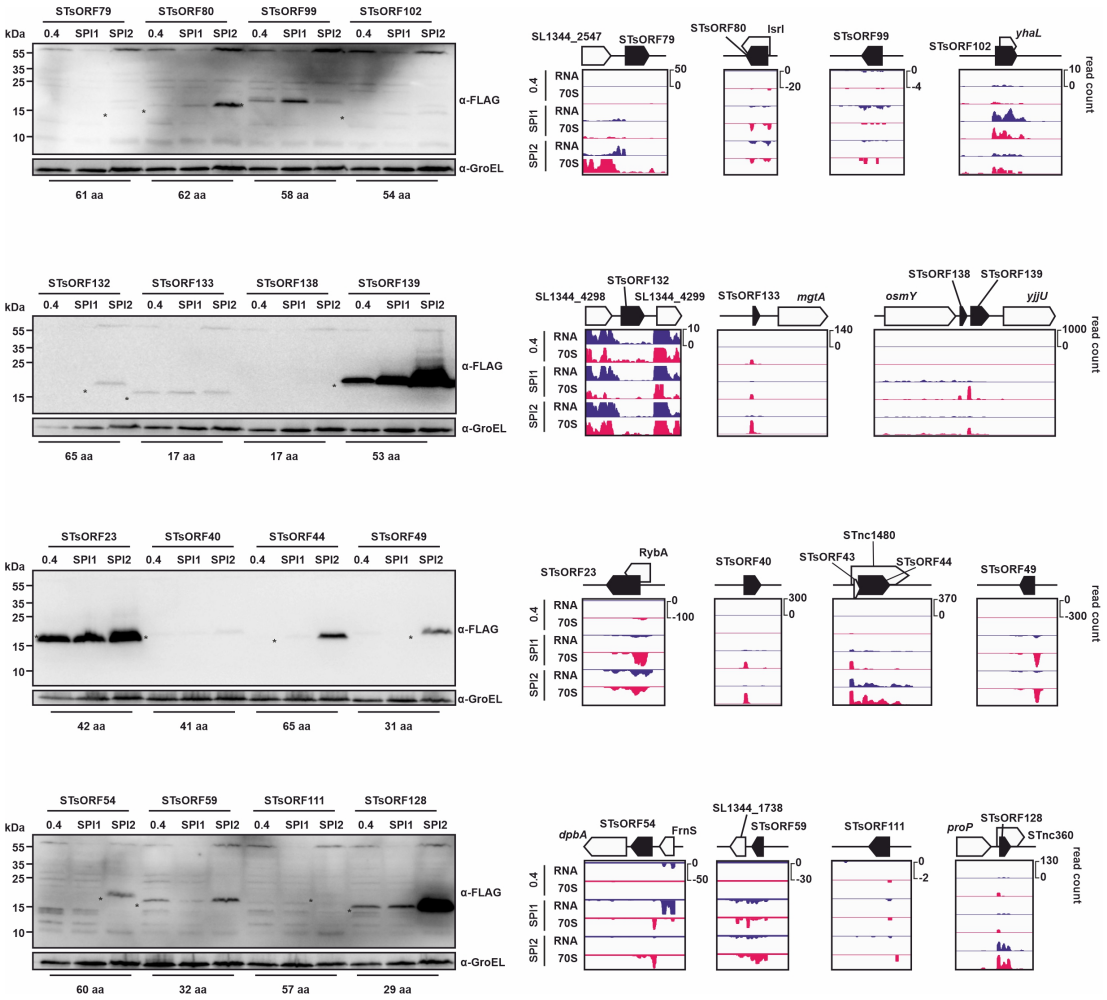



**Figure 2.3: Validation of the candidates identified by both sPepFinder and Ribo-seq.** For each STsORF indicated, the C-terminally SPA-tagged strain was grown in LB to $OD_{600}$ of 0.4 (0.4), in SPI-1-inducing conditions (SPI1), and SPI-2-inducing conditions (SPI2). Protein samples were collected and analysed via western blot (left). The FLAG antibody was used to detect the SPA-tagged proteins, while GroEL was probed for as loading control. The corresponding coverage plots from Ribo-seq are shown on the right, with the localization of neighbouring genes on top. The growth conditions are the same as for western blot analysis. RNA indicates reads from total RNA-seq, while 70S indicates reads derived from the footprints. The panel is representative of two biological replicates.

Another work sought to annotate new genes in the *Salmonella* strain 14028s (Baek et al., 2017). Cross-referencing these predictions provided further support for 7 candidates called by sPepFinder and 2 by Ribo-seq. On the other hand, 66 new short genes predicted in *Salmonella* 14028s were not identified in *Salmonella* SL1344, which can be explained by differences in the experimental setup or data

analysis, as well as genetic differences between the two strains (Henry et al., 2005; Clark et al., 2011).

So far, sPepFinder and Ribo-seq together led to the annotation of 139 STsORFs (Table 8.1). For the following analyses, these 139 candidates were merged with the already known 470 small proteins of *Salmonella*, for a total of 609 sORFs.

## 2.2 Drawing a global picture of small protein expression, essentiality, and interactome

To accelerate the identification of infection-related small proteins, the updated *Salmonella* annotation was used to analyse new and existing datasets that attribute to each small protein an expression pattern in infection (dual RNA-seq), requirement for virulence (TraDIS), and engagement in inter-molecular interactions (Grad-seq). In this way, STsORFs will be ranked to pinpoint candidates with the highest potential of being involved in infection for follow-up functional characterization.

### 2.2.1 Dual RNA-seq

Dual RNA-seq informs on the expression pattern of each gene throughout infection (Section 1.4.1, Fig. 2.1). To identify sORFs induced intracellularly, data generated from a *Salmonella* infection of epithelial cells (Westermann et al., 2016) were reanalysed with the updated STsORF annotation. For every bacterial gene, the expression levels at each time point are compared to the level in the inoculum. Among the mRNAs of the 470 previously annotated small proteins, 280 were significantly differentially expressed during infection (FDR $< 0.05$). Two of the most induced genes, *ssaI* (82 aa) and *ssaS* (88 aa), encode known members of the SPI-2 T3SS, confirming that intracellular expression levels can be considered as a readout for an involvement in infection. The two genes were followed by the one encoding the uncharacterized protein YjiS (54 aa; Fig. 2.4a).

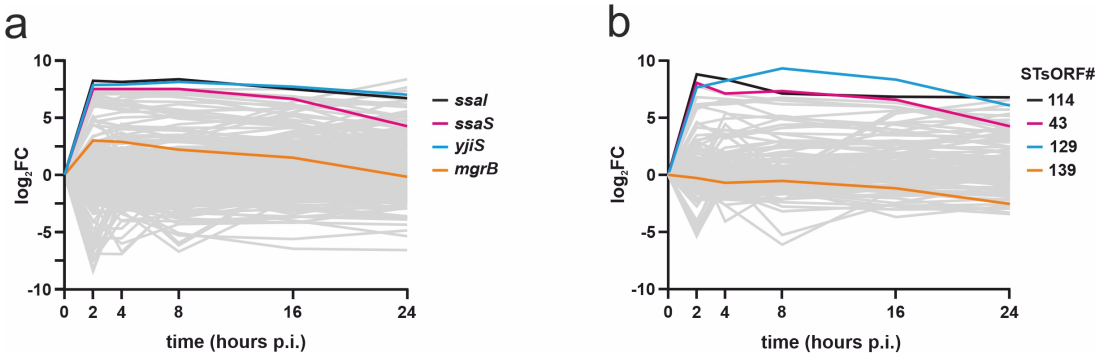



**Figure 2.4: Dual RNA-seq shows the intracellular expression pattern of sORFs. a** Expression pattern of known small proteins during infection of epithelial cells. **b** Expression pattern of STsORFs during infection of epithelial cells. In both panels, the colored genes are the most highly induced ones at 2 h p.i., as well as *mgrB* (left panel) and STsORF139 (right panel). Data were taken from (Westermann et al., 2016).

With respect to the newly annotated STsORFs, 101 were differentially expressed intracellularly as compared to the inoculum (FDR $< 0.05$), while three more were detected but did not pass the statistical filtering for differential expression (Fig. 2.4b).

The most strongly upregulated genes reached a rate of induction comparable to that of *ssaI* and *ssaS* (Fig. 2.4a and b). The expression of the top three STsORFs, STsORF114, 43, and 129, was confirmed by the SalComMac database featuring transcriptomic data from *Salmonella* grown in different conditions ((Srikumar et al., 2015); Fig. 2.5). STsORF114 (16 aa) is the homologue of the *Salmonella* strain LT2 MgtP, regulator of the expression of the downstream genes (Lee and Groisman, 2012b). STsORF43 (13 aa) overlaps with the sRNA STnc1480 (Fig. 2.5a), together with STsORF44 (65 aa) which was previously validated (Fig. 2.2c). STsORF129 (32 aa), on the other hand, appears to be independently transcribed from a TSS internal to the *phoN* gene (Fig. 2.5b). These STsORFs, with their strong intracellular induction, represent ideal candidates for validation of translation. Similarly to the small proteins validated in Fig. 2.2, these three STsORFs were tested via epitope tagging and western blotting. Unlike for STsORF129, the translation of STsORF114 and STsORF43 could be validated (Fig. 2.5c).

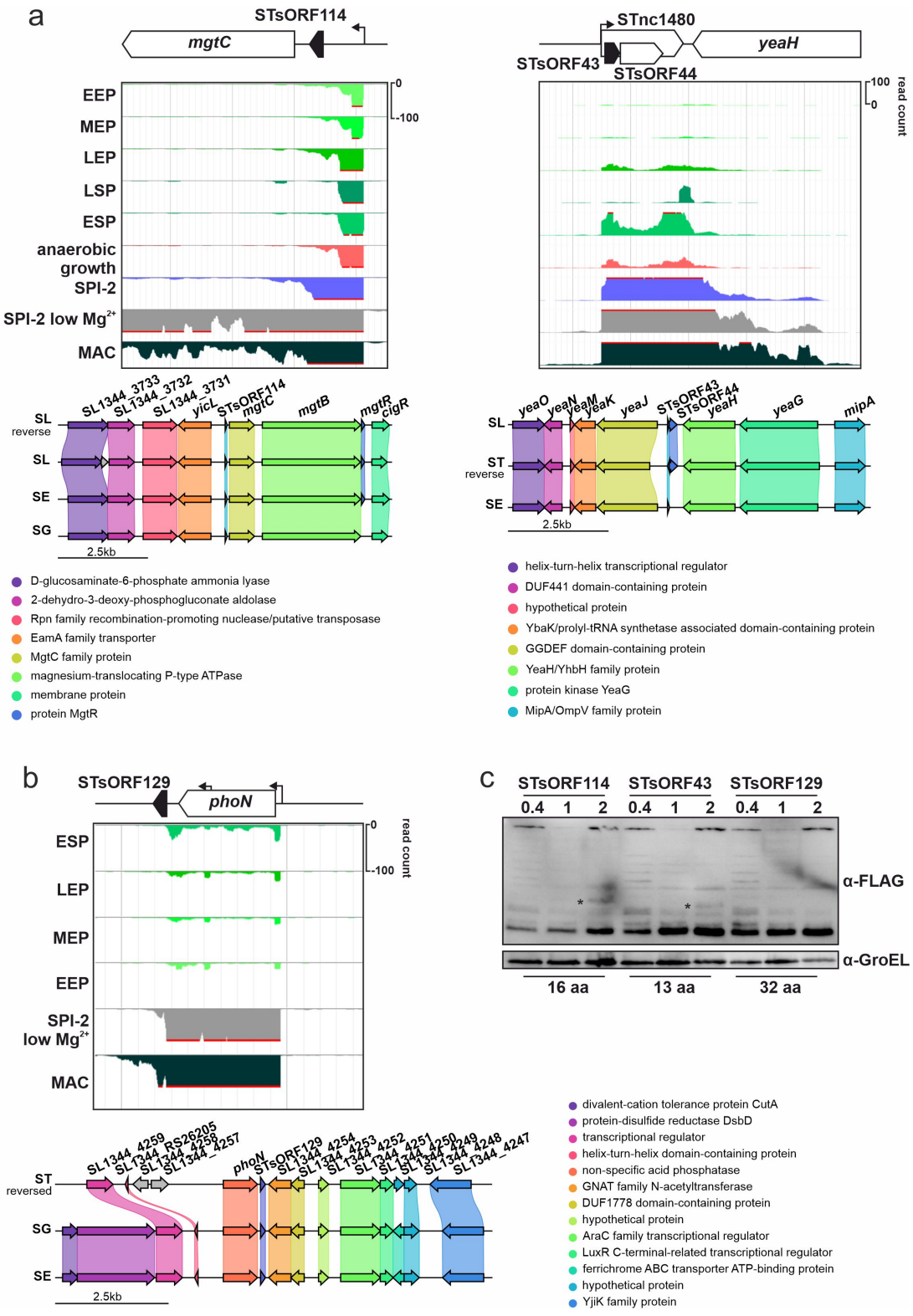



**Figure 2.5:** (Continued on next page.)

**Figure 2.5:** (Previous page.)**STsORFs that are highly induced intracellularly. a**, **b** Genomic context, read coverages (Srikumar et al., 2015), and conservation at the nucleotide level of the small protein candidates STsORF114, STsORF43, and STsORF129. In both panels, the growth conditions are: EEP: early exponential phase; MEP: mid exponential phase; LEP: late exponential phase; LSP: late stationary phase; ESP: early stationary phase; MAC: macrophages. The graphs below each coverage plots indicate which annotated neighbouring genes are conserved. The genomes were retrieved with a BLAST search of the STsORFs, and the alignments were performed with clinker (Gilchrist and Chooi, 2020). The genomes are *Salmonella* Typhimurium: "SL"; *Salmonella* Typhi: "ST"; *Salmonella* enteritidis: "SE"; *Salmonella* Gallinarum: "SG". These are the only genomes in which homologs could be identified. **c**, Validation via western blotting of the SPA-tagged small proteins from panels a and b. The strains were grown either in LB medium to an $OD_{600}$ of 0.4, in SPI-1, or SPI-2 conditions, marked as "0.4", "1", and "2", respectively.

Given that intracellular induction in expression is a hallmark of genes involved in adaptation and survival in the host cell, the extent of expression of *yjiS*, STsORF43, and STsORF44 puts them at the top of the list when ranking small genes whose impact on infection should be evaluated.

### 2.2.2 TraDIS

While intracellular induction is a good indication of a gene requirement for pathogenesis, it is not direct proof. This type of information can instead be derived from the transposon-directed insertion sequencing (TraDIS) approach (Section 1.4.2).

To this end, a library of ~100,000 transposon mutants of *Salmonella* was generated and used to infect RAW264.7 macrophages. The abundance of each mutant was quantified in the pool both in the inoculum and at 20 hours post infection (p.i.). For each gene, a decrease in abundance indicates requirement for infection, while an increase suggests anti-virulence properties. Selection during infection resulted in differential abundance changes ranging from $\log_2$fold-change (FC) of -6.5 to 5.9, although the majority of genes fell in the $|\log_2 FC| < 1$ interval (Fig. 2.6a). Mutants of SPI-2 genes are expected to be counterselected, as required for intracellular survival. This was confirmed by the lower abundance of mutants of genes encoding SPI-2 proteins such as SsaV and SseC ($\log_2 FC = -1.2$), and SsaO ($\log_2 FC = -1.4$). A strong negative selection was however observed for metabolic genes such as *purA* ($\log_2 FC = -6.5$, involved in purine biosynthesis), or structural components of the membrane such as the major lipoprotein *lpp* ($\log_2 FC = -4.8$).

Conversely, disruption of several genes promoted virulence, as exemplified by the *rfa/rfb* clusters ($\log_2$FC up to 5.8). This particular case is supported by a previous work indicating that these operons, whose products are involved in lipopolyssaccharide O-antigen assembly, promote *Salmonella* intracellular survival (Zenk et al., 2009).

Overall, the library included transposon insertions in 454 sORFs out of 609. The differential abundance of each mutant is shown in Fig. 2.6b. Only a limited number of sORFs had a $|\log_2\text{FC}|$ higher than 1, and for even fewer genes this value was significant (FDR $< 0.05$). As expected, disruption of *sseA* (89 aa) decreased *Salmonella* fitness, since it encodes a structural component of the SPI-2 T3SS. The magnitude of decrease in abundance after infection resembled that of larger infection-relevant proteins such as SsaV and SseC (Fig. 2.6a). This suggests that TraDIS can also highlight small genes involved in virulence.

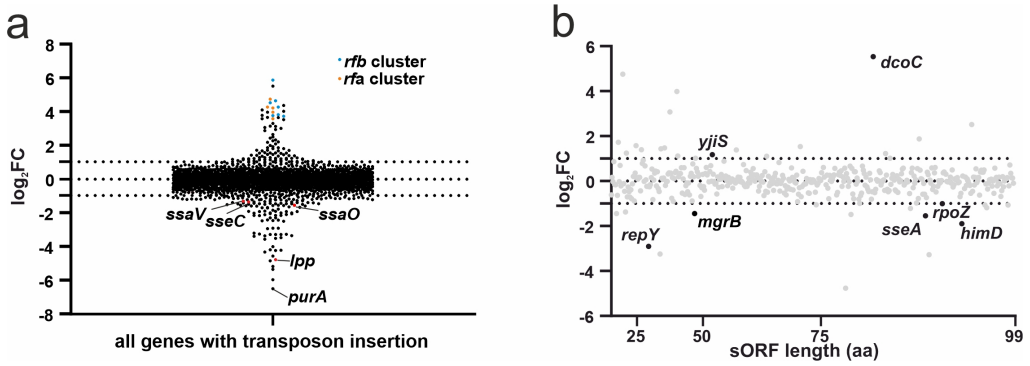



**Figure 2.6: Small proteins with a transposon insertion assayed for macrophages infection. a** Differential abundance of *Salmonella* genes targeted by transposon insertion. The $\log_2$FC indicates the differential abundance of each mutant after infection as compared to its level in the inoculum. Relevant genes are indicated and mentioned throughout the text. **b** The plot shows the difference in abundance at 20h p.i. vs in the inoculum, displayed on the y-axis as a $\log_2$FC, of all the sORFs where a transposon insertion occurred. Each gene is represented with a gray dot. For the ones that passed the statistical cutoff (FDR $< 0.05$) the name is included and they are marked in black.

Among the other genes that passed the statistical filtering are *rpoZ* (91 aa), important for cell growth as it is a subunit of the RNA polymerase. Given that the PhoPQ TCS is essential for virulence (Cho et al., 2017), it is reasonable that the absence of its regulator MgrB (47 aa) leads to an infection defect. Along the same lines, *repY* (29 aa) and *himD* (94 aa) were also required for virulence. They

encode for the regulator of the plasmid pCol1B9 replication and the β-subunit of the integration host factor, respectively, with HimD already described as a positive regulator of *Salmonella* infection (Mangan et al., 2006; Yoon et al., 2009). While these small proteins had a negative impact on *Salmonella* fitness, disruption of two other genes had an opposite outcome. The absence of either *yjiS* (54 aa) or *dcoC* (81 aa), whose products are so far uncharacterised, rendered *Salmonella* more virulent. *dcoc*, annotated as a probable subunit of the oxaloacetate decarboxylase, was the sORF with the most extreme change in abundance ($\log_2$FC of 5.5). However, there is currently no evidence of its expression in any condition tested in SalComMac (Srikumar et al., 2015).

None of the new STsORF mutants had significant changes, despite showing strong levels of differential abundance. STsORF138, encoded together with STsORF139 from the 3'UTR of *osmY* (Fig. 2.2c), had a $\log_2$FC of 4.7 (q-value = 0.18). Of note, STsORF139 had a $\log_2$FC of 0.0075 (q-value = 1), and *osmY* had a $\log_2$FC of 0.23 (q-value = 0.85). This opens up two scenarios for interpretation of this phenotype: either STsORF138 is the only gene in the operon required for infection (although its translation could not be validated (Fig. 2.2)), or its disruption impacts the expression of its neighbouring genes. The average number of insertions in STsORF138 throughout the two replicates of inoculum and 20h p.i. was 0.75, while in *osmY* was 5.5 and for STsORF139 was 2. Thus, the low insertion rate could explain the low confidence in the dysregulation.

Overall, TraDIS data provide a global snapshot of genes required for infection, including short ORFs. Uncharacterized genes such as YjiS or STsORF138-STsORF139 represent excellent candidates for follow-up studies, given their potential impact on infection.

### 2.2.3 Grad-seq

Neither dual RNA-seq nor TraDIS provide direct proof of a protein's existence. For this, one useful dataset to analyse is gradient profiling by sequencing, or Grad-seq (Section 1.4.3). Based on the separation of soluble cytosolic complexes on a glycerol gradient, its fractionation and analysis by mass spectrometry, it informs on the sedimentation pattern of each protein detected and allows to make educated

guesses regarding molecules engaged in inter-molecular interactions. In the present work, Grad-seq data from Gerovac et al. (2020) were re-analysed with the updated small proteins annotation.

Of the 4'657 proteins annotated in *Salmonella*, 2'225 were detected in the gradient. None of the new STsORFs was recovered, while 170 out of 470 known small proteins were identified. These include both validated small proteins, as well as ones that are currently marked as predicted or hypothetical, annotated based on genome analysis but lacking experimental validation. This is the case of SL1344_-0350 or YdcY (64 and 77 aa, respectively; Fig. 2.7). It can be inferred that a given small protein is interacting with other molecules if it sediments beyond the first two fractions. As a proof of principle, the small protein RpoZ (91 aa) localized well within the gradient, with a profile overlapping with the other subunits of the RNA polymerase (Fig. 2.7). Another Grad-seq study applied to *E. coli* grown in LB medium to an $OD_{600}$ of 2.0, the same condition used for the *Salmonella* Grad-seq, reported the same behavior of RpoZ co-sedimenting with other RNA polymerase subunits (Hör et al., 2020a).

The focus was directed to uncharacterized small proteins, which were 89 of the 170 detected. Of these, 82 migrated beyond the first two fractions, which is a good indication of engagement in interactions (Fig. 2.7). For example, SL1344_-1543 (93 aa) was highly abundant in the pellet, the fraction containing the largest complexes such as the ribosomes. Instead, SL1344_2732 (96 aa) peaked in a region that overlaps with the RNA polymerase. While these small proteins correlate with known larger complexes, direct interaction should be proven with further analysis.

*E. coli* Grad-seq data (Hör et al., 2020a) were mined to find homologs of uncharacterized small proteins from the *Salmonella* gradient. A total of 2'145 proteins were detected in the textitE. coli gradient, 134 of which were shorter than 100 aa. Fifty-three of these showed conservation with the uncharacterized small proteins of interest. The comparison of their profiles is shown in Fig. 2.8: in both cases, the majority of the small proteins sediments among the first five fractions. It is interesting how the distribution differed in a few cases. In *Salmonella*, YqiC (99 aa) overlaps with the RNA polymerase, while in *E. coli* it rather peaks at the $2^{nd}$ fraction, instead of the $6^{th}$. YqiC was annotated in *Salmonella* based on conser-

vation with UbiK, a ubiquinone biosynthesis factor in *E. coli*. The ubiquinone pathway is highly conserved in bacteria (Meganathan, 2001), although it is known to be important for *Salmonella* virulence (Aussel et al., 2014). Thus, it is possible that YqiC might have acquired different interactors, given the difference in the sedimentation profile.

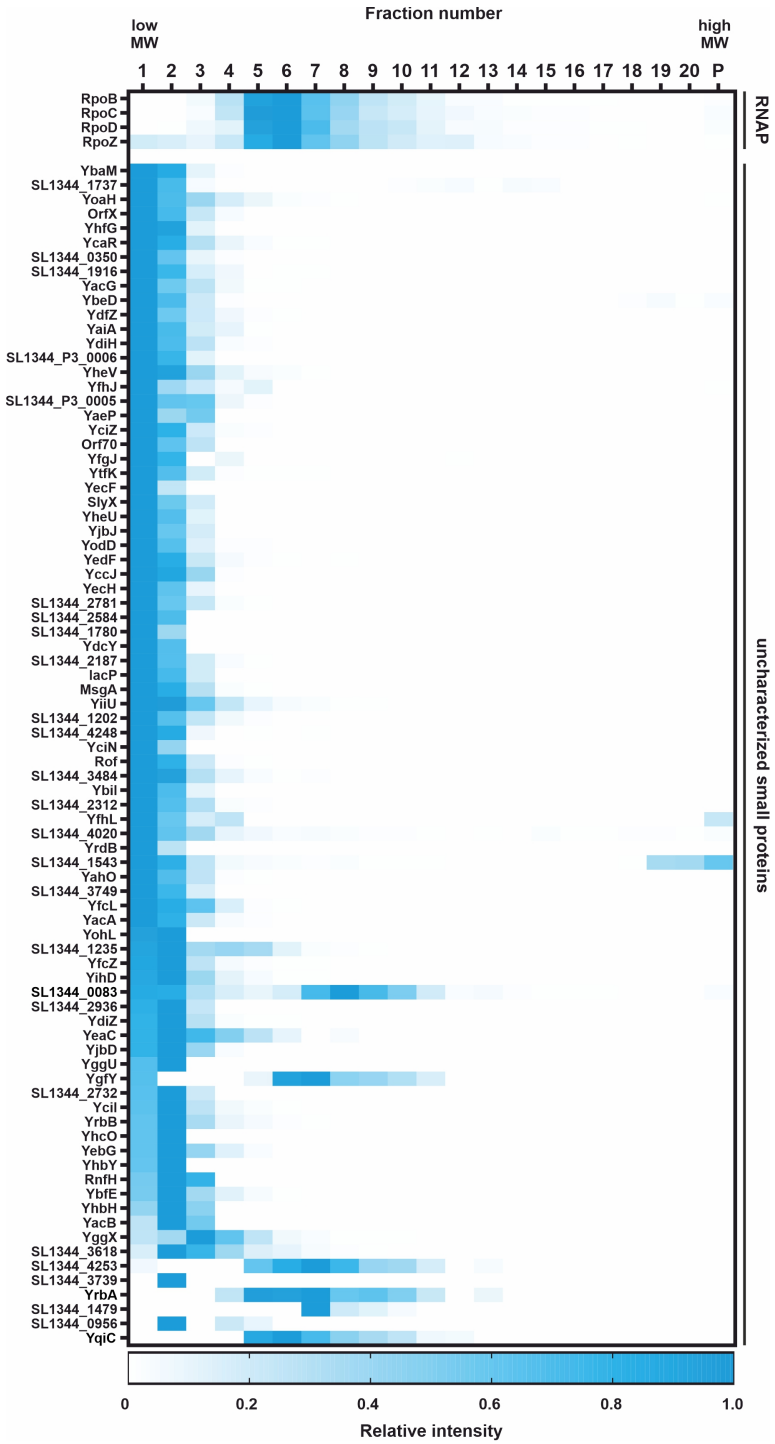



**Figure 2.7: Sedimentation profile of uncharacterized small proteins detected with Grad-seq.** The heat map shows the sedimentation profile of the four subunits of the RNA polymerase detected via mass spectrometry, with the small protein RpoZ having a distribution similar to the larger proteins in the same complex. The proteins below are the 82 uncharacterized small proteins that sedimented beyond the first fractions.

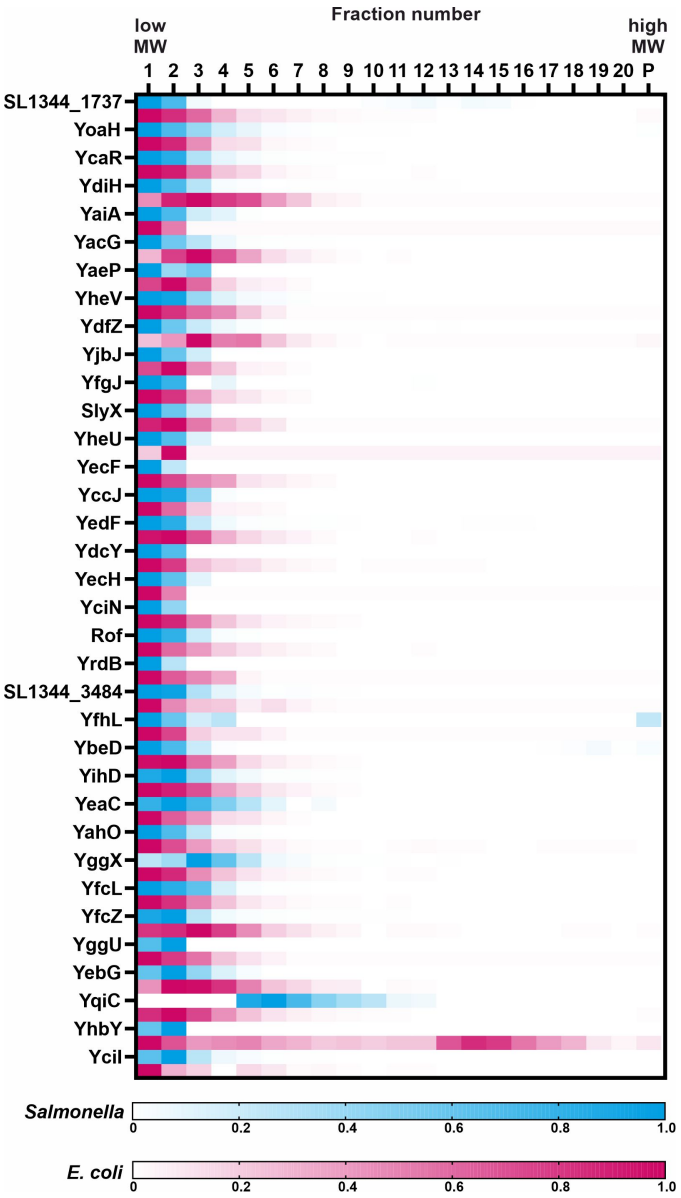



**Figure 2.8: Comparison of the sedimentation profile of uncharacterized small proteins in *Salmonella* whose homologues were detected in an *E. coli* Grad-seq.** The heat map shows the sedimentation profile of the small, uncharacterized proteins from *Salmonella* (blue) and of their homologs detected in an *E. coli* Grad-seq experiment (fucsia, (Hör et al., 2020a))

.

YhbY (97 aa) is another small protein with a marked difference in its distribution. It peaks in fraction two in the *Salmonella* gradient, while in the *E. coli* one it localizes both at the top and around fraction 14. This protein was recently shown

to mediate the assembly of the ribosomal subunits in *E. coli* (Gagarinova et al., 2016), while no characterization has been reported for *Salmonella*. The distribution in *E. coli* fits with the literature, since it localizes in the region between the 30S and 50S subunits (Hör et al., 2020a). Considering that the ribosome assembly pathway is highly conserved, it is surprising to see such a difference in these two closely-related species.

Inferring the actual complex in which a protein is present requires refined downstream analysis of Grad-seq data, as well as thorough independent validation. This notwithstanding, this approach clearly indicates that a number of small proteins are engaged in inter-molecular interactions worth investigating.

## 2.3 Shortlisting interesting small proteins

The functions of several of the aforementioned small proteins are not known. Therefore, an approach to identify the most promising candidates is important to fast-track their characterization. In terms of infection relevance, the combination of dual RNA-seq and TraDIS can highlight genes that are highly expressed in infection and whose presence may be important for pathogenesis.

For example, the uncharacterized protein YjiS is highly induced when *Salmonella* is inside epithelial cells (Fig. 2.4a) and its absence resulted in a mild hypervirulence in RAW 264.7 macrophages infection (Fig. 2.6). Among the new small proteins, STsORF139 stood out for its high abundance upon growth in SPI-2 conditions (Fig. 2.2). Encoded from the 3'UTR of *osmY* together with the small protein candidate STsORF138 (Fig. 2.2c), it depicts an interesting situation where STsORF139 is highly abundant as a protein, but is not required for infection, while disruption of its upstream CDS results in a strong phenotype in the TraDIS setup ($\log_2$FC of 4.7 (q-value = 0.18)).

These are examples of the top candidates provided with these techniques. Mining dual RNA-seq data for more nuances of induction rates, for example, or strong but not significant TraDIS phenotypes will deliver even more information still worth considering.

## 2.4 Discussion

Small protein annotation is challenging in eukaryotes and prokaryotes alike (Storz et al., 2014). The development of tools to identify sORFs in bacterial genomes based on features such as the strength of the SD sequence or homology to protein motifs can only cover cases of conservation of already known small proteins. Since they evolve more quickly than larger genes (Carvunis et al., 2012), many sORFs are missed. With this in mind, two technologies have been employed in this work to implement the annotation of small proteins in the model bacterial pathogen *Salmonella* Typhimurium.

Machine learning tools hold great potential in this context, being able to identify previously unrecognized features of sORFs and thus leading to the annotation of new classes of genes. sPepFinder was developed specifically for this purpose (Li and Chao, 2020). After training on small proteins annotated in ten bacterial species, it predicted 113 high-confidence new small proteins in *Salmonella*. To complement these data, Ribo-seq was applied to the wild-type strain in three conditions, two of which are infection-relevant. While there was small overlap between Ribo-seq- and sPepFinder-derived sORFs, the two approaches resulted in a total of 139 candidates, for which the nomenclature "STsORF" followed by sequential numbers has been used. Independent validation of the 16 STsORFs called by both approaches, and three highly induced genes in infection, via epitope tagging and detection on western blot confirmed the translation of 17 CDSs. Different explanations can account for the lack of validation of two candidates. Tagging at the C-terminus could cause the protein to be unstable or non functional, and thus rapidly degraded. Alternatively, the protein level might be too low for detection via western blot. These issues could be overcome by changing the tag for detection to a smaller one and fuse it to the N-terminus of the protein, or employing mass spectrometry.

Despite being two advanced approaches to tackle the annotation of new small proteins, both sPepFinder and Ribo-seq come with limitations. While sPepFinder might require a larger training set to improve its detection rate, it is nevertheless plausible that some of the predictions it generated have been discarded because no

corresponding expression on the *Salmonella* transcriptomic database SalComMac (Srikumar et al., 2015) was detected. A larger set of environmental conditions used to sample *Salmonella* RNA might reveal if and when discarded sPepFinder candidates might be expressed. On the other hand, ways to improve the confidence with which Ribo-seq identifies sORFs, such as using drugs to stall the ribosomes at the translation start site, can improve its predictions (Meydan et al., 2019). Overall, this could result in a larger number of sORFs called by both approaches.

Merging the new STsORFs (139) with the annotated *Salmonella* small proteome (470) resulted in a total of 609 proteins shorter than 100 aa. This new number constitutes the 13% of the *Salmonella* genome. It is reasonable to believe that this number is destined to increase in the future for multiple reasons. i) In *Salmonella* and *E. coli* alike, the distribution of proteins length indicates a marked under-representation of genes between 10 and 50 aa, as compared to the number of proteins in the other intervals (Appendix Fig. 8.1). ii) Several candidates from sPepFinder and Ribo-seq were discarded due to low confidence. Given the success in the independent validation, it is possible that true, new small proteins have been excluded due to the stringent filtering. iii) Several small proteins might not be expressed in the conditions tested so far. iv) Evidence that sORFs can overlap, partially or entirely, with larger genes, has been accumulating (Adams et al., 2021). This genomic arrangement is difficult to confidently identify without tailored transcriptomic data, but nevertheless holds promise to reveal several new small proteins.

This increase in the number of small proteins highlights one of the parallels that can be drawn between the annotation of sRNAs and small proteins. The initial, serendipitous identifications were followed up by more systematic searches for non coding RNAs in *Salmonella*, leading up to at least 70 sRNAs as estimated in 2009 (Vogel, 2009). Just four years later, this number skyrocketed to 280 sRNAs (Kröger et al., 2013), thanks to the use of the rising high-throughput technique RNA-seq. Recently, a catalogue of the validated sRNAs listed a total of 172 (Hör et al., 2020b). Thus, a good fraction of the 280 predicted sRNAs still await investigation. Another common theme to both class of gene products is the high induction in stress conditions, as shown by the high occurrence of STsORFs whose

abundance increases in infection-mimicking conditions (Fig. 2.3, Fig. 2.5). This is also a common feature of sRNAs (Hör et al., 2020b).

To generate a comprehensive description of small proteins in the context of infection, the updated annotation was first used to re-analyse existing dual RNA-seq data (Westermann et al., 2016). In this way, the expression pattern of each bacterial gene throughout infection is provided. Strong intracellular induction is a hallmark of genes whose products are fundamental for surviving the harsh conditions encountered inside the host cell, or to actively interfere with host processes to create an environment favourable to *Salmonella* survival and proliferation (LaRock et al., 2015). In this context, genes encoded in the SPI-2 island are the ones most extensively investigated. Dual RNA-seq data might instead point out other interesting sORFs with a similar behaviour. For example, *yjiS* is currently uncharacterized and yet it reaches induction levels comparable to those of SPI-2 genes (*ssaI* and *ssaS*; Fig. 2.4a). Similarly, some STsORFs were also highly induced during infection (Fig. 2.4b). These are the first candidates one should turn to when focusing on genes required for intracellular survival. Nevertheless, also highly repressed genes might point at a function in the stages preceding infection, similarly to the situation for SPI-1-encoded proteins, which are required for invasion of the host cells rather than intracellular survival.

TraDIS is an excellent approach to complement dual RNA-seq data, since cross-referencing the two sets shows the rate of intracellular expression of bacterial genes and provides a preliminary indication of their requirement for virulence. Despite the fact that random transposon insertions are more likely to occur in larger genes than shorter ones, TraDIS already proved itself useful in the characterization of sRNAs (Barquist et al., 2013). Therefore, it has been considered in this work for the characterization of sORFs. Overall, the changes in abundance of all mutants in the *Salmonella* library challenged with infection of RAW264.7 macrophages ranged between -6.5 and 5.9 (Fig. 2.6a). The most extreme cases were attributed to the *rfa/rfb* loci in terms of increased benefit for virulence upon disruption. Conversely, metabolic and structural genes such as *purA* or *lpp*, respectively, gave the strongest defect in intracellular survival. A large fraction of sORFs was targeted by transposon insertions, but disruption resulted in a significant impact on bac-

terial abundance in only seven of these genes (Fig. 2.6b). Among these was a SPI-2-encoded gene (*sseA*) which, upon disruption, had the strongest impact on virulence among all the SPI-2 genes ($\log_2$FC = -1.5). This effect is nevertheless not as pronounced as that of genes such as *purA* or *lpp*, despite these latter ones not being directly involved in infection. This observation indicates that the absence of structural members of the T3SS or effectors impairs virulence but not as strongly as the depletion of genes involved in other general metabolic pathways, whose impact on bacterial fitness is exacerbated when the pathogen enters the nutrient-limited host cell. Contrary to this, two uncharacterized small proteins (*yjiS* and *dcoC*) showed potential for being new anti-virulence factors since the absence of either had a positive effect on bacterial abundance (Fig. 2.6b). While *dcoC*, a putative subunit of the oxaloacetate decarboxylase, showed a high $\log_2$FC (5.5), no expression of its gene could be detected so far (Kröger et al., 2013). Intriguingly, dual RNA-seq indicated *yjiS* was one of the top three intracellularly induced genes. Despite showing extreme fold-changes in the TraDIS data, none of the other STsORF mutants passed the threshold for statistical significance. For example, interruption of the candidate STsORF138 increased bacterial abundance of more than 25 fold (q-value = 0.18). One explanation for this could be a low rate of transposon insertion that compromised the statistical significance. Independent validation of the data with of clean deletion mutants will be necessary to confirm TraDIS phenotypes. Furthermore, this is an essential step to overcome population effects that might amplify or dampen the phenotype of a given mutant.

The power of Grad-seq lies in the capacity to globally assess the sedimentation behavior of proteins and RNAs, information that can be used to infer their involvement in intermolecular complexes (Gerovac et al., 2020). This is particularly straightforward in the case of small proteins. Based on their size, in the setup used for the Grad-seq data analysed here (Gerovac et al., 2020), they are not expected to migrate beyond the first fractions, which contain molecules not engaged in any interaction. Therefore, the attention is drawn to cases such as SL1344_0350 or YdcY (Fig. 2.7). Another benefit of Grad-seq is the direct proof of a protein's existence. Nevertheless, limitations in mass spectrometry protocols currently applied can overlook small proteins. Therefore, variations of standard

methods to include a step to enrich small proteins will aid in this challenge. When using Grad-seq data to complement dual RNA-seq and TraDIS, it is notable that some of the most interesting candidates are not detected in Grad-seq. For example, YjiS was missing because it is not expressed in the growth condition used that is SPI-1 inducing, while *yjiS* is specifically expressed in intracellular conditions (Fig. 2.4a). MgrB, on the other hand, is known to be an inner membrane protein (Lippa and Goulian, 2009), whereas the lysis step currently implemented in the Grad-seq protocol enriches cytosolic and periplasmic proteins. Nevertheless, interesting candidates to work with should be selected among the ones that sedimented beyond the first two fractions. A comparison with the homologues in *E. coli* with a different distribution (Fig. 2.8) can suggest which small proteins might have an interaction that has diverged in the two bacterial species.

It is expected that the combination of these techniques can highlight the top high-priority candidates when approaching the characterization of new or hypothetical small proteins potentially involved in infection. For example MgrB, a regulator of the PhoPQ TCS (Lippa and Goulian, 2009), has a clear involvement in infection which lacks thorough investigation. On the other hand, the uncharacterized protein YjiS is an ideal example of an anti-virulence gene that is highly induced during infection. Among the new small proteins, STsORF139 stands out as extremely abundant, especially in SPI-2 conditions (Fig. 2.2c). Furthermore, it is encoded in the 3'UTR of *osmY* together with the candidate STsORF138, which had a remarkable impact on infection when disrupted. Contrary to YjiS, a standalone gene about which nothing is known in *Salmonella* or *E. coli*, STsORF139 has the potential of being active in the same conditions in which OsmY is induced, namely high osmolarity stress (Yim and Villarejo, 1992).

Overall, these complementary methods enhance confidence in the identification of small proteins which are likely involved in a given process - in this work, *Salmonella* pathogenesis. As high-throughput techniques are being applied to an increasing number of model organisms and understudied prokaryotes, re-analysis of these data with an eye on small proteins can deliver a great deal of information.

# 3 MgrB

Parts of the work presented in this chapter have been published in:

**Venturini E.**, Svensson S.L., Maaß S., Gelhausen R., Eggenhofer F., Li L., Cain A.K., Parkhill J., Becher D., Backofen R., Barquist L., Sharma C.M., Westermann A.J., and Vogel J., 2020. A global data-driven census of *Salmonella* small proteins and their potential functions in bacterial virulence. *microLife* 1, no. 1, uqaa002.

The small protein MgrB is known to regulate the activity of the PhoPQ TCS (Lippa and Goulian, 2009) which, in *Salmonella* and *E. coli* alike, mediates the adaptation to acidic pH and low magnesium (Monsieurs et al., 2005). Furthermore, *Salmonella* PhoPQ is also a key player in the response to antimicrobial peptides (Dalebroux and Miller, 2014). This, together with the requirement of *mgrB* for infection suggested by TraDIS (Fig. 2.6b), makes MgrB an ideal small protein to be functionally investigated in the context of infection.

## 3.1 The impact of MgrB on *Salmonella* virulence

### 3.1.1 MgrB is required for infection of epithelial cells and macrophages

To assess the impact MgrB has on infection, and to confirm the phenotype observed in TraDIS (Fig. 2.6), a strain lacking the *mgrB* coding sequence was generated

($\Delta mgrB$). Compared to wild-type *Salmonella*, no change in growth in either LB or minimal media was observed (Fig. 3.1).

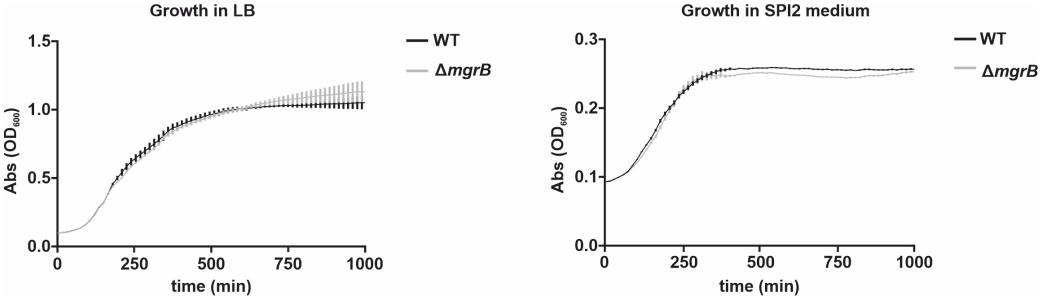



**Figure 3.1: $\Delta$*mgrB* *Salmonella* shows no growth defect.** *Salmonella* wild-type and $\Delta mgrB$ strains were grown in LB medium or SPI-2 medium. The graph shows the average of triplicates, with error bars showing standard error from the mean.

To validate the TraDIS phenotype caused by the absence of MgrB (Fig. 2.6b), the $\Delta mgrB$ strain and a strain with a plasmid carrying *mgrB* with its native promoter ($mgrB^+$) were compared to the wild-type strain during infection of macrophages and epithelial cells. A colony forming unit (c.f.u.) assay was performed for both host cell lines with a multiplicity of infection (m.o.i.) of 10 for RAW264.7 macrophages and an m.o.i. of 25 for HeLa cells (Fig. 3.2a). In parallel, the bacterial load for single host cells was measured via flow cytometry, making use of strains expressing the green fluorescent protein (GFP; Fig. 3.2b, c).

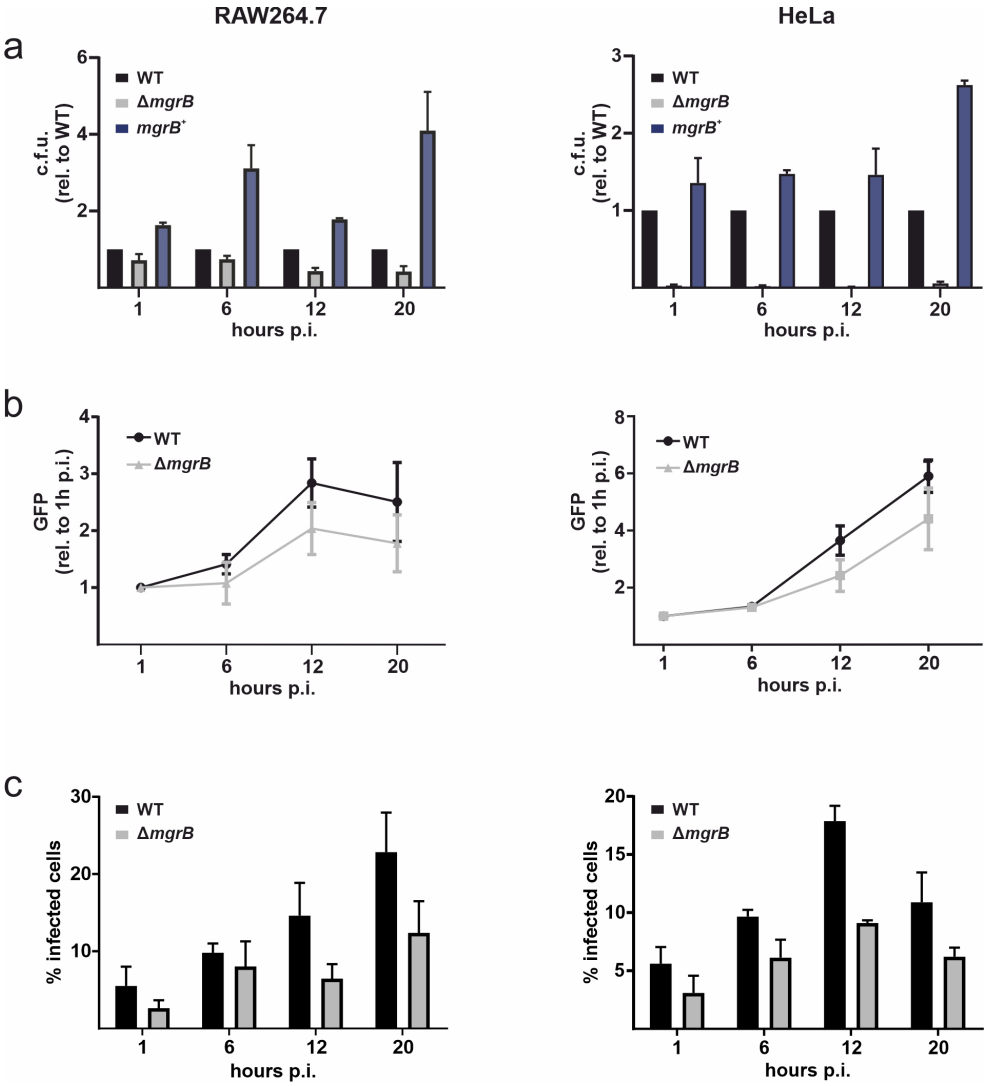



**Figure 3.2: MgrB is required for infection of epithelial cells and macrophages.** Wild-type, Δ*mgrB*, and *mgrB*$^+$ *Salmonella* were used to infect RAW264.7 macrophages (m.o.i. of 10; left side) or HeLa cells (m.o.i. of 25; right side). **a** C.f.u. quantification from gentamicin protection assays. **b** Amount of intracellular bacteria, quantified with GFP$^+$ strains, inside the host cells relative to the 1h time point. **c** Percentage of infected (GFP$^+$) cells. In all panels averages from three replicates are shown, and the bars indicate standard errors from the mean. Wild-type and Δ*mgrB Salmonella*, when compared to the *mgrB*$^+$ strain, carry the empty vector control.

The infection assays show that MgrB is required for infection of both host cell lines, an effect that is specific since the *mgrB*$^+$ strain (over)complements this phenotype (Fig. 3.2a). Furthermore, flow cytometry data show not only that

$\Delta mgrB$ *Salmonella* is unable to survive and/or replicate in the host cells, but also that the overall population of infected cells is smaller (Fig. 3.2b, c). Altogether, this can account for the low bacterial load observed in the c.f.u. assay and validates the TraDIS phenotype.

### 3.1.2 Lack of MgrB dampens the level of SPI-1 effectors

It is known that while *Salmonella* is passively taken up by macrophages, it requires its SPI-1 T3SS to secrete effector proteins to actively invade epithelial cells (Fàbrega and Vila, 2013). Given that the $\Delta mgrB$ mutant was markedly impaired at the first infection stages in HeLa cells as compared to RAW264.7 macrophages, differences in the secretome were analysed. For this, the *Salmonella* wild-type, $\Delta mgrB$, and $mgrB^{+}$ strains were grown in SPI-1-inducing conditions, that is in LB medium to an $OD_{600}$ of 2, and both the whole proteome and the secretome were visualized via Coomassie staining.

Two bands were absent from the supernatant in the $\Delta mgrB$ strain (Fig. 3.3a). Based on the molecular weight of these bands, they could be attributed to the two effectors SipA (74kDa) and SipC (43kDa). This was tested via western blotting: in both cases, the two proteins were less abundant in the supernatant of the $\Delta mgrB$ mutant, with SipC being also lower intracellularly (Fig. 3.3b). A more sensitive analysis carried out, for example, via mass spectrometry, can indicate if this regulation is common to other secreted effectors.

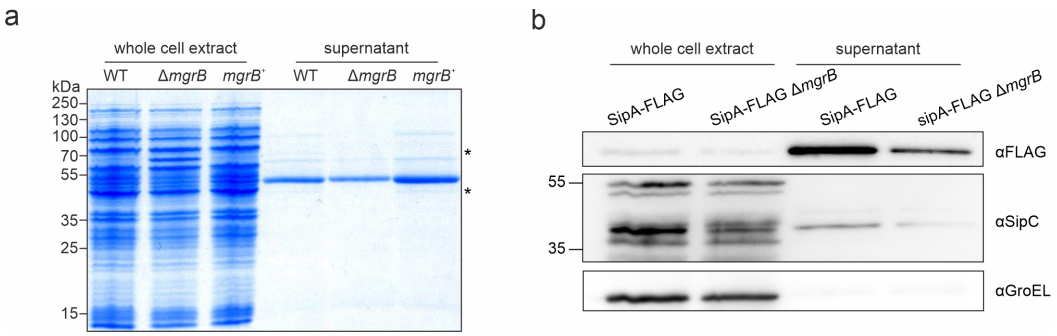



**Figure 3.3: MgrB affects the production of effectors in SPI-1-grown *Salmonella*.** **a** The supernatant of wild-type, Δ*mgrB*, and *mgrB*$^{+}$ *Salmonella* grown in SPI-1-inducing conditions was visualized via SDS-PAGE and Coomassie staining. The wild-type and Δ*mgrB* strains carry the empty vector control. The asterisks mark the bands that change in intensity in the strains. **b** *Salmonella* with a chromosomal FLAG-tagged version of SipA with or without *mgrB* were grown and analysed similar to panel **a**. Here, SDS-PAGE was followed by western blot analysis to visualize SipA-FLAG or SipC. GroEL was probed for as a loading control.

The lower amount of effector proteins can explain the marked invasion defect of epithelial cells of *Salmonella* Δ*mgrB* (Fig. 3.2a). Overall, these data indicate MgrB as a regulator of pathogenesis, as suggested from the integration of dual RNA-seq and TraDIS data (Section 2.2), through an unknown mechanism that might rely on PhoPQ.

## 3.2 The impact of MgrB on the transcriptome of *Salmonella* grown in infection-relevant conditions

### 3.2.1 MgrB affects the transcriptome of *Salmonella* grown in SPI-2-, but not SPI-1-inducing conditions

To shed light on the molecular mechanism behind the infection phenotype observed, the transcriptomes of *Salmonella* wild-type and Δ*mgrB* were analysed. The two strains were grown in both SPI-1- and SPI-2-inducing conditions, and total RNA was isolated and sequenced.

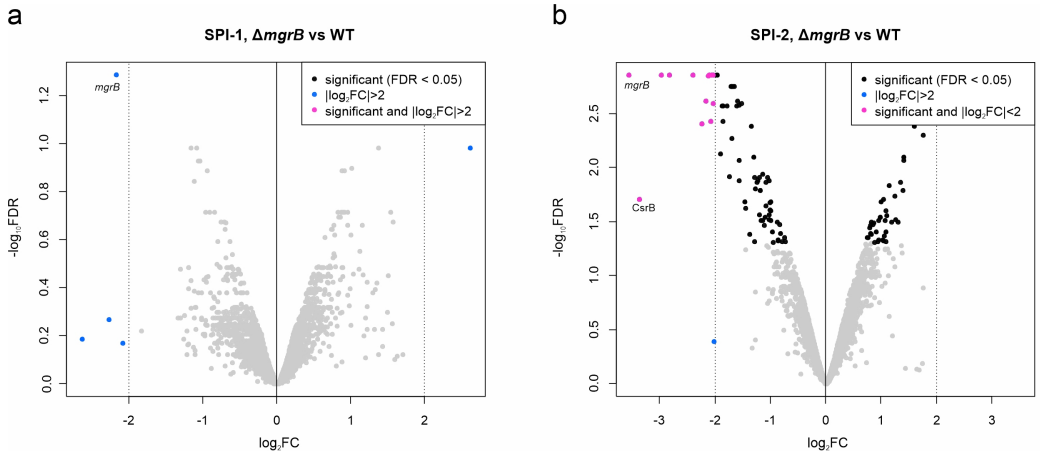



**Figure 3.4: Changes in the transcriptome of SPI-1- and SPI-2-grown *Salmonella* in response to MgrB.** Total RNA from wild-type and Δ*mgrB Salmonella* grown in **a** SPI-1- or **b** SPI-2-inducing conditions was analysed. The volcano plots indicate difference in abundance of single transcripts in Δ*mgrB* vs wild-type *Salmonella* (two biological replicates were collected). Changes in abundance ($\log_2$FC, x-axis) were considered significant with an FDR $< 0.05$.

While no major dysregulation occurred in SPI-1 condition (Fig. 3.4a), twelve genes were downregulated in the absence of *mgrB* in SPI-2 medium ($\log_2$FC $<$ -2 and FDR $< 0.05$; Fig. 3.4b). Several of these mRNAs belong to motility and chemotaxis pathways (Table 3.1). A notable exception in terms of biological function is the sRNA CsrB, which is involved together with the sRNA CsrC in regulating the global RNA-binding protein (RBP) CsrA by sequestering it through its several CsrA binding sites (Vakulskas et al., 2015).

**Table 3.1:** Genes dysregulated ($|\log_2 FC| > 2$ and FDR $< 0.05$) in $\Delta mgrB$ vs wild-type *Salmonella* grown in SPI-2 medium.

| **Gene name** | **$\log_2$FC** | **q-value** |
|---|---|---|
| CsrB | -3.36 | 0.019 |
| *flgB* | -2.97 | 0.001 |
| FUR_125 | -2.83 | 0.001 |
| *tar* | -2.41 | 0.001 |
| *flgC* | -2.25 | 0.003 |
| *flgD* | -2.17 | 0.002 |
| *cheR* | -2.13 | 0.001 |
| *motB* | -2.12 | 0.001 |
| *fliC* | -2.08 | 0.003 |
| *fliD* | -2.07 | 0.001 |
| *fliS* | -2.04 | 0.001 |
| STnc3600 | -2.03 | 0.002 |

The RNA-seq data gathered for *Salmonella* grown in SPI-2 medium were independently validated via northern blotting. As suggested by RNA-seq analysis, *fliC* was downregulated in the $\Delta mgrB$ strain, an effect that was overcomplemented in the $mgrB^+$ strain (Fig. 3.5a). CsrB levels were also lower in the $\Delta mgrB$ strain as compared to the wild-type strain; on the other hand, the levels of CsrC were not as affected by the absence of MgrB, indicating a specific effect on the regulation of only one of these two closely-related sRNAs (Fig. 3.5a). The sRNA MicA was also probed due to its inverse correlation to that of the genes considered so far ($\log_2$FC = 1.41 and q-value = 0.008). In this case, northern blot analysis did not show any major change in the level of MicA, in conflict with RNA-seq data (Fig. 3.5a). Further analysis showed that the half-lives of both CsrB and CsrC are longer when *mgrB* is expressed at higher levels in the complementation strain compared to its level in the wild-type (Fig. 3.5b, c). On the contrary, MicA had an extended half-life in the $\Delta mgrB$ strain compared to the wild-type or $mgrB^+$ strains (Fig. 3.5d), suggesting that MgrB has an impact on the stability of MicA that is overcome by a faster turnover, since the alteration of its steady-state level indicated by RNA-seq were not validated via northern blotting.

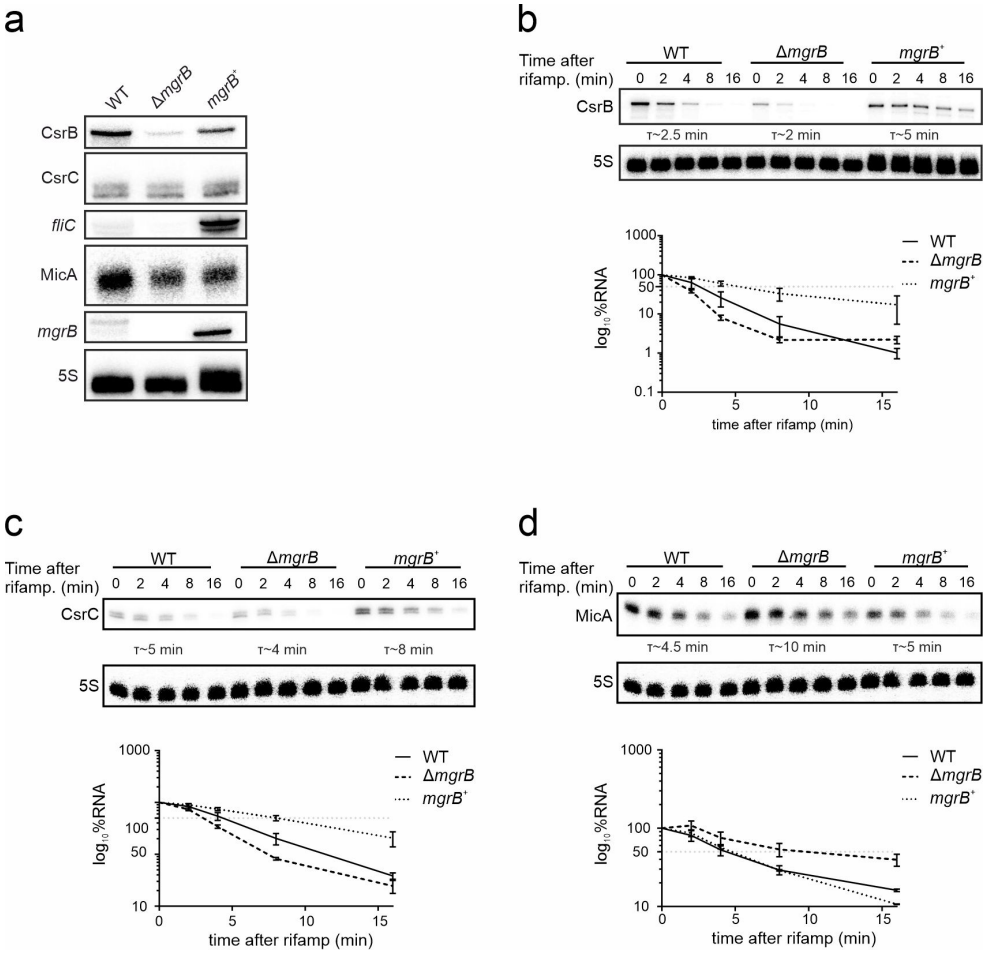



**Figure 3.5: Validation of RNA-seq analysis. a** Northern blot validation of RNA-seq data. 5S RNA was probed for as a loading control. **b**, **c**, **d** Northern blot of rifampicin assays of wild-type, Δ*mgrB*, and *mgrB*$^{+}$ *Salmonella* grown in SPI-2 medium. The plots below the blots show averages of three biological replicates, with bars for standard errors, of the $\log_{10}$RNA levels after addition of rifampicin (t=0).

Since MgrB is conserved between *E. coli* and *Salmonella*, it was tested if the *E. coli* protein could complement the regulation of CsrB in *Salmonella*. When expressed in the Δ*mgrB Salmonella* background from a plasmid under a constitutive promoter, *E. coli mgrB* could stabilize CsrB, supporting a regulation that is conserved between the two species (Fig. 3.6). On the other hand, MicA was downregulated in the presence of *Salmonella mgrB* but not in the presence of *E. coli mgrB*.

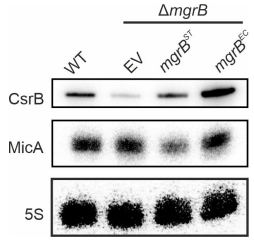



**Figure 3.6: MgrB from *E. coli* can complement CsrB and MicA regulation by *Salmonella* MgrB.** The levels of CsrB, MicA, and 5S RNA as a loading control, were analysed in the wild-type and Δ*mgrB* strains carrying the empty vector (EV) control, as well as in the Δ*mgrB* plasmid carrying either *Salmonella* (ST) or *E. coli* (EC) *mgrB* expressed from a constitutive promoter.

So far, the RNA-seq data indicate that MgrB has an impact on the gene expression of *Salmonella* grown in SPI-2- but not SPI-1-inducing conditions. In particular, MgrB has a positive effect on the expression of motility and chemotaxis genes, as well as the sRNA CsrB. None of these genes can account for the strong impairment at infection of the Δ*mgrB* strain. Therefore, further work is necessary to elucidate the underlying mechanism.

## 3.3 The impact of MgrB on the proteome of *Salmonella* grown in an infection-relevant condition

In bacteria, protein levels do not correlate with the level of the corresponding mRNA (Taniguchi et al., 2010). Therefore, independent proteome analysis can shine light on possible (post-)transcriptional regulation in response to external stimuli or the absence of a given gene. To evaluate the impact of MgrB at the proteome level, *Salmonella* wild-type, Δ*mgrB* (both strains carrying the empty vector control), and *mgrB*$^+$ were analysed via mass spectrometry. For this, cells were grown in SPI-2 medium to an $OD_{600}$ of 0.3 and samples were collected from three biological replicates. A total of 58 proteins were dysregulated in the Δ*mgrB* strain. In particular, 31 proteins were more abundant in the mutant than in the wild-type strain, while 27 were less abundant in the mutant than the wild-type. These 58 proteins are significantly (p-value $< 0.05$) dysregulated in the deletion

mutant compared to the wild-type. In the complementation strain, instead, their level was comparable to that in the wild-type strain, or over-complemented. This effect is likely due to the higher levels of *mgrB* in the complementation strain than in wild-type *Salmonella* (Fig. 3.5a). The strain lacking *mgrB* had lower levels of FliC and SopB than *Salmonella* wild-type, which was independently tested via western blotting. The regulation was confirmed by the over-complementation on the level of the two proteins in the $mgrB^+$ strain (Fig. 3.7a).

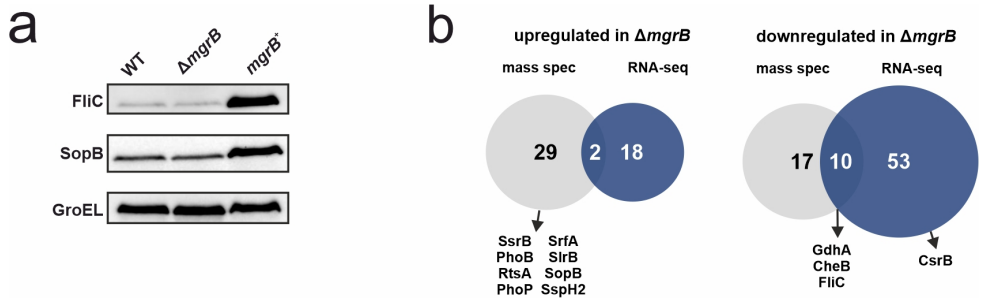



**Figure 3.7: Validation of mass spectrometry data, and their overlap with RNA-seq data. a** Mass spectrometry data were validated via western blot. For this, wild-type, Δ*mgrB* (both carrying the empty vector control), and $mgrB^+$ *Salmonella* were grown in SPI-2 medium. FliC and SopB were probed to validate the stabilization from MgrB, while GroEL was probed as loading control. **b** Venn diagrams showing the overlap between genes dysregulated in Δ*mgrB* vs wild-type *Salmonella* in mass spectrometry and RNA-seq. Proteins and sRNA of interest from each group are indicated.

The proteomics data were cross-referenced with RNA-seq data. This showed only partial overlap, mostly in the downregulation of motility and chemotaxis proteins (FliC and CheB, Fig. 3.7b). Interestingly, members of different TCSs were among the proteins upregulated in Δ*mgrB*: SsrB, PhoB, and PhoP itself. This may be in line with the recent finding that MgrB can interact with sensor kinases other than PhoQ itself (Yadavalli et al., 2020). Altogether, this supports the hypothesis that small proteins are highly versatile in their targetome, having the potential to regulate multiple partners.

### 3.3.1 MgrB is required for *Salmonella* motility

Since MgrB stabilized the expression of several motility-related genes, the ability of the Δ*mgrB* strain to swim was compared to that of wild-type *Salmonella*.

As expected, the mutant was impaired, a phenotype that was restored in the complementation strain (Fig. 3.8). This phenotype reflects the widespread down-regulation of motility-related genes shown by RNA-seq (Table 3.1).

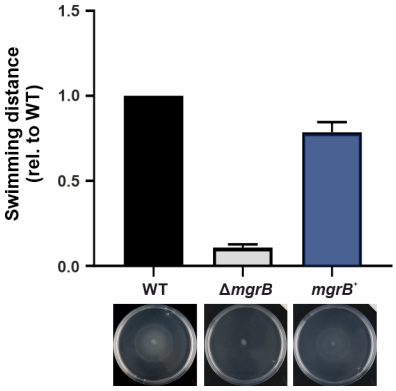



**Figure 3.8: *Salmonella* is unable to swim without MgrB.** *Salmonella* wild-type, Δ*mgrB* (both strains carrying the empty vector control), and *mgrB*$^{+}$ were spotted at the center of a plate with 0.3% agar. The swimming distance from the site of inoculation was measured and calculated relative to that of the wild-type strain (n=3, error bars indicate the standard error). A representative image of each strain is shown below the respective bar.

## 3.4 Discussion

The PhoPQ TCS is well known to regulate bacterial response to environmental challenges such as low pH, low magnesium, or antimicrobial peptides (Groisman, 2001). It can directly and indirectly activate a large set of genes. It is therefore reasonable that bacteria have evolved a set of accessory regulators to avoid its unnecessary or prolonged activation, which would be energetically expensive and detrimental.

Two small proteins in *E. coli* have been identified with such a role: SafA and MgrB (Eguchi et al., 2012; Lippa and Goulian, 2009). Of these, only MgrB is conserved in *Salmonella* (Lippa and Goulian, 2009). Investigation of the mechanism of MgrB activity suggests that, by localizing to the membrane, it inhibits PhoQ auto-phosphorylation, which results in lower amounts of PhoP in its active form (Salazar et al., 2016). This knowledge, together with the observation that disruption of MgrB results in lower *Salmonella* infection (Fig. 2.6) placed MgrB as a top candidate for functional characterization.

TraDIS data were independently confirmed with a clean deletion mutant of the MgrB CDS. This revealed a defect in infection of both epithelial and phagocytic cells (Fig. 3.2). The lower levels of effector proteins in the Δ*mgrB* mutant (Fig. 3.3) can account for the defect at invading epithelial cells, which is an active mechanism, unlike the passive uptake by macrophages. It is possible that there is a greater effect of *mgrB* deletion in SPI-1 conditions than reported here, which would be shown by mass spectrometry analysis of the entire proteome.

Interestingly, the lower abundance of proteins in the secreted fraction was not accompanied by a dysregulation at the mRNA level from *Salmonella* grown in SPI-1-inducing conditions (Fig. 3.4a). This suggests a post-transcriptional effect that might be uncoupled from the activity of PhoPQ. Since MgrB is an inner membrane protein (Lippa and Goulian, 2009), it might bind and regulate the SPI-1 T3SS, an hypothesis that should be tested with interactome studies to complement mass spectrometry data. For this, a tagged version of MgrB should be tested to exclude loss of functionality with a readout such as virulence or regulation of one of the targets identified in this work. If the tagged strain behaved like wild-type *Salmonella*, this could be used for co-immunoprecipitation and mass spectrometry analysis. This hypothesis could also be addressed with a targeted approach such as a bacterial two-hybrid assay testing selected members of the SPI-1 T3SS. Alternatively, MgrB could regulate the chaperone-mediated secretion of SPI-1 effectors.

The RNA-seq data derived from cells grown in SPI-2 medium showed a marked downregulation of the sRNA CsrB and motility genes (Fig. 3.4b, 3.5). The latter was reflected in a swimming defect of Δ*mgrB Salmonella* (Fig. 3.8), as well as in proteomic data (Fig. 3.7). The requirement of flagella for *Salmonella* infection is well established (Carsiotis et al., 1984; Weinstein et al., 1984). This goes along with the infection defect of Δ*mgrB Salmonella* in RAW264.7 macrophages. In particular, non flagellated *Salmonella* were shown to have a poor survival rate inside mouse-derived macrophages, but to be taken up by macrophages as efficiently as the wild-type (Carsiotis et al., 1984). This, together with the lower escape rate of non-flagellated cells from dying macrophages (Sano et al., 2007), points to flagella as the major intermediary between MgrB and the infection outcome.

Mass spectrometry analysis also highlighted interesting and unexpected proteins with increased levels in the absence of MgrB. These were transcription regulators of three TCSs: PhoP from PhoPQ, SsrB from SsrAB, and PhoB from PhoBR. All these TCSs are known to positively regulate their own expression (Shin et al., 2006; Feng et al., 2004; Gao and Stock, 2018). However, their levels were not altered in the RNA-seq data of $\Delta$*mgrB Salmonella* grown in SPI-2 medium (*phoP*: $\log_2$FC = 0.02; *ssrB*: $\log_2$FC = 0.45; *phoB*: $\log_2$FC = 0.15).

Overall, this suggests that MgrB might interfere with protein abundance in a post-transcriptional or post-translational manner, and is supported by the observation that, in *E. coli*, MgrB can interact with other sensor kinases (Yadavalli et al., 2020). Furthermore, lack of MgrB did not broadly affect the known targetome of PhoPQ, which is an additional hint at MgrB having other, unknown targets in the bacterial cell.

# 4 YjiS

The small protein YjiS (54 aa) is so far uncharacterized in *Salmonella*, although other proteins containing the domain of unknown function (DUF) 1127 have been recently studied in $\alpha$-proteobacteria (Kraus et al., 2020; Grützner et al., 2021). In *Salmonella*, its transcription is highly induced during infection (Fig. 2.4), an observation that is supported by other datasets (Colgan et al., 2016; Srikumar et al., 2015). Furthermore, its interruption rendered *Salmonella* more virulent (Fig. 2.6). Overall, this makes YjiS an ideal candidate as a novel regulator of infection.

## 4.1 Regulation of *yjiS* expression

To place *yjiS* in a regulatory context, transcriptomic data from a set of *Salmonella* mutants lacking different transcriptional regulators was interrogated (Colgan et al., 2016). Overall, *yjiS* was less abundant in strains lacking SPI-2 regulators (*hilD*, *phoPQ*, *ssrAB*), compared to the wild-type (Colgan et al., 2016). Therefore, no single transcriptional regulator could be confidently identified as a factor influencing *yjiS* expression. The RBP ProQ was recently shown to regulate the expression of genes involved in infection (Westermann et al., 2019). Considering this, the level of *yjiS* was analysed in *proQ* deletion ($\Delta$*proQ*) and complementation (*proQ*$^{+}$) mutants. This indicated that ProQ has a specific and positive influence on the level of *yjiS* (Fig. 4.1a). By contrast, the global RBP Hfq did not affect the level of *yjiS* mRNA (Fig. 4.1a).

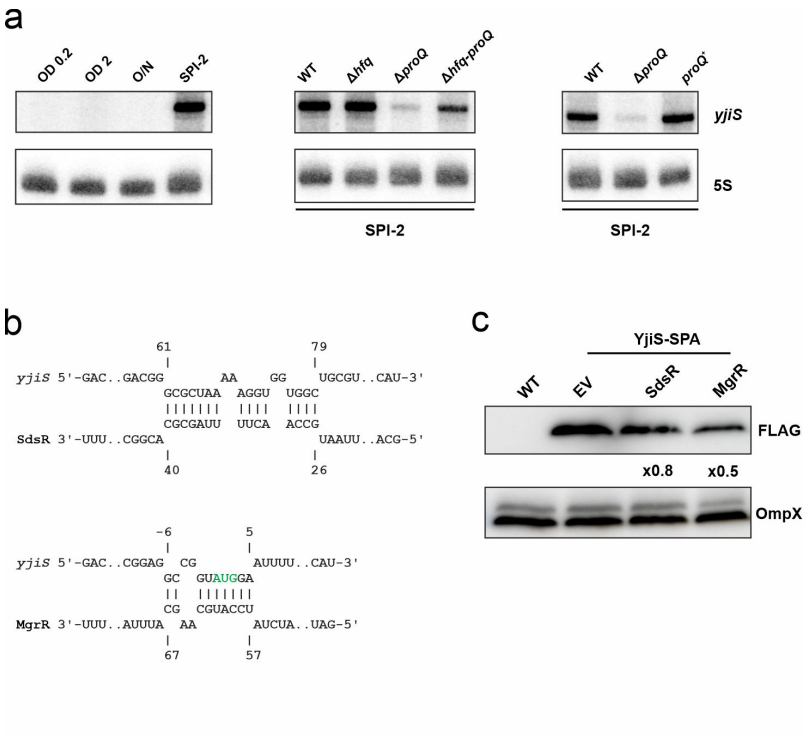



**Figure 4.1: Expression of *yjiS* is regulated by ProQ and MgrR. a** Northern blot analysis of RNA collected from *Salmonella* wild-type (left panel) grown in LB medium to an $OD_{600}$ of 0.2 or 2.0, over-night (ON), or in SPI-2 medium. Since *yjiS* is expressed in SPI-2 condition, its level was also analysed in SPI-2 medium in the absence of the RBPs Hfq and ProQ ($\Delta hfq$, $\Delta proQ$, $\Delta hfq\Delta proQ$). Lack of *proQ* was complemented in the $proQ^+$ strain expressing *proQ* from a plasmid, and *yjiS* levels compared to wild-type or $\Delta proQ$ carrying the empty vector control. 5S RNA was probed for as a loading control. **b** Predicted interaction between *yjiS* and SdsR or MgrR (the numbers indicate the distance from the TSS in the sRNA, and the distance from the start codon in *yjiS*). In the interaction with MgrR, the start codon of *yjiS* is in green. **c** Western blot analysis of the levels of YjiS, tagged with a SPA tag at the C-terminus and detected with the FLAG antibody, in response to SdsR or MgrR overexpression. The wild-type strain was included as a negative control, and OmpX was probed for as a loading control. The numbers below the upper blot indicate the fold-change in YjiS levels relative to the empty vector control (EV), and are the average of three biological replicate.

RIL-seq (RNA interaction by ligation and sequencing) is a method for the global identification of the interactions between sRNAs and their targets that a given RBP mediates (Melamed et al., 2016). An Hfq-RIL-seq dataset generated from *Salmonella* grown in SPI-2 medium (Matera *et al.*, unpublished) highlighted two sRNAs possibly binding and regulating *yjiS*: MgrR and SdsR. The putative binding sites on *yjiS* engage the known seed sequences of the sRNAs (Fig. 4.1b, (Melamed et al., 2016; Fröhlich et al., 2016)). To test their influence on YjiS, the

small protein was chromosomally tagged at the C-terminus with a SPA tag, and coupled with a plasmid-borne overexpression of either MgrR or SdsR. This showed a mild repression by SdsR and a stronger one by MgrR (Fig. 4.1c). MgrR is known to be destabilized by the absence of *hfq* (Moon and Gottesman, 2009). Lack of de-repression of *yjiS* in the Δ*hfq* (Fig. 4.1a) could be explained by a regulation of MgrR on *yjiS* translation, given that it binds the region around its start codon (Fig. 4.1b).

Overall, *yjiS* expression pattern resembles that of SPI-2-encoded genes. Its mRNA levels are stabilized by ProQ, and its protein levels are repressed by MgrR. MgrR was shown to bind Hfq but not ProQ (Holmqvist et al., 2018), suggesting that the two regulations are independent from each other.

## 4.2 YjiS is associated with the inner membrane

The strain with SPA-tagged YjiS was used to investigate the subcellular localization of YjiS. Sub-fractionation of cells grown in SPI-2 medium was performed, separating the cytosol from the two membranes. This showed that YjiS localizes to the inner membrane (Fig. 4.2a).

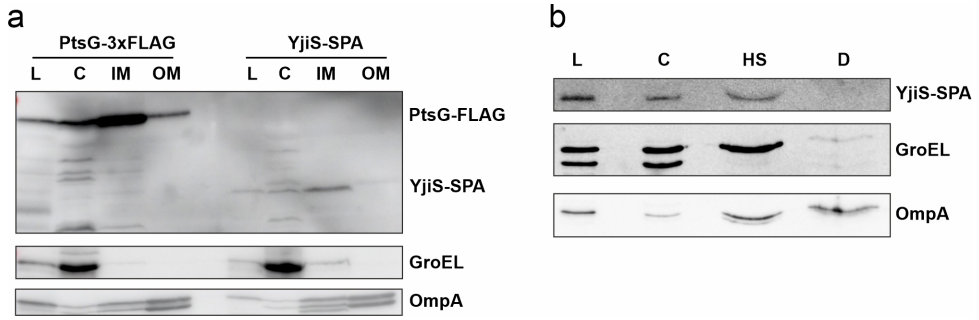



**Figure 4.2: YjiS associates with the inner membrane. a** YjiS-SPA tag or PtsG-3xFLAG *Salmonella* (the PtsG-3xFLAG strain was taken from (Papenfort et al., 2012)) were grown in LB medium to an $OD_{600}$ of 0.4 (PtsG) or SPI-2-inducing conditions (YjiS), collected, and subfractionated. The content of each fraction (lysate (L), cytosol (C), inner membrane (IM), outer membrane (OM)) was analysed via western blotting. PtsG and YjiS could be detected with an $\alpha$-FLAG antibody, with PtsG serving as a control for the IM fraction. **b** *Salmonella* expressing YjiS-SPA tag was grown in SPI-2-inducing conditions and separated into cytosolic (C) and membrane fraction. The latter was first treated with a high-salt (HS) buffer to detach membrane-bound proteins, and subsequently with a detergent (D) to solubilize membrane-inserted protein. Each fraction, alongside with the lysate (L), was analysed via western blotting. GroEL and OmpA were probed for as controls of the C and membranes fractions, respectively.

Since YjiS lacks any known membrane localization sequence, it is a possibility that it interacts with another inner-membrane protein rather than being itself inserted in it. To investigate this, another cell fractionation experiment was performed. Unlike previously, the total membrane fraction was not separated into inner and outer membrane, but was first treated with a high-salt buffer to detach proteins not inserted in the membrane. Once these were recovered, the remaining sample was treated with a detergent to solubilize proteins that are inserted in the membrane. With this approach, YjiS was detected only in the high-salt fraction, indicating that it is not itself inserted in the membrane (Fig. 4.2b).

## 4.3 YjiS as a virulence suppressor

### 4.3.1 The absence of YjiS makes *Salmonella* more virulent in macrophage infection

One of the reasons for focusing on YjiS as a virulence regulator was that its disruption by transposon insertion made *Salmonella* more virulent (Fig. 2.6b). To

independently validate the TraDIS data, a deletion mutant lacking the YjiS CDS was generated ($\Delta yjiS$) and compared to the wild-type strain in infection. The bacterial load of the two strains was quantified at different time points after infection of RAW264.7 macrophages in a gentamicin protection assay. This confirmed the TraDIS data, with the $\Delta yjiS$ being more than twice as abundant as *Salmonella* wild-type at 20h p.i. (Fig. 4.3a).

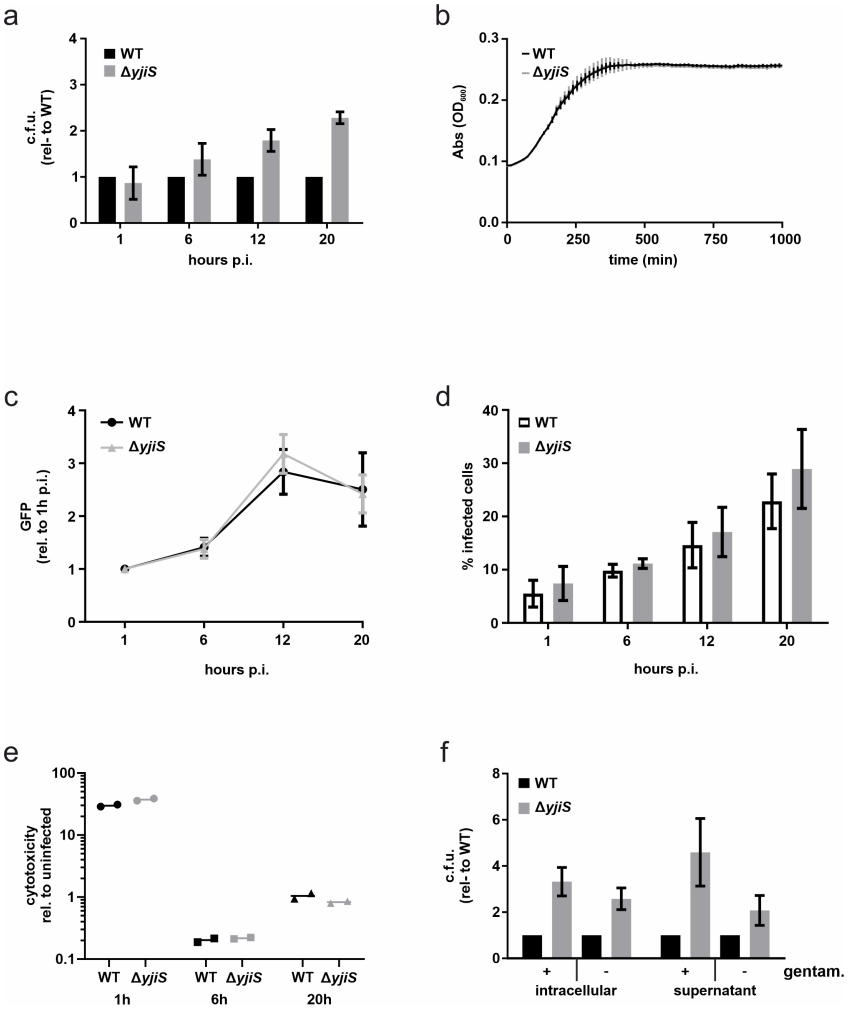



**Figure 4.3: YjiS is a virulence suppressor that prevents bacterial escape from macrophages.** **a** C.f.u. assay of wild-type and Δ*yjiS Salmonella* infecting RAW264.7 macrophages. Data were collected from three biological replicates and the error bars indicate the standard error from the mean. **b** *In vitro* growth of wild-type and Δ*yjiS Salmonella* in SPI-2 medium. The error bars indicate the standard error from the mean of data collected from three biological replicates. **c** Amount of intracellular bacteria per single infected RAW264.7 macrophage. The bacteria were quantified via flow cytometry analysis using GFP-expressing wild-type or Δ*yjiS Salmonella*. Data were collected from three biological replicates and the error bars indicate the standard error from the mean. **d** The percentage of infected RAW264.7 macrophages was quantified using GFP-expressing wild-type or Δ*yjiS Salmonella*. The error bars indicate the standard error from the mean of data collected from three biological replicates. **e** Quantification of LDH in the supernatant of RAW264.7 infection with wild-type or Δ*yjiS Salmonella*. The points are from the two biological replicates performed. **f** C.f.u. assay of RAW264.7 macrophages infected with either wild-type or Δ*yjiS Salmonella*. The infection was performed in the presence (+) or absence (-) of gentamicin in the supernatant. Both intracellular bacteria and bacteria in the supernatant were collected. Data were collected from three biological replicates and the error bars indicate the standard error from the mean.

There might be several explanations for this phenotype, namely the Δ*yjiS* mutant is i) hyper-replicating; ii) inducing less cytotoxicity in the host; iii) able to escape and re-invade host cells; or iv) a combination of the previous possibilities.

Hyper-replication was tested by evaluating the growth of wild-type and Δ*yjiS* *Salmonella* in SPI-2 medium, which mimics the host intracellular environment. This showed no difference in growth (Fig. 4.3b). Similarly, the amount of bacteria per single infected host cell was quantified using $GFP^+$ strains and flow cytometry. Also this approach did not highlight marked differences between the strains (Fig. 4.3c). Despite this, the overall population of infected cells was higher for the Δ*yjiS* mutant rather than for the wild-type Δ*yjiS* at 20 hours p.i. (Fig. 4.3d). This could be due to a lower rate of host cell death induced by the absence of YjiS. However, a cytotoxicity assay testing the amount of lactate dehydrogenase (LDH) in the infection medium as a readout of cell death showed no difference in the death rate of host cells (Fig. 4.3e). This lead to the hypothesis that YjiS might prevent *Salmonella* escape from the host and re-infection of the bystanders. To test this, bacteria present in the infection supernatant at 20 hours p.i. were quantified with a c.f.u. assay. The amount of Δ*yjiS* *Salmonella* recovered was four times higher than the wild-type *Salmonella* (Fig. 4.3e). This assay is usually carried out in the presence of gentamicin, a bacteriostatic antibiotic, to limit the amount of bacteria in the supernatant. To evaluate the impact of gentamicin, the level of extracellular bacteria was also quantified in an infection assay in the absence of the antibiotic. In this case as well, YjiS appeared to prevent *Salmonella* escape (Fig. 4.3e).

Overall, these observations suggest that YjiS has a role in preventing virulence that limits *Salmonella* exit from the host cell and re-infection of the neighbouring ones.

### 4.3.2 Requirement of conserved residues for YjiS activity

The infection phenotype is an ideal readout for the requirement of selected residues within the YjiS amino acid sequence. Given the high prevalence of arginines in the DUF1127 annotated in YjiS, these were selected for an arginine to alanine mutation. Furthermore, the "DIGL" motif is also highly conserved (Fig. 4.4a). Based on this, a library of alanine mutants was constructed on a medium copy

plasmid encoding *yjiS* with its native promoter and terminator. Each mutated version of the plasmid was transformed in the Δ*yjiS* background and compared to a strain with the wild-type *yjiS* plasmid in a gentamicin protection assay. Given that the phenotype for the absence of YjiS was observed at 20h p.i., this time point was chosen to test the mutants. This indicated that the expression of *yjiS* from a plasmid can complement the phenotype of the Δ*yjiS* strain (Fig. 4.4b), proving that the hypervirulence is a direct phenotype of YjiS.

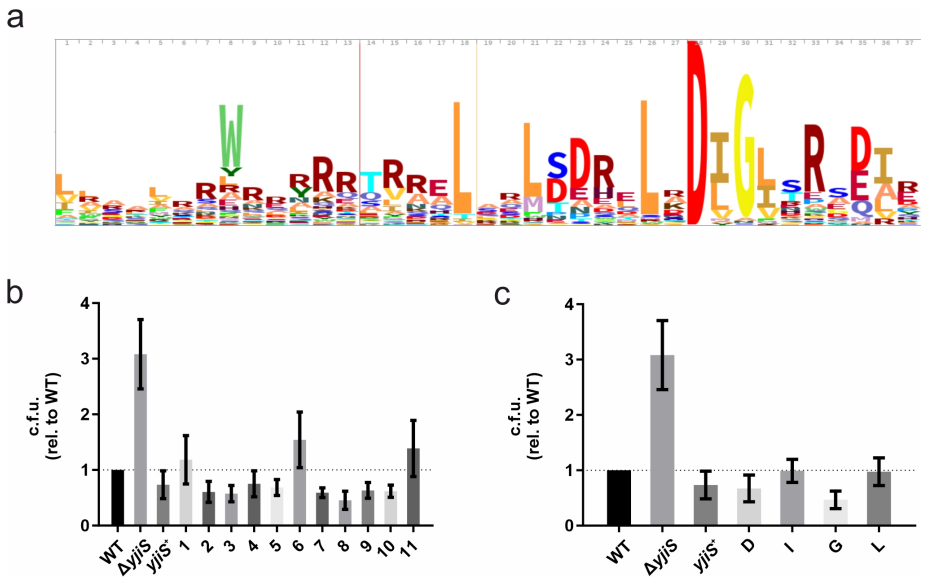



**Figure 4.4: Alanine scanning of conserved residues of YjiS. a** Amino acid frequency in the residues of the DUF1127 taken from Pfam (Mistry et al., 2021). The size of each letter is directly proportional to its frequency. **b** Infection assay of RAW264.7 macrophages using wild-type and Δ*yjiS* *Salmonella* with the empty vector control, Δ*yjiS* with a complementation plasmid with wild-type *yjiS* (*yjiS*$^{+}$), or with an arginine to alanine mutation of each of the 11 arginines in YjiS. **c** C.f.u. assay of RAW264.7 macrophage infected with wild-type, Δ*yjiS*, or *yjiS*$^{+}$ as in the previous planel, or alanine mutants of the DIGL motif. In panels **b** and **c** data were collected at 20 hours p.i. from three biological replicates. The error bars indicate the standard error from the mean.

The majority of the arginine mutants over-complemented the hyper-virulence phenotype (Fig. 4.4b). Only mutants of the 6$^{\text{th}}$ and 11$^{\text{th}}$ arginines did not fully complement the lack of *yjiS*, although they did not reach the level of the Δ*yjiS* strain (Fig. 4.4b). Similarly to the majority of the arginine mutants, none of the DIGL mutants failed to complement the deletion (Fig. 4.4c). This indicates that no single arginine or DIGL residue is responsible for YjiS activity.

## 4.4 Investigating the molecular mechanism of YjiS activity

### 4.4.1 YjiS does not majorly impact the transcriptome of *Salmonella*

To assess the impact of YjiS on the transcriptome of *Salmonella*, total RNA from wild-type and Δ*yjiS* strains grown in SPI-2 medium was collected and sequenced. Differential analysis indicated that eight genes passed the statistical filtering of $|\log_2 FC| > 2$ and FDR $< 0.05$ (Fig. 4.5). The upregulated transcripts belong to motility-related genes, suggesting a destabilizing effect of the small proteins on their half-life or turnover. However, attempts at validating these data via transcriptional reporters or motility assays were negative (Fig. 4.5b, c).

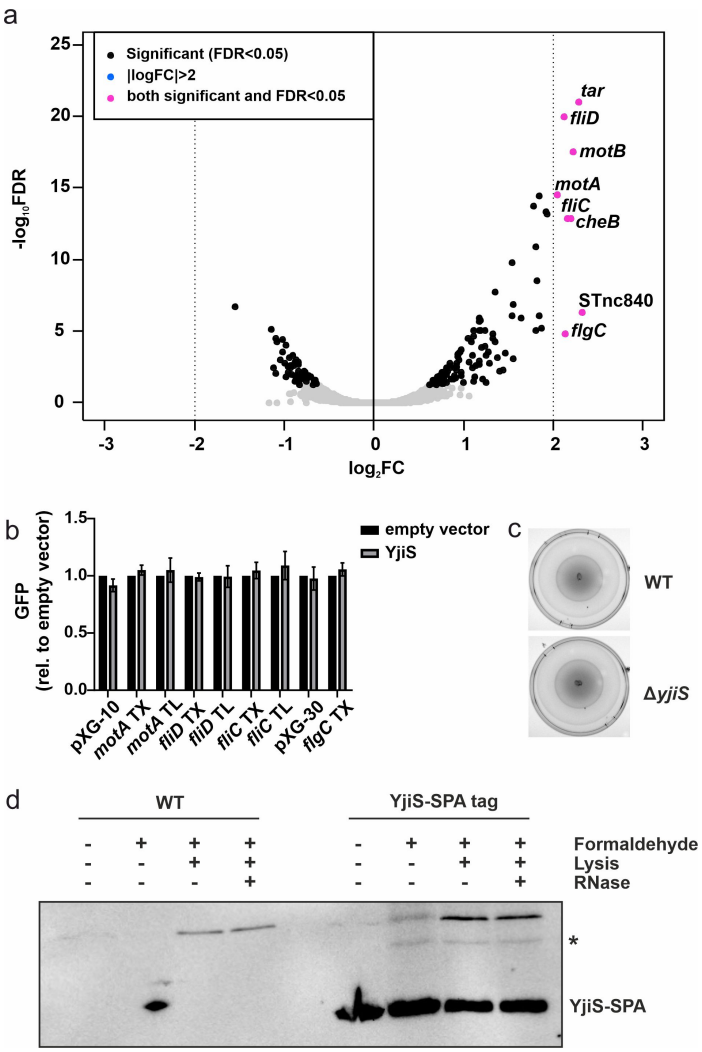



**Figure 4.5: Changes in the transcriptome of *Salmonella* grown in SPI-2-inducing conditions in response to YjiS. a** Total RNA from wild-type and Δ*yjiS Salmonella* grown in SPI-2-inducing conditions was analysed. The volcano plot indicates difference in abundance of single transcripts in Δ*yjiS* vs wild-type *Salmonella* (two biological replicates were collected). The change in abundance ($\log_2$FC, x-axis) was considered significant with an FDR $< 0.05$. **b** Analysis of the changes in the fluorescence of transcriptional (TX) and translational (TL) reporters of transcripts dysregulated in **a**. The reporters were transformed in Δ*yjiS Salmonella* and paired either with the empty vector or with a vector overexpressing *yjiS*. The GFP signal was calculated relative to the empty vector control. pXG-10 and pXG-30 were included to evaluate unspecific changes in the empty fluorescence reporters. Data were collected from three biological replicates and the error bars indicate the standard error from the mean. **c** *Salmonella* wild-type and Δ*yjiS* were spotted at the center of a plate with 0.3% agar to test differences in motility. The image is representative of three biological replicates. **d** *Salmonella* wild-type or YjiS-SPA tag were treated with formaldehyde, lysed with mechanical beating, or treated with RNase when indicated (+). The samples were analysed via western blotting and YjiS-SPA tag was detected with the FLAG antibody, and the wild-type strain was included as a background control.

The overall positive charge of YjiS suggested an interaction with nucleic acids. Furthermore, other members of the DUF1127-containing small proteins family have been shown to bind RNA, although in $\alpha$-proteobacteria (Kraus et al., 2020; Grützner et al., 2021). Since sequence analysis showed that YjiS from *Salmonella* is not a homologue of these proteins (Kraus et al., 2020), it is unclear if YjiS binds RNA also in *Salmonella*. RNA binding was tested crosslinking cell lysates of *Salmonella* wild-type and SPA-tagged YjiS with formaldehyde. Subsequent differential treatment with RNase should indicate if any shift observed is due to binding of RNA or other proteins. Indeed, crosslinking caused a shift that was maintained also after treatment with RNase, thus excluding RNA as a binding partner (Fig. 4.5d).

### 4.4.2 Mass-spectrometry analysis of the interactome of YjiS

Given that YjiS does not seem to bind RNA, the possibility of it binding another protein was tested via co-immunoprecipitation (Co-IP; Fig. 4.6a) followed by mass spectrometry. This was performed with the SPA-tagged YjiS strain as well as the wild-type as background control. Two replicates were analysed, yielding seven high-confidence partners recovered in both replicates of the YjiS-SPA tag strain but not in the wild-type, as well as other proteins enriched relative to the wild-type (Fig. 4.6b). Overall, these included several known inner membrane proteins: FtsN, PspA, and YdiJ (exclusively detected in the YjiS-SPA tag strain), as well as SsrA, MsbA, Trg, PntB, and SL1344_3112 (enriched relative to the wild-type).

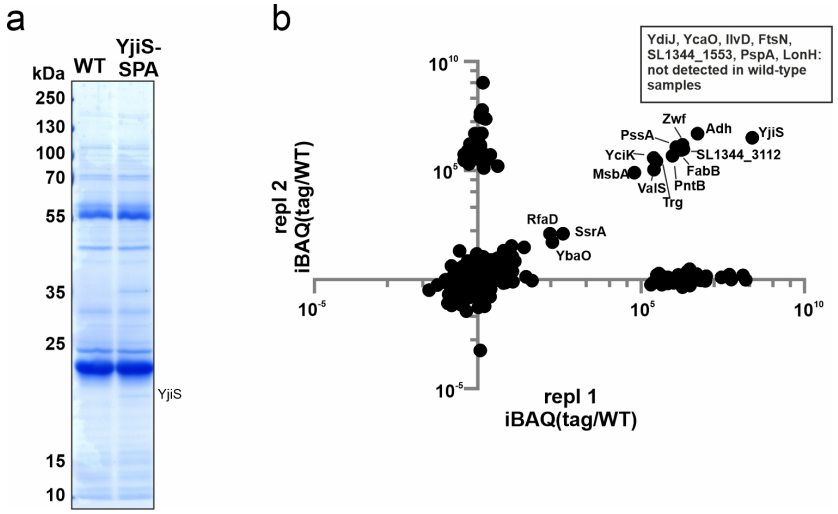



**Figure 4.6: Candidate interaction partners of YjiS. a** Coomassie-stained SDS-PAGE of the CoIP performed on wild-type and YjiS-SPA tag *Salmonella* grown in SPI-2 medium. The image is representative of three biological replicates. **b** Graph showing the proteins detected via mass spectrometry of the CoIP performed on *Salmonella* wild-type or SPA-tagged YjiS (n=2). Each axis represents the abundance (iBAQ) in either replicate of the proteins detected in the YjiS-SPA tag strain relative to the wild-type. The proteins in the rectangle at the top right are the ones detected only in the YjiS-SPA tag CoIP but not in the wild-type control.

Among the putative interactors of YjiS, only SsrA, sensor kinase of the SsrAB TCS, has a known role in virulence (Pérez-Morales et al., 2017). Of note, the level of *yjiS* was decreased in *Salmonella* mutants lacking *ssrA*, *ssrB*, or both as shown by the SalCom Regulon database (Colgan et al., 2016), suggesting that YjiS might be a target of SsrAB involved in a regulatory loop with the TCS. However, further work is required to investigate the outcome of this and the other interactions.

## 4.5 Discussion

The expression pattern of *yjiS* resembles that of SPI-2-encoded genes (Fig. 2.4), as already observed in a global analysis of *Salmonella* transcriptome during infection (Hautefort et al., 2008). Given that high expression during infection is a good indication of a virulence-related function (Hébrard et al., 2011), and that lack of YjiS promotes *Salmonella* virulence (Fig. 2.6), YjiS was chosen as a high priority candidate for functional characterization.

Independent validation of the TraDIS data confirmed that lack of YjiS makes

*Salmonella* more virulent when infecting mouse macrophages (Fig. 4.3a). Different mechanisms possibly causing this phenotype were tested. In this way, hyper-replication or different induction of host cell death in the absence of *yjiS* were excluded as underlying mechanisms (Fig. 4.3b-e). Instead, the quantification of extracellular bacteria indicated that YjiS prevents *Salmonella* escape from the host cells and re-infection of bystanders (Fig. 4.3a, f). Whether the escape is an active mechanism, with the bacteria localizing to the eukaryotic membrane and disrupting it to be taken up by other macrophages, or rather indirect through induction of apoptosis and uptake of resulting apoptotic vescicles by other macrophages, has not been addressed yet. Both these mechanisms are largely uncharacterised, despite being essential for bacterial survival and spread throughout the host organism (Flieger et al., 2018).

To identify the molecular mechanism of YjiS activity, both RNA and protein interactors were considered. Binding of RNA could be excluded (Fig. 4.5). Instead, co-immunoprecipitation and mass spectrometry analysis revealed seven proteins that were identified only in the strain with tagged YjiS, and not in the background control: PspA, LonH, FtsN, YdiJ, YcaO, IlvD, and SL1344_1553 (Fig. 4.6b). The known localization to the inner membrane of several of these proteins supports the observation that YjiS itself localizes to the inner membrane, although it is not itself inserted in it (Fig. 4.2). No study has identified a single YjiS interactor as directly influencing virulence outcome. Nevertheless, one interesting candidate at the center of speculation is PspA: the Psp (phage shock protein) system has been recognized in several bacterial species but no consensus has been reached when describing its role in the cell. One common theme is its involvement in the response to membrane stresses (Flores-Kim and Darwin, 2016). In *Salmonella* infection, PspA is necessary to import metal ions that, as an anti-infection mechanism, are actively secreted from the *Salmonella* containing vacuole (Darwin, 2013; Karlinsey et al., 2010). An inhibitory role of YjiS on PspA would be in accordance with higher levels of PspA increasing *Salmonella* virulence (Hassani et al., 2009).

It is counterintuitive that an mRNA as abundant as *yjiS* could code for a protein that represses virulence, rather than promote it. The class of negative

regulators of virulence should not be confused with antivirulence genes, whose expression is conflicting with infection. Such genes are lost during evolution, or their expression has been downregulated during virulence (Bliven and Maurelli, 2012). This is the case, for example, of the repressor *lacI* which is missing in *Salmonella* (Eswarappa et al., 2009). On the other hand, few examples of genes whose absence makes *Salmonella* more virulent are known. One such case is Lrp (leucine-responsive regulatory protein), which represses SPI-1 and SPI-2 regulators and its absence results in hypervirulent *Salmonella*, both in cell culture and in animals (Baek et al., 2009).

One possibility for a bacterial pathogen to have evolved such mechanisms is to avoid over-stimulating the host response and be rapidly cleared, but rather adapt to the environments encountered, while spreading from host cell to host cell and into different organs. In this context, the phenotype observed for YjiS is in accordance with another transposon-based investigation of *Salmonella* gene requirement for infection carried out in animals (Chaudhuri et al., 2013). The lack of *yjiS* attenuated virulence in pig, cattle, and chicken. In light of this, YjiS might dampen virulence to escape clearance in the long term. This mechanism could be active in only a subpopulation of cells, a mechanism that should be tested at a single cell level.

Alternatively, YjiS might carry out a function in conditions not tested so far, and its absence is just incidentally making *Salmonella* better at infection. This condition will likely share similarities with the intracellular environment, which is so far the only condition in which expression of *yjiS* was detected. Similarly, the targets of YjiS might not be present in cells grown in SPI-2 medium. Along these lines, an additional possibility is that YjiS senses a specific environment inside the host organism that is not reflected in the experimental setup used so far, which were growth in SPI-2 medium or infection of RAW264.7 macrophages.

An interesting aspect of YjiS at the level of the amino acid composition is the presence of the DUF1127 domain, shared with other small proteins. While the hallmark of this domain is often considered to be its high prevalence of arginines (Kraus et al., 2020; Grützner et al., 2021), the most highly conserved residues are the "DIGL" motif located in the second half of the DUF1127. The requirement

of arginines and the DIGL motif were tested using the infection outcome as a readout, but no single residue played an essential role in the function of YjiS as it relates to virulence (Fig. 4.4).

Overall, while its mechanism of activity is unclear, this work has described YjiS as a new potential regulator of *Salmonella* virulence.

# 5 Conclusions and outlook

The field of small proteins as a class of understudied and underannotated molecules has opened up in the last decade, with several studies showing how small proteins are fundamental in the regulation of diverse biological processes (Orr et al., 2020). These were key aspects that drove the design of the present work, which initially addressed the identification of new sORFs in *Salmonella* Typhimurium, and subsequently provided a global description of small protein expression and requirement for infection. This was aimed at identifying new regulators of virulence.

sORF predictions were generated combining a computational approach (sPepFinder, (Li and Chao, 2020)) with an experimental one (Ribo-seq). While Ribo-seq is the most powerful experiment to identify mRNAs being translated, it comes with limitations that can be overcome by complementing it with *in silico* predictions. The two approaches identified 139 STsORFs, which add up to the small proteins currently annotated in *Salmonella* (470) to a total of 609. Among the 139 STsORFs, 17 of the 19 candidates chosen for validation were confirmed. This high success rate promises that even more new STsORFs will be confirmed in the future.

It is reasonable to expect that performing Ribo-seq on bacteria grown in conditions rarely studied will further expand the list of small proteins. With this, new features of amino acid composition or noncanonical start codons will generate a more comprehensive training set used for machine learning tools such as sPepFinder, leading to the identification of new sORFs. Furthermore, evidence indicates that bacterial genomes might be more densely organized than what was believed so far, with alternative sORFs encoded inside larger CDSs (Meydan et al., 2019; Yuan et al., 2017). These alternative ORFs are often discarded as potential new genes because they are indistinguishable from the larger ORFs in standard

transcriptomic data. Indeed, only a handful of these are known in bacteria (Meydan et al., 2019; Yuan et al., 2017), which makes it difficult to identify recurrent patterns that might help with their recognition. Altogether, these findings suggest that even more small proteins in *Salmonella*, now accounting for 13 % of all its CDSs, are yet to be found.

Here, data analysis performed with the updated annotation highlighted genes of interest in the context of infection, namely ones that were highly transcribed in the host and whose absence affected infection outcome. Cross-referencing these data lead to the identification of MgrB and YjiS as new regulators of virulence, albeit in different ways. While MgrB is necessary to invade and replicate in host cells, the presence of YjiS lowers bacterial load at late infection stages, a feature rarely observed in genes highly expressed in the host cell. Similarly, more small proteins which were highly induced in infection (dual RNA-seq) or had an influence on virulence (TraDIS) await functional characterization. Among these, a notable example is STsORF139, which was an extremely abundant new small protein.

Placed into context with the available literature on small proteins, the independent functional investigations carried out here lead to generalisations on recurring features of small proteins. Both YjiS and MgrB, as well as the new STsORF139, are inner membrane(-associated) small proteins (Appendix Fig. 8.2). This is true for a large number of other small proteins (Hemm et al., 2008, 2010; Fontaine et al., 2011). MgrB, YjiS, and STsORF139 are most highly expressed in minimal medium which, due to its limiting nutrients availability and its low pH, can be considered a stress condition. Another common theme is their interaction with larger proteins, whether TCSs (MgrB) or various inner membrane proteins (YjiS); given STsORF139 localization in the 3'UTR of *osmY*, encoding an inner membrane protein induced upon high osmolarity stress, it is possible that STsORF139 will be involved in regulating OsmY. The rationale behind this hypothesis is that a small protein encoded in a polycistronic transcript often acts in the same pathway as the other products of the mRNA. Knowledge regarding the function of the other proteins can directly indicate a condition or readout to be tested for investigating the small protein candidate.

These observations open up a scenario in which small proteins are key players

in fine-tuning bacterial cells response to a stress condition by transiently regulating the activity of larger proteins that are localized at the site of stress sensing, the cell membrane. Such a mechanism allows a fast response to external stimuli. A perfect example of this it the small protein SgrT, encoded by the dual function sRNA SgrS, where both products block import of sugar in the presence of sugar phosphate stress. While SgrS halts the translation of new importers (Papenfort et al., 2013; Rice and Vanderpool, 2011), SgrT inhibits the activity of the ones already produced (Lloyd et al., 2017). It is intriguing to think that in the future well-characterized master regulators will reveal, upon closer inspection, their regulation by new small proteins.

Given the intrinsic challenges in handling small proteins with standard molecular biology techniques such as tagging or western blotting, recent technological advances in, for example, mass spectrometry, or variation of established protocols to adapt them to the case of small proteins will greatly expedite their functional characterization.

Overall, this work demonstrates how orthogonal approaches can catalyse a fast and successful identification of candidates to characterize. Furthermore, an annotation that includes small proteins candidates can be used to mine published datasets which are nowadays generated at an unprecedented pace for classic organisms as well as new, emerging model bacteria. In this way, new key regulators will be identified, possibly providing targets for the treatment of infections or, in the case of commensal bacteria, the manipulation of the microbiota.

# 6 Materials and Methods

## 6.1 Equipment, consumables, and reagents

### 6.1.1 Instruments

**Table 6.1:** List of instruments and software used.

| Instruments and software | Manufacturer |
|---|---|
| BD Accuri C6 | BD Biosciences |
| Cell culture hood Safe 2020 Class II | Thermo Scientific |
| Centrifuge Eppendorf 5415R | Eppendorf |
| Centrifuge Eppendorf 5424 | Eppendorf |
| Centrifuge Hereaus Multifuge X3R | Thermo Scientific |
| Electroporator MicroPulser | Bio-Rad |
| FLA eraser for imaging plates | GE Healthcare |
| FlowJo | BD Biosciences |
| GelStick Imager | Intas |
| Genome browser IGV 2.8.2 | Thorvaldsdóttir et al. (2013) |
| Gradient station *ip* | Biocomp Instruments |
| Graphics and statistics software Prism 8.1.0 | GraphPad Software, Inc. |
| Heat block Thermomixer comfort | Eppendorf |
| Horizontal gel electrophoresys Perfect Blue Mini S, M, L | PeqLab |
| Hybridization oven HP-1000 | UVP |
| ImageQuant LAS 4000 imaging system | GE Healthcare |

**Table 6.1** *continued from previous page*

| Instrument and software | Manufacturer |
|---|---|
| Image processing software ImageJ 1.52u | Schneider et al. (2012) |
| Incubator for bacterial plates | Memmert |
| Incubator for eukaryotic plates HERA-cell 150i | Thermo Scientific |
| Light microscope Eclipse TS100 | Nikon |
| Mixer mill MM400 | Retsch |
| Phosphoimager Typhoon FLA 7000 | GE Healthcare |
| Plate reader Infinite M Plex | TECAN |
| Power supply peqPOWER E250, E300 | PeqLab |
| RealTime CFX96 System | Bio-Rad |
| Research plus pipettes | Eppendorf |
| Rotator SB2 | Stuart |
| Scale 572 | Kern |
| Semi-dry electroblotter Perfect Blue SEDEC M | PeqLab |
| Shaking incubator Innova 44 | New Brunswick Scientific |
| Sonicator Sonoplus HD 70 | Bandelin |
| Spectrophotometer NanoDrop 2000 | PerkinElmer |
| Spectrophotometer Ultraspec 10 Cell Density Meter | Amersham Biosciences |
| Spreadsheet editor Excel 365 ProPlus | Microsoft corporation |
| Table-top ultracentrifuge optima MAX-XP | Beckman Coulter |
| Thermal cycler MJ Mini | Bio-Rad |
| Ultracentrifuge Optima XP-80 | Beckman Coulter |
| Ultracentrifuge rotor SW 40 Ti | Beckman Coulter |
| Ultracentrifuge rotor SW 60 Ti | Beckman Coulter |
| UV crosslinker (254 nm) | Vilber |
| Vector graphics editor CorelDRAW X7 | Corel Corporation |

**Table 6.1** *continued from previous page*

| **Instrument and software** | **Manufacturer** |
|---|---|
| Vertical gel electrophoresis Perfect Blue Twin | PeqLab |
| Victor3 1420 Multilabel Counter | PerkinElmer |
| Vortex Genie 2 | Scientific Industries |
| Waterbath 1092 | GFL |
| Western blot imaging system Amersham ImageQuant 800 | Cytiva |

### 6.1.2 Consumables

**Table 6.2:** List of consumables used.

| Consumables | Manufacturer |
|---|---|
| 0.45 µm polyethersulfone membrane | Millipore |
| 6-well plates | Corning |
| 12-well plates | Corning |
| 96-well plates | Nunc |
| Bolt 4–12% Bis-Tris gels | Thermo Fischer Scientific |
| Centrifuge tubes | Sarstedt |
| Cuvettes | Sarstedt |
| Disposale glass pipettes | Kimble |
| Electroporation cuvettes | Cell projects |
| Fisherbrand cell scrapers | Fischer Scientific |
| G-25 MicroSpin columns | GE Healthcare |
| Glass beads 0.1 mm | Roth |
| Glass bottles | Schott |
| Hard-Shell 96-Well PCR Plates | Bio-Rad |
| Hybond-XL membranes | GE Healthcare |
| LTQ Orbitrap Elite | Thermo Fischer Scientific |
| Neubauer counting chamber | HBG Henneberg-Sander |
| PCR tubes | Thermo Fischer Scientific |
| Petri dishes | Corning |
| Phase lock gel tubes 2ml | 5 Prime |
| Phosphor screen | Fujifilm |
| Pipetboy acu-jet pro | BRAND |
| Pipette tips | Sarstedt |
| PVDF membrane | GE Healthcare |
| Safe-lock tubes 1.5 mL, 2 mL | Eppendorf |
| Sierological pipettes (plastic) | Greiner bio-one |
| T-75 flasks | Corning |
| Ultracentrifugation tubes | Seton |

### 6.1.3 Chemicals and reagents

**Table 6.3:** List of chemicals and reagents used.

| Chemicals and reagents | Manufacturer |
|---|---|
| 1,4-Dithiothreitol (DTT) | Roth |
| Acetone | Roth |
| Albumin fraction V | Roth |
| Anti-FLAG M2 magnetic beads | Sigma |
| Carbenicillin | Roth |
| Chloramphenicol | Roth |
| Dimethyl dulfoxide (DMSO) | Roth |
| DMEM | Gibco |
| DNA loading buffer (6x) | Thermo Fisher Scientific |
| dNTPs | Thermo Fisher Scientific |
| Dynabeads protein A/G | Thermo Fisher Scientific |
| Dynabeads MyOne Streptavidin T1 beads | Invitrogen |
| ECL western blot detection reagent | GE Healthcare |
| EDTA | Roth |
| EGTA | Roth |
| Ethanol | Roth |
| Ethanol absolute | Merck |
| Formaldehyde | Roth |
| Gel loading buffer II | Ambion |
| GeneRuler 1 kb DNA Ladder | Fermentas |
| Gentamicin sulfate salt | Sigma |
| Glycerol (99%) | Sigma |
| Glycine | Roth |
| GlycoBlue | Thermo Fisher Scientific |
| $\gamma$ -32P-ATP | Hartmann Analytic |
| Igepal | Sigma |
| Isopropanol | Roth |
| Kanamycin sulfate | Roth |

**Table 6.3** *continued from previous page*

| **Chemicals and reagents** | **Manufacturer** |
|---|---|
| L(+)-arabinose | Roth |
| LDS sample buffer | Thermo Fischer Scientific |
| MES buffer | Thermo Fischer Scientific |
| Methanol | Roth |
| Milk powder | Roth |
| N-lauroylsarcosine sodium salt | Sigma-Aldrich |
| PageRuler Plus prestained protein ladder | Thermo Fischer Scientific |
| PBS | Gibco |
| Phenylmethyl sulfonylfluoride (PMSF) | Serva |
| Propidium iodide | Sigma |
| pUC mix marker, 8 | Fermentas |
| RedSafe | ChemBio |
| Rifampicin | Fluka |
| RNA ladder high and low range | Fermentas |
| Roti-Aqua P/C/I | Roth |
| Roti-Blue | Roth |
| Roti-Hybri-Quick | Roth |
| Rotiphorese gel 40 (19:1) | Roth |
| Rotiphorese gel 40 (37.5:1) | Roth |
| RPMI | Gibco |
| SimplyBlue Coomassie | Thermo Fischer Scientific |
| Sodium desoxycholate | Sigma-Aldrich |
| SYBR Gold Nucleic Acid Gel Stain | Invitrogen |
| Tetracylcin | Roth |
| Trichloroacetic acid | Sigma |
| Triton X-100 | Sigma |
| TRIzol | Invitrogen |

### 6.1.4 Enzymes and kits

**Table 6.4:** List of enzymes used.

| **Enzyme** | **Manufacturer** |
|---|---|
| Calf intestinal phosphatase (CIP) | NEB |
| DNaseI | Thermo Fischer Scientific |
| Lysozyme | Roth |
| Micrococcal nuclease (MNase) | NEB |
| Phusion DNA polymerase | NEB |
| Polynucleotide kinase (PNK) | Thermo Fischer Scientific |
| Restriction enzymes | Thermo Fischer Scientific |
| RNA Fragmentation Reagent | Ambion |
| RNase A/T1 mix | Thermo Fischer Scientific |
| RNase inhibitor | Thermo Fischer Scientific |
| SUPERaseIN RNase inhibitor | Ambion |
| SuperScript II reverse transcriptase | Thermo Fischer Scientific |
| Trypsin-EDTA | Gibco |
| T4 DNA ligase | NEB |
| *Taq* DNA polymerase | NEB |

**Table 6.5:** List of commercial kits used.

| **Commercial kits** | **Manufacturer** |
|---|---|
| CytoTox 96(R) Non-Radioactive Cytotoxicity Assay | Promega |
| MEGAscript T7 Kit | Ambion |
| NucleoSpin Gel and PCR clean-up | Macherey-Nagel |
| NucleoSpin Plasmid EasyPure | Macherey-Nagel |
| Takyon No ROX SYBR 2X MasterMix blue dTTP | Eurogentec |
| TakyonOne-Step Kit Converter | Eurogentec |

### 6.1.5 Antibodies

**Table 6.6:** List of antibodies used.

| Antibody (source) | Working dilution | Provider |
|---|---|---|
| $\alpha$-FLAG (mouse) | 1:1,000 in 3% BSA | Sigma |
| $\alpha$-GroEL (rabbit) | 1:10,000 in 3% BSA | Sigma |
| $\alpha$-FliC (mouse) | 1:1,000 in 3% BSA | InvivoGen |
| $\alpha$-SopB (rabbit) | 1:3,000 in 3% BSA | Kolbe, Hamburg, Germany |
| $\alpha$-SipC (rabbit) | 1:3,000 in 3% BSA | Kolbe, Hamburg, Germany |
| $\alpha$-OmpA (rabbit) | 1:50,000 in 3% BSA | Belin, Stockholm, Sweden |
| $\alpha$-OmpX (rabbit) | 1:5,000 in 3% BSA | Linke, Tuebingen, Germany |
| $\alpha$-mouse (goat) | 1:1,000 in 3% BSA | Thermo Scientific |
| $\alpha$-rabbit (goat) | 1:10,000 in 3% BSA | Thermo Scientific |

## 6.2 Media, buffers, and solutions

### 6.2.1 Media

**Complete DMEM**: DMEM (Gibco); 10% fetal calf serum (FCS, Biochrom), 2 mM L-glutamine (Gibco); 1 mM sodium pyruvate (Gibco). 1% penicillin/streptomycin was added when required.

**Complete RPMI**: RPMI (Gibco); 10% fetal calf serum (FCS, Biochrom), 2 mM L-glutamine (Gibco); 1 mM sodium pyruvate (Gibco). 1% penicillin/streptomycin was added when required.

**Lennox broth (LB) medium**: 10% (w/v) tryptone; 0.5% (w/v) yeast extract; 85.6 mM NaCl.

**LB agar**: 10% (w/v) tryptone; 0.5% (w/v) yeast extract; 85.6 mM NaCl, 1.2% (w/v) agar.

**SPI-2 medium**: MES 170 mM, KCl 5 mM, $(NH_4)_2SO_4$ 7.5 mM, $K_2SO_4$ 0.5 mM, $KH_2PO_4$ 1 mM, $MgCl_2$ 8 $\mu$M, casaminoacids 0.1%. The medium is brought to pH 5.8 with KOH.

### 6.2.2 Buffers and solutions

**Lysis buffer for cell fractionation**: 20 mM Tris HCl pH 7.5, 150 mM KCl, 1 mM $MgCl_2$, 1 mM PMSF.

**Lysis buffer for co-immunoprecipitation**: 1 mM $MgCl_2$, 20 mM Tris HCl pH 7.5, 1% igepal, 1mM DTT, 1mM PMSF, 10 U/ml DNaseI.

**Lysis buffer for the isolation of the inner membrane**: lysis buffer with 0.5% N-lauroylsarcosyn.

**Lysis buffer for the isolation of the outer membrane**: lysis buffer with 2% triton X-100.

**Developer solution**: 60 g $Na_2CO_3$, 4 mg $Na_2S_2O_3$ x 5 $H_2O$, 0.5 ml formaldehyde (37%), $H_2O$ to 1 l.

**DNA loading dye (5x stock)**: 10 mM Tris-HCl pH 7.6, 60% (v/v) glycerol, 60 mM EDTA pH 8.0, 0.025% (w/v) bromophenol blue.

**Ethanol/sodium acetate 30:1**: 29 ml ethanol, 1 ml 3M sodium acetate pH 5.2 or 6.5.

**Fixing solution**: 500 ml ethanol, 120 ml acetic acid, 0.5 ml formaldehyde (37%), $H_2O$ to 1 l.

**PBS (10x)**: 2 g KCl, 2.4 g $KH_2PO_4$, 80 g NaCl, 14.4 g $Na_2HPO_4$, adjust the pH to 7.4, $H_2O$ to 1 l.

**Proteinase K buffer**: 0.5% (w/v) SDS, 50 mM Tris-HCl, 0.4 M EDTA pH8.

**Polyacrilamide (PAA) gel electrophoresis solution (6%), for RNA**: 100 ml 10x TBE, 420 g urea, 150 or 100 ml Rotiphorese gel 40 (19:1) for 6% or 4% gels, $H_2O$ to 1 l.

**PAA running gel electrophoresis solution, for proteins**: 3.75 ml lower buffer, 3 or 3.25 ml Rotiphorese gel 40 (37.5:1) and 3.25 or 3 ml $H_2O$ for 12 or 15% gels, 75 $\mu$l 10% (w/v) APS, 7.5 $\mu$l TEMED.

**PAA stacking gel electrophoresis solution, for proteins**: 1.25 ml upper buffer, 1 ml Rotiphorese gel 40 (37.5:1), 7.5 ml $H_2O$, 90 $\mu$l 10% (w/v) APS, 9 $\mu$l TEMED.

**Protein loading dye (5x)**: 15 g SDS, 46.95 ml 1M Tris-HCl, pH 6.8, 75 ml glycerol, 11.56 g DTT, 0.075 g bromophenol blue, $H_2O$ to 150 ml.

**Ribo-seq lysis buffer**: 100 mM $NH_4Cl$, 10 mM $MgCl_2$, 20 mM Tris-HCl pH 8.0, 0.1% NP-40, 0.4% Triton X-100, 50 U/ml DNase I, 500 U RNase Inhibitor, 1 mM chloramphenicol.

**RNA loading dye (2x)**: 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene

cyanol, 18 $\mu$M EDTA pH 8, 0.13% (w/v) SDS, 95% formamide.

**SDS running buffer (10x)**: 30.275 g Tris base, 144 g glycine, 10 g SDS, $H_2O$ to 1 l.

**Sensitizer**: 0.2 g $Na_2S_2O_3$ x $5H_2O$, $H_2O$ to 1 l.

**Silver staining solution**:2 g $AgNO_3$, 0.75 ml formaldehyde (37%), $H_2O$ to 1 l.

**Silver staining stop solution**: 10 g glycine, $H_2O$ to 1 l.

**SSC buffer (20x stock)**: 3M NaCl, 0.3M sodium citrate pH 7.

**Stop mix**: 95% ethanol, 5% acidic phenol.

**Sucrose buffer**: 100 mM NHtextsubscript4Cl, 10 mM MgCltextsubscript2, 5 mM CaCltextsubscript2, 20 mM Tris-HCl pH 8.0, 1 mM chloramphenicol.

**TAE buffer (50x stock)**: 242 g Tris base, 51.7 ml acetic acid, 10mM EDTA pH 8, $H_2O$ to 1 l.

**TBE buffer (10x stock)**: 108 g Tris base, 55 g boric acid, 20mM EDTA pH 8, $H_2O$ to 1 l.

**TBS buffer (10x stock)**: 24.11 g Tris base, 72.6 g NaCl, adjust to pH 7.4 with HCl, $H_2O$ to 1 l.

**TBS-T buffer (10x stock)**: 1 TBS, 0.1% (v/v) Tween-20.

**TE (1x)**: 100 mM Tris-HCl pH 8, 10 mM EDRA pH8.

**Transfer buffer (10x stock)**: 3 g Tris base, 14.4 g glycine, 200 ml methanol, $H_2O$ to 1 l.

**Tris running gel solution**: 1.5 M Tris-HCl, pH 8.8, 0.4% (w/v) SDS.

**Tris stacking gel solution**: 0.5M Tris-HCl pH 6.8, 0.4% (w/v) SDS.

## 6.3 Bacterial strains, plasmids, and oligonucleotides

### 6.3.1 Bacterial strains

**Table 6.7:** List of bacterial strains used.

| Name | Genotype | Marker | Comment | Reference |
|---|---|---|---|---|
| JVS-00584 | Δ*hfq* | | *Salmonella* lacking *hfq* | Sittka et al. (2007) |
| JVS-01574 | wild-type *Salmonella* SL1344 | | | laboratory strain collection |
| JVS-03858 | GFP$^+$ | Cm | constitutive expression of GFP from the chromosome | Papenfort et al. (2009) |
| JVS-04117 | *ptsG*::3xFLAG | Kan | *Salmonella ptsG* tagged at the C-terminus with 3xFLAG | laboratory strain collection |
| JVS-05709 | wild-type *E. coli* MG1655 | | | laboratory strain collection |
| JVS-10315 | Δ*proQ* | | *Salmonella* lacking *proQ* | Smirnov et al. (2016) |
| JVS-10318 | Δ*hfq*Δ*proQ* | | *Salmonella* lacking *hfq* and *proQ* | Westermann et al. (2019) |
| JVS-11532 | Δ*yjiS* | Kan | *Salmonella* lacking *yjiS* | this study |
| JVS-11583 | Δ*mgrB* | Kan | *Salmonella* lacking *mgrB* | this study |
| JVS-11608 | *sipA*::3xFLAG | Kan | *Salmonella sipA* tagged at the C-terminus with 3xFLAG | laboratory strain collection |
| JVS-11669 | Δ*yjiS* | | cured *Salmonella* lacking *yjiS* | this study |
| JVS-11857 | Δ*mgrB* | | cured *Salmonella* lacking *mgrB* | this study |
| JVS-12294 | *yjiS*::SPA | Kan | *Salmonella yjiS* tagged at the C-terminus with SPA tag | this study |
| JVS-12374 | *yjiS*::SPA | | cured *Salmonella yjiS* tagged at the C-terminus with SPA tag | this study |

**Table 6.7** *continued from previous page*

| Name | Genotype | Marker | Comment | Reference |
|---|---|---|---|---|
| JVS-12601 | STsORF49::SPA | Kan | *Salmonella* STsORF49 tagged at the C-terminus with SPA tag | this study |
| JVS-12603 | STsORF54::SPA | Kan | *Salmonella* STsORF54 tagged at the C-terminus with SPA tag | this study |
| JVS-12605 | STsORF79::SPA | Kan | *Salmonella* STsORF79 tagged at the C-terminus with SPA tag | this study |
| JVS-12607 | STsORF80::SPA | Kan | *Salmonella* STsORF80 tagged at the C-terminus with SPA tag | this study |
| JVS-12609 | STsORF99::SPA | Kan | *Salmonella* STsORF99 tagged at the C-terminus with SPA tag | this study |
| JVS-12611 | STsORF102::SPA | Kan | *Salmonella* STsORF102 tagged at the C-terminus with SPA tag | this study |
| JVS-12613 | STsORF128::SPA | Kan | *Salmonella* STsORF128 tagged at the C-terminus with SPA tag | this study |

**Table 6.7** *continued from previous page*

| Name | Genotype | Marker | Comment | Reference |
|---|---|---|---|---|
| JVS-12615 | STsORF132::SPA | Kan | *Salmonella* STsORF132 tagged at the C-terminus with SPA tag | this study |
| JVS-12616 | STsORF133::SPA | Kan | *Salmonella* STsORF133 tagged at the C-terminus with SPA tag | this study |
| JVS-12617 | STsORF138::SPA | Kan | *Salmonella* STsORF138 tagged at the C-terminus with SPA tag | this study |
| JVS-12619 | STsORF139::SPA | Kan | *Salmonella* STsORF139 tagged at the C-terminus with SPA tag | this study |
| JVS-12623 | STsORF23::SPA | Kan | *Salmonella* STsORF23 tagged at the C-terminus with SPA tag | this study |
| JVS-12624 | STsORF40::SPA | Kan | *Salmonella* STsORF40 tagged at the C-terminus with SPA tag | this study |
| JVS-12626 | STsORF44::SPA | Kan | *Salmonella* STsORF44 tagged at the C-terminus with SPA tag | this study |

**Table 6.7** *continued from previous page*

| Name | Genotype | Marker | Comment | Reference |
|---|---|---|---|---|
| JVS-12628 | STsORF59::SPA | Kan | *Salmonella* STsORF59 tagged at the C-terminus with SPA tag | this study |
| JVS-12630 | STsORF111::SPA | Kan | *Salmonella* STsORF111 tagged at the C-terminus with SPA tag | this study |
| JVS-12632 | STsORF114::SPA | Kan | *Salmonella* STsORF114 tagged at the C-terminus with SPA tag | this study |
| JVS-12634 | STsORF43::SPA | Kan | *Salmonella* STsORF43 tagged at the C-terminus with SPA tag | this study |
| JVS-12636 | STsORF129::SPA | Kan | *Salmonella* STsORF129 tagged at the C-terminus with SPA tag | this study |
| JVS-12918 | Δ*mgrBsipA*::3xFLAG | Kan | *Salmonella* Δ*mgrB* was transdcuced with the lysate of *Salmonella* with *sipA*::3xFLAG | this study |

### 6.3.2 Plasmids

**Table 6.8:** List of plasmids used.

| Name | Insert and purpose | Parental plasmid | Resistance | Reference |
|---|---|---|---|---|
| pJV300 | empty vector control | pZE12 | Amp | Sittka et al. (2007) |
| pKD4 | template for gene deletion | | Amp | Datsenko and Wanner (2000) |
| pKD46 | expression of ΛRED recombinase | | Amp | Datsenko and Wanner (2000) |
| pJL-148 | template for SPA tag | | Kan | Babu et al. (2009) |
| pCP20 | expression of Flp recombinase | | Amp | Cherepanov and Wackernagel (1995) |
| pKP8-35 | empty vector control | pBAD | Amp | Papenfort et al. (2012) |
| pXG-0 | GFP negative vector control | | Cm | Urban and Vogel (2007) |
| pXG-1 | GFP positive vector control | | Cm | Urban and Vogel (2007) |
| pXG-10 | GFP empty vector control | | Cm | Urban and Vogel (2007) |
| pXG-30 | GFP empty vector control | | Cm | Urban and Vogel (2007) |
| pEV-003 | translational fusion of *motA* | pXG-10 | Cm | this study |
| pEV-004 | transcriptional fusion of *motA* | pXG-10 | Cm | this study |
| pEV-005 | translational fusion of *fliD* | pXG-10 | Cm | this study |
| pEV-006 | translational fusion of *fliC* | pXG-10 | Cm | this study |
| pEV-007 | transcriptional fusion of *fliC* | pXG-10 | Cm | this study |
| pEV-008 | pulse-expression of *yjiS* | pBAD | Amp | this study |
| pEV-010 | expression of *yjiS* under its native promoter and with its native terminator | pZE12 | Amp | this study |
| pEV-011 | transcriptional fusion of *fliD* | pXG-10 | Cm | this study |
| pEV-018 | constitutive expression of *mgrB* | pZE12 | Amp | this study |

**Table 6.8** *continued from previous page*

| Name | Insert and purpose | Parental plasmid | Resistance | Reference |
|---|---|---|---|---|
| pEV-022 | constitutive expression of *E. coli* *mgrB* | pZE12 | Amp | this study |
| pEV-024 | expression of *mgrB* under its native promoter and with the native terminator | pZE12 | Amp | this study |
| pEV-034 | arginine ($1^{st}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-035 | arginine ($2^{nd}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-036 | arginine ($3^{rd}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-037 | arginine ($4^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-038 | arginine ($5^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-039 | arginine ($6^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-040 | arginine ($7^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-041 | arginine ($8^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |

**Table 6.8** *continued from previous page*

| Name | Insert and purpose | Parental plasmid | Resistance | Reference |
|---|---|---|---|---|
| pEV-042 | arginine ($9^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-043 | arginine ($10^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-044 | arginine ($11^{th}$) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-045 | aspartic acid (DIGL motif) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-046 | isoleucin (DIGL motif) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-047 | glycine (DIGL motif) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-048 | leucine (DIGL motif) to alanine mutation of *yjiS* based on pEV-010 | pZE12 | Amp | this study |
| pEV-100 | constitutive expression of MgrR | pZE12 | Amp | this study |
| pZE12-proQ | expression of *proQ* under its native promoter | pZE12 | Amp | Smirnov et al. (2016) |
| pKF-68,3 | pulse-expression of SdsR | pBAD | Amp | Fröhlich et al. (2016) |
| pEH-526 | translational reporter of *flgC* | pXG-30 | Cm | |

### 6.3.3 Oligonucleotides

**Table 6.9:** List of oligonucleotides used.

| **Name** | **Purpose** | **Sequence (5' →3')** |
|---|---|---|
| pZE-A | Sense to sequence pZE-12 | GTGCCACCTGACGTCTAAGA |
| pZE-XbaI | Antisense to sequence pZE-12 | TCGTTTTATTTGATGCCTCTAGA |
| pZE-Cat | Sense to sequence pXG plasmids | TGGGATATATCAACGGTGGT |
| JVO-155 | Antisense to sequence pXG plasmids | CCGTATGTAGCATCACCTTC |
| JVO-322 | Northern blot probe for 5S rRNA | CTACGGCGTTTCACTTCTGAGTTC |
| JVO-1371 | Northern blot probe for MicA | GGATGATGATAACAAATGCGCGTCT |
| JVO-1592 | Northern blot probe for *fliC* | GGATGATGATAACAAATGCGCGTCT |
| JVO-11221 | Northern blot probe for CsrB | GGCGTGTCCTGCTGAAGCGTCATCATCCTGACGT |
| JVO-11222 | Northern blot probe for CsrC | CTGAGTCGTTGATCCTGTTAGCGTCCTCGCTA |
| JVO-15068 | Antisense to check deletion of *mgrB* | GTCGTTCATTTCACCACCTC |
| JVO-15134 | Sense to check deletion of *mgrB* | CACACCAGGGCAAACAATAC |

**Table 6.9** *continued from previous page*

| **Name** | **Purpose** | **Sequence (5' →3')** |
|---|---|---|
| JVO-15138 | Sense to delete *mgrB* | ATGATAATTTAGCGACATAAGATGTAAGATCGGAGAGTGGAGATTACAAAGATGACGATGAC |
| JVO-15139 | Antisense to delete *mgrB* | CTACCACTGCTACACGGGAAGGAAATCTCTGGTGTAAAACGTCATATGAATATCCTCCTTAG |
| JVO-15370 | Sense for *motA* transcriptional fusion | gtttttGACGTCTTGAATTGTCGAGCTGTAAC |
| JVO-15371 | Sense for *motA* translational fusion | gtttttATGCATAACTGCTGGATGAACAGATCG |
| JVO-15372 | Antisense for *motA* transcriptional/translational fusion | gtttttGCTAGCGATGGCCTTCCCGTTGTTGC |
| JVO-15373 | Sense for *fliD* transcriptional fusion | gtttttGACGTCAACTACTGTTTGCAATCAAA |
| JVO-15374 | Sense for *fliD* translational fusion | gtttttATGCATGCACCTGAATTTCGAACTT |
| JVO-15375 | Antisense for *fliD* transcriptional/translational fusion | gtttttGCTAGCTTTGGTAATTGGCGTTAAGC |
| JVO-15376 | Sense for *fliC* transcriptional fusion | gtttttGACGTCACGGTGAGAAACCGTGGGCA |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-15377 | Sense for *fliC* translational fusion | gtttttATGCATGCAATCTGGAGGCAAAGTTT |
| JVO-15378 | Antisense for *fliC* transcriptional/translational fusion | gtttttGCTAGCAATCGCCTGACCTGCCGCAT |
| JVO-15408 | Northern blot probe for *yjiS* | CCGCTTACGCAGCCACCACC |
| JVO-15683 | Sense to clone *mgrB* in pZE12 under pL | P-ATAAGATGTAAGATCGGAGAG |
| JVO-15684 | Antisense to clone *mgrB* in pZE12 | gttttttctagaATTATCGTTCTGCAACCAGACGAATG |
| JVO-15687 | Sense to clone *mgrB* in pZE12 with its native promoter | gttttttctcgagCAATTGCGCGAATGATCAAAC |
| JVO-15817 | Sense to clone *E. coli mgrB* in pZE12 under pL | P-ATAAGGTAGGTGAAACGGAGATTG |
| JVO-15818 | Antisense to clone *E. coli mgrB* in pZE12 | gttttttctagaTCACCACCTCTGAGTAGAGG |
| JVO-15689 | Northern blot probe for *mgrB* | CACAGCAACAGGCATACTAC |

**Table 6.9** *continued from previous page*

| **Name** | **Purpose** | **Sequence (5' →3')** |
|---|---|---|
| JVO-16959 | Sense to add SPA-tag at the C-terminus of *yjiS* | GCAGTTAAGGGATATCGGGTTGGAAAGGAAGGATGTGGAATCCATGGAAAAGAGAAGA |
| JVO-16960 | Antisense to add SPA-tag at the C-terminus of *yjiS* | CATAAAATCCTTTAATTACTATTGCCCTTGAATAGATTCAAGTTCCTATTCCGAAGTTC |
| JVO-16972 | Sense to mutate Arg #1 in *yjiS* | GAATTTTACGAGAACGCCCCCAGACAACCATTCATTGGCTTTGTGCAG |
| JVO-16973 | Antisense to mutate Arg #1 in *yjiS* | CTGCACAAAGCCAATGAATGGTTGTCTGGGGGCGTTCTCGTAAAATTC |
| JVO-16974 | Sense to mutate Arg #2 in *yjiS* | GAATTTTACGAGAACCGCCCCGCACAACCATTCATTGGCTTTG |
| JVO-16975 | Antisense to mutate Arg #2 in *yjiS* | CAAAGCCAATGAATGGTTGTGCGGGGCGGTTCTCGTAAAATTC |
| JVO-16976 | Sense to mutate Arg #3 in *yjiS* | CATTCATTGGCTTTGTGCAGCTCGCACGGGCGCTAAAAAGGTGG |
| JVO-16977 | Antisense to mutate Arg #3 in *yjiS* | CCACCTTTTTAGCGCCCGTGCGAGCTGCACAAAGCCAATGAATG |
| JVO-16978 | Sense to mutate Arg #4 in *yjiS* | CCATTCATTGGCTTTGTGCAGCTCAGAGCGGCGCTAAAAAGGTGG |
| JVO-16979 | Antisense to mutate Arg #4 in *yjiS* | CCACCTTTTTAGCGCCGCTCTGAGCTGCACAAAGCCAATGAATGG |

**Table 6.9** *continued from previous page*

| **Name** | **Purpose** | **Sequence (5' →3')** |
|---|---|---|
| JVO-16980 | Sense to mutate Arg #5 in *yjiS* | GGCGCTAAAAGCGTGGTGGCTGCGTAAGCGGG |
| JVO-16981 | Antisense to mutate Arg #5 in *yjiS* | CCCGCTTACGCAGCCACCACGCTTTTAGCGCC |
| JVO-16982 | Sense to mutate Arg #6 in *yjiS* | GCTAAAAAGGTGGTGGCTGGCTAAGCGGGCCTGTCAGGC |
| JVO-16983 | Antisense to mutate Arg #6 in *yjiS* | GCCTGACAGGCCCGCTTAGCCAGCCACCACCTTTTTAGC |
| JVO-16984 | Sense to mutate Arg #7 in *yjiS* | GTGGCTGCGTAAGGCGGCCTGTCAGGCGCTG |
| JVO-16985 | Antisense to mutate Arg #7 in *yjiS* | CAGCGCCTGACAGGCCGCCTTACGCAGCCAC |
| JVO-16986 | Sense to mutate Arg #8 in *yjiS* | CCTGTCAGGCGCTGGCACGAATGAGTGACGAGCAG |
| JVO-16987 | Antisense to mutate Arg #8 in *yjiS* | CTGCTCGTCACTCATTCGTGCCAGCGCCTGACAGG |
| JVO-16988 | Sense to mutate Arg #9 in *yjiS* | CCTGTCAGGCGCTGCGAGCAATGAGTGACGAGCAG |
| JVO-16989 | Antisense to mutate Arg #9 in *yjiS* | CTGCTCGTCACTCATTGCTCGCAGCGCCTGACAGG |
| JVO-16990 | Sense to mutate Arg #10 in *yjiS* | GAATGAGTGACGAGCAGTTAGCGGATATCGGGTTGGAAAGG |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-16991 | Antisense to mutate Arg #10 in *yjiS* | CCTTTCCAACCCGATATCCGCTAACTGCTCGTCACTCATTC |
| JVO-16992 | Sense to mutate Arg #11 in *yjiS* | GCAGTTAAGGGATATCGGGTTGGAAGCGAAGGATGTGGAATAGTGAATC |
| JVO-16993 | Antisense to mutate Arg #11 in *yjiS* | GATTCACTATTCCACATCCTTCGCTTCCAACCCGATATCCCTTAACTGC |
| JVO-17000 | Sense to mutate D in *yjiS* DIGL motif | GAGTGACGAGCAGTTAAGGGCTATCGGGTTGGAAAGGAAGGATGTG |
| JVO-17001 | Antisense to mutate D in *yjiS* DIGL motif | CACATCCTTCCTTTCCAACCCGATAGCCCTTAACTGCTCGTCACTC |
| JVO-17002 | Sense to mutate I in *yjiS* DIGL motif | GTGACGAGCAGTTAAGGGATGCCGGGTTGGAAAGGAAGGATGTG |
| JVO-17003 | Antisense to mutate I in *yjiS* DIGL motif | CACATCCTTCCTTTCCAACCCGGCATCCCTTAACTGCTCGTCAC |
| JVO-17004 | Sense to mutate G in *yjiS* DIGL motif | GACGAATGAGTGACGAGCAGTTAAGGGATATCGCGTTGGAAAGGAAGGATGTG |
| JVO-17005 | Antisense to mutate G in *yjiS* DIGL motif | CACATCCTTCCTTTCCAACGCGATATCCCTTAACTGCTCGTCACTCATTCGTC |
| JVO-17006 | Sense to mutate L in *yjiS* DIGL motif | GTGACGAGCAGTTAAGGGATATCGGGGCGGAAAGGAAGGATG |
| JVO-17007 | Antisense to mutate L in *yjiS* DIGL motif | CATCCTTCCTTTCCGCCCCGATATCCCTTAACTGCTCGTCAC |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18622 | Sense to add SPA-tag at the C-terminus of STsORF23 | GATCAACGGCAAACCGGAGCAGGATAACCCTTCCACGAAATCCATGGAAAAGAGAAGA |
| JVO-18623 | Antisense to add SPA-tag at the C-terminus of STsORF23 | GTTGAGGATGCAAAATAAAGCGGCGTACCGCACCGCCGCGAGTTCCTATTCCGAAGTTC |
| JVO-18624 | Sense to validate SPA-tag at the C-terminus of STsORF23 | AAATCCTAACGATGCCAGGC |
| JVO-18625 | Antisense to validate SPA-tag at the C-terminus of STsORF23 | TCATTACCTGCCTGATTATGCG |
| JVO-18626 | Sense to add SPA-tag at the C-terminus of STsORF40 | TAATAAAGATATCGAATTTTTCTACGTCGGAGACGATAGCTCCATGGAAAAGAGAAGA |
| JVO-18627 | Antisense to add SPA-tag at the C-terminus of STsORF40 | TACCTGAAGATATGACACCCGTTGCCAACTGACTGGAATTAGTTCCTATTCCGAAGTTC |
| JVO-18628 | Sense to validate SPA-tag at the C-terminus of STsORF40 | AATTCCCTGAAGAACGGGAG |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18629 | Antisense to validate SPA-tag at the C-terminus of STsORF40 | CAGCAACTGGCGGTAAGAC |
| JVO-18630 | Sense to add SPA-tag at the C-terminus of STsORF44 | TCTGCAATTACCGCCACCTCCTGGTGAGTATTGCTGGCATTCCATGGAAAAGAGAAGA |
| JVO-18631 | Antisense to add SPA-tag at the C-terminus of STsORF44 | CCTTACATGCGCCGACTTATCACTGAAATACTATGGATTTAAGTTCCTATTCCGAAGTTC |
| JVO-18632 | Sense to validate SPA-tag at the C-terminus of STsORF44 | TTTACCATAGCGATGAGAATTCC |
| JVO-18633 | Antisense to validate SPA-tag at the C-terminus of STsORF44 | ACGATTGTCTCCAGCGCG |
| JVO-18634 | Sense to add SPA-tag at the C-terminus of STsORF49 | CGGATTTCTGGCCGCCTGGTTCAGTCACAAATGGGATGATTCCATGGAAAAGAGAAGA |
| JVO-18635 | Antisense to add SPA-tag at the C-terminus of STsORF49 | GCAGCAAGGGGCCGGCTCTGCGAGGGATTATCTCCGTTCAAGTTCCTATTCCGAAGTTC |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18636 | Sense to validate SPA-tag at the C-terminus of STsORF49 | CCGGGATGAAAGCCAGC |
| JVO-18637 | Antisense to validate SPA-tag at the C-terminus of STsORF49 | ATCTCCAGGTCAGCCATACG |
| JVO-18638 | Sense to add SPA-tag at the C-terminus of STsORF54 | GTCGCATCAAACGCCCGGATCTGCGAGGATACGCGCCTGCTCCATGGAAAAGAGAAGA |
| JVO-18639 | Antisense to add SPA-tag at the C-terminus of STsORF54 | AAAAGTGGTCACAATTTACTCTCGCGGTAACGACTCATAGAGTTCCTATTCCGAAGTTC |
| JVO-18640 | Sense to validate SPA-tag at the C-terminus of STsORF54 | ATGGTTGCATCGCAGTAAATGG |
| JVO-18641 | Antisense to validate SPA-tag at the C-terminus of STsORF54 | CCACACAAGGTCAGAATTTTGG |
| JVO-18642 | Sense to add SPA-tag at the C-terminus of STsORF59 | ACCCGAAGACGAAGGAAAAAATCCAAAGAAAAACCAAAAGTCCATGGAAAAGAGAAGA |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18643 | Antisense to add SPA-tag at the C-terminus of STsORF59 | GCCTCCAACGAACCATAGTTATTAAAGATACCCTACAGCAAGTTCCTATTCCGAAGTTC |
| JVO-18644 | Sense to validate SPA-tag at the C-terminus of STsORF59 | ACCAACGTGGCGTGCAGG |
| JVO-18645 | Antisense to validate SPA-tag at the C-terminus of STsORF59 | TATGTGTCAACTCTTCGCTGTTC |
| JVO-18646 | Sense to add SPA-tag at the C-terminus of STsORF79 | GAACGCTCTGCTACGCCCGTTCACCCATTGGACGCTGGATTCCATGGAAAAGAGAAGA |
| JVO-18647 | Antisense to add SPA-tag at the C-terminus of STsORF79 | ACACTTGTCCAGCGCCTCCACACCGTCAGCGGTGAATCAGAGTTCCTATTCCGAAGTTC |
| JVO-18648 | Sense to validate SPA-tag at the C-terminus of STsORF79 | TTCTTCTTTGCAATGCTTTGTC |
| JVO-18649 | Antisense to validate SPA-tag at the C-terminus of STsORF79 | AACTTGTTCAATCAATTCCGGG |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18650 | Sense to add SPA-tag at the C-terminus of STsORF80 | AGAGGCTGAGATTCTGGCGGGACCGTTAGAACCTCCAATTTCCATGGAAAAGAGAAGA |
| JVO-18651 | Antisense to add SPA-tag at the C-terminus of STsORF80 | AAAAAAGGCCGCCGAATGGCAGCCTCAAATGGAATATGTAAGTTCCTATTCCGAAGTTC |
| JVO-18652 | Sense to validate SPA-tag at the C-terminus of STsORF80 | AGTGCGTCATTGGGAATTATCAG |
| JVO-18653 | Antisense to validate SPA-tag at the C-terminus of STsORF80 | TGCTACAACACAGTAACCAGTTC |
| JVO-18654 | Sense to add SPA-tag at the C-terminus of STsORF99 | CCAGACGAGCGAAGGCGACAGCGTGATGAATGAGCAGAAATCCATGGAAAAGAGAAGA |
| JVO-18655 | Antisense to add SPA-tag at the C-terminus of STsORF99 | TTTAGCCAGGCAACATCCGGCTGCCCGTCAAGGATATGCGAGTTCCTATTCCGAAGTTC |
| JVO-18656 | Sense to validate SPA-tag at the C-terminus of STsORF99 | AGTATTGTTTTCCGGTATTGGTC |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18657 | Antisense to validate SPA-tag at the C-terminus of STsORF99 | GCGGAAATTCTTGGCGGC |
| JVO-18658 | Sense to add SPA-tag at the C-terminus of STsORF102 | GGACGCCGAAGCCCGGCTGGCAGAAGAAGAGGCCGCAGCCTCCATGGAAAAGAGAAGA |
| JVO-18659 | Antisense to add SPA-tag at the C-terminus of STsORF102 | TCAGCCGGATCGGGCGCTTGCGCCGCCATCCGGCATGATTAGTTCCTATTCCGAAGTTC |
| JVO-18660 | Sense to validate SPA-tag at the C-terminus of STsORF102 | TCCTGGCGGCGCAACAAC |
| JVO-18661 | Antisense to validate SPA-tag at the C-terminus of STsORF102 | ATGAAAGTCTCTACCAGCGC |
| JVO-18662 | Sense to add SPA-tag at the C-terminus of STsORF102 | ACCGCCATATTCACGACGGTATCATGCAATCACAGGAGTCTCCATGGAAAAGAGAAGA |
| JVO-18663 | Antisense to add SPA-tag at the C-terminus of STsORF102 | AAGATAAAAATCAATAAATTCAACAGACCAAGATATCAAAAGTTCCTATTCCGAAGTTC |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18664 | Sense to validate SPA-tag at the C-terminus of STsORF102 | CCGGGAGTTATACCTTGCC |
| JVO-18665 | Antisense to validate SPA-tag at the C-terminus of STsORF102 | CATGTGGCTGGATTTTGCCG |
| JVO-18666 | Sense to add SPA-tag at the C-terminus of STsORF102 | GATTGTCAGCAGTTTTATCTATGTGTGGGTCACGACGTATTCCATGGAAAAGAGAAGA |
| JVO-18667 | Antisense to add SPA-tag at the C-terminus of STsORF102 | ACCCGTGCCTGGGTGCTGTTGAAAAAGGCATAACAGGCGTAGTTCCTATTCCGAAGTTC |
| JVO-18668 | Sense to validate SPA-tag at the C-terminus of STsORF102 | GCCTGGCTGGTGGAAAGC |
| JVO-18669 | Antisense to validate SPA-tag at the C-terminus of STsORF102 | TCTGTTGCTGCCGGAAAGC |
| JVO-18670 | Sense to add SPA-tag at the C-terminus of STsORF132 | GGGAGTTGTGGTCTACTACATGTTGAGGAAAACGATTGGCTCCATGGAAAAGAGAAGA |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18671 | Antisense to add SPA-tag at the C-terminus of STsORF132 | TTTTTGCACTCAATCTGCGGCAAATCCGACCACTTTTTGCAGTTCCTATTCCGAAGTTC |
| JVO-18672 | Sense to validate SPA-tag at the C-terminus of STsORF132 | TCGTCATCGCTCACAAGGTC |
| JVO-18673 | Antisense to validate SPA-tag at the C-terminus of STsORF132 | AAGGATCAGCGGACAGGC |
| JVO-18674 | Sense to add SPA-tag at the C-terminus of STsORF133 | ACCCACCCCTCTCCCGCGATGGAGAATTTTCCTTTTCCGGTCCATGGAAAAGAGAAGA |
| JVO-18675 | Antisense to add SPA-tag at the C-terminus of STsORF133 | TGTCACTGTCTTACACACCGGTAAGACAGCAGAGGCAGGCAGTTCCTATTCCGAAGTTC |
| JVO-18676 | Sense to validate SPA-tag at the C-terminus of STsORF133 | GTTGATTCATCACCGCTTCG |
| JVO-18677 | Antisense to validate SPA-tag at the C-terminus of STsORF133 | AGGGTTTCTGTGCCGGAAG |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18678 | Sense to add SPA-tag at the C-terminus of STsORF138 | ACTAAAAATATCGCGAATGAGTAGCCTGAGCGCTCATATTTCCATGGAAAAGAGAAGA |
| JVO-18679 | Antisense to add SPA-tag at the C-terminus of STsORF138 | GAAACATAAGCCTCTCCTTTACCATAGTTAATGTCGACCGAGTTCCTATTCCGAAGTTC |
| JVO-18680 | Sense to validate SPA-tag at the C-terminus of STsORF138 and STsORF139 | TGCTGTTGCGACAGGTTCTG |
| JVO-18681 | Antisense to validate SPA-tag at the C-terminus of STsORF138 and STsORF139 | CGTAACCGGTATTCGCTGG |
| JVO-18682 | Sense to add SPA-tag at the C-terminus of STsORF139 | GCTCTTCCTGGTCAGCCTGTTCATGGGCCGTAAACGACCCTCCATGGAAAAGAGAAGA |
| JVO-18683 | Antisense to add SPA-tag at the C-terminus of STsORF139 | CTGGACTGGCTTGATAGCAATATATCTTTACAGATGCCTGAGTTCCTATTCCGAAGTTC |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18684 | Sense to add SPA-tag at the C-terminus of STsORF114 | GTTTAAACACGCTTTATTTCCTCCGCCGTTAACACGACGCTCCATGGAAAAGAGAAGA |
| JVO-18685 | Antisense to add SPA-tag at the C-terminus of STsORF114 | TGCTATATTTAGCACACGACAAATTTCTGCCCTGAGGCAAAGTTCCTATTCCGAAGTTC |
| JVO-18688 | Sense to validate SPA-tag at the C-terminus of STsORF114 | TTGAAAATATCTTATCGCCGTGC |
| JVO-18689 | Antisense to validate SPA-tag at the C-terminus of STsORF114 | CTCCCTCAGCGGATATTGG |
| JVO-18690 | Sense to add SPA-tag at the C-terminus of STsORF43 | AGTGGTAAGGATAATAATGTACTTCCGTTCGGAGGCACTATCCATGGAAAAGAGAAGA |
| JVO-18691 | Antisense to add SPA-tag at the C-terminus of STsORF43 | CAATGCAAAAAGAAAAATTGTTGCCATTTCTTTAAATCGTTAGTTCCTATTCCGAAGTTC |
| JVO-18694 | Sense to validate SPA-tag at the C-terminus of STsORF43 | TGTAGCGGCGGTTTGCAG |

**Table 6.9** *continued from previous page*

| Name | Purpose | Sequence (5' →3') |
|---|---|---|
| JVO-18695 | Antisense to validate SPA-tag at the C-terminus of STsORF43 | ACCAGACCTTATGGCGCG |
| JVO-18696 | Sense to add SPA-tag at the C-terminus of STsORF129 | TTTGCAGTCCGTCCAGACCGTCATGACGTCGTTGTCCGGCTCCATGGAAAAGAGAAGA |
| JVO-18697 | Antisense to add SPA-tag at the C-terminus of STsORF129 | TCGCTGTGGATAGTCAGTGAGCCAGACCGGTTTTACCGCTAGTTCCTATTCCGAAGTTC |
| JVO-18700 | Sense to validate SPA-tag at the C-terminus of STsORF129 | AGCCAGAATATTCTCGCCAG |
| JVO-18701 | Antisense to validate SPA-tag at the C-terminus of STsORF129 | TAGATAACTGGCTGGAACAGC |
| JVO-19031 | Sense to clone MgrR in pZE12 under pL | P-GATCCGTTATCCTCGCAAGAAT |
| JVO-19032 | Antisense to clone MgrR in pZE12 | tttttctagaCAAGGGGCCGCTTTTGTCTG |

## 6.4 Methods

### 6.4.1 Bacterial cells

#### Cultivation

*Salmonella enterica* serovar Typhimurium SL1344, which in this work is referred to as wild-type, was routinely grown on LB agar plates at 37°C. One colony was used to inoculate 2 ml of LB medium for an overnight culture, grown at 37°C shaking at 220 rpm. This was diluted 1:100 the following day for the main culture, grown at 37°C shaking at 220 rpm. Cells grown in this way to an $OD_{600}$ of 2.0 were considered to be in SPI-1-inducing conditions. For growth in SPI-2-inducing conditions, cells were collected when in SPI-1 conditions, washed twice with PBS and once with SPI-2 medium, diluted 1:0 in fresh SPI-2 medium and grown at 37°C shaking at 220 rpm until they reached an $OD_{600}$ of 0.3.

When required, antibiotics were added to the plates or to the liquid medium: 100 $\mu$g/ml ampicillin (Amp), 20 $\mu$g/ml chloramphenicol (Cm), 50 $\mu$g/ml kanamycin (Kan).

#### Growth curves

Cells scraped from an LB agar plate after overnight incubation at 37°C and resuspended in PBS were used to inoculate growth curves. The suspension was measured to have each well of a 96-well plate at an initial $OD_{600}$ of 0.005. The growth in either LB medium or SPI-2 medium was followed with a plate reader Infinite M Plex.

#### Preparation of electro-competent *Salmonella*

To prepare electro-competent *Salmonella*, cells were diluted 1:100 from an overnight culture and grown in LB medium at 37°C shaking at 220 rpm to an $OD_{600}$ of 0.4. Cells were then cooled on ice for 30 min, centrifuged for 20 min at 4,000 rpm, 4°C. The pellet was washed twice with ice-cold $H_2O$. After the last wash, the pellet was resuspended in 100 $\mu$l for 1 $OD_{600}$. 100 $\mu$l of cells were mixed with ~10 ng of plasmid DNA or ~1 $\mu$g of PCR in pre-cooled 1mm electroporation cuvettes. Electroporation was performed with the parameters: 200 Ω, 2.5 kV, 25 $\mu$F. Bacteria were then resuspended in 500 $\mu$l LB medium and recovered at 37°C shaking at 220 rpm for 1 h. After recovery, cells were plated on LB plates with the appropriate antibiotic.

### One-step gene inactivation or tagging

For recombination in the chromosome, the protocols were based on (Datsenko and Wanner, 2000) for gene inactivation and (Uzzau et al., 2001) for tagging. Briefly, a strain carrying the pKD46 plasmid was grown overnight at 28°C shaking at 220 rpm, and the next day diluted 1:300 in fresh LB medium with 0.2% L-arabinose. Cells were collected when reaching an $OD_{600}$ of 0.5, made electrocompetent and transformed as described in Section 6.4.1. After electroporation, cells were recovered and plated on LB/Kan plates. The clones were verified via PCR. The resistance cassette was removed as described in (Datsenko and Wanner, 2000).

### P22 transduction

P22 phages were used to transduce the genomic mutations to the recipient strain. To generate the P22 lysate, 10 ml of LB/Kan were inoculated with one colony of the strain with the mutation to be transferred, together with ~20 $\mu$l of P22 wild-type phage, and grown overnight at 37°C shaking at 220 rpm. The next day, 2 ml of culture were centrifuged for 10 min at 13,000 rpm at room temperature. The lysate (supernatant) was transferred to a glass tube with 400 $\mu$l of chloroform. The tubes were vortexed for 10 seconds and stored at 4°C.

The lysate was then transduced to the recipient strain by mixing 100 $\mu$l of the culture of the recipient strain at an $OD_{600}$ of 1 with 10 $\mu$l, 20 $\mu$l, 40 $\mu$l, or 60 $\mu$l of phage lysate. The mix was incubated at room temperature for 20 min, then the transduction was stopped by addition EGTA to a final concentration of 10 mM. Cells were recovered growing at 28°C shaking at 220 rpm with 500 $\mu$l fresh LB for 1 h, then plated on LB plates. The clones were verified via PCR.

## 6.4.2 DNA methods

### Polymerase chain reaction (PCR)

Fragments of DNA for cloning or recombination in the chromosome were amplified with Phusion DNA polymerase, while control PCRs were performed with *Taq* polymerase. If required, the purification of the product was carried out with the NucleoSpin Gel and PCR clean-up kit according to the manufacturer's instructions.

### DNA electrophoresis

PCR products or digested plasmids were separated on 1-2% (w/v) agarose gels in 1x TAE for 30-60 min at 120-160 V. DNA was mixed with 6x DNA loading buffer for loading. For staining, RedSafe was mixed to liquid agarose (10 $\mu$l for 50 ml agarose).

### Plasmid and genomic DNA preparation

Plasmid DNA was extracted with the NucleoSpit Plasmid QuickPure kit according to the manufacturer's instructions.

Genomic DNA was extracted from 200 $\mu$l of overnight culture, which were centrifuged and the pellet resuspended in 100 $\mu$l of water. The sample was then boiled at 95°C for 5 min, vortexed briefly, and centrifuged at 13,000 rpm for 5 min at room temperature. The supernatant was moved to a new tube and mixed with 1 v of chloroform, then vortexed for 30 sec. The sample was centrifuged again at 13,000 rpm for 10 min at room temperature. Approximately 75% of the supernatant, now containing DNA, was moved to a new tube and stored at $-20$°C.

### Restriction digestion and ligation

Digestion with restriction enzymes was performed using enzymes and buffers from NEB and Fermentas, following the manufacturer's instructions. Ligation was performed using the T4 DNA ligase according to the manufacturer's instructions.

## 6.4.3 Eukaryotic cells

### Cultivation

The eukaryotic cells used in this study were human cervix carcinoma (HeLa-S3, ATCC CCL-2.2) and mouse leukaemic monocyte/macrophage (RAW264.7, ATCC TIB-71). Cells were routinely grown in complete DMEM or complete RPMI, respectively (Section 6.2.1), in T-75 flasks in a 5% $CO_2$, humidified atmosphere at 37°C.

### Infection

The infection of both HeLa cells and RAW264.5 macrophages was carried out as follows. Two days before infection, 2x$10^5$ cells were seeded in six-well plates with 2 mL of

complete medium without pen/strep. On the same day, the bacterial strains required were streaked on LB agar plates, incubated overnight at 37°C, and used for overnight cultures the day before infection.

On the day of infection, the host cells were counted to calculate the number of bacteria needed to reach the required multiplicity of infection (m.o.i.). To infect epithelial cells, *Salmonella* was diluted 1:100 from the overnight culture and grown to an $OD_{600}$ of 2. A number of cells to reach an m.o.i. of 25 was collected via centrifugation at 12,000 rpm for 2 min at room temperature, and resuspended in complete DMEM. The HeLa supernatant was replaced with DMEM with bacteria, the plates centrifuged for 10 min at 250 g, room temperature, and then incubated for 30 min in 5% $CO_2$, humidified atmosphere at 37°C. The medium was then replaced with DMEM containing gentamicin at a final concentration of 50 mg/ml, marking the time point 0. After 30 min, the medium was replaced with DMEM containing gentamicin at a final concentration of 10 mg/ml and the plates were incubated until the desired time point.

The infection of RAW264.7 was carried out with an m.o.i. of 10 in complete RPMI. The necessary amount of bacteria based on the number of eukaryotic cells and the m.o.i. was collected, resuspended in RPMI medium and incubated with 10% mouse serum for 20 min at room temperature. After this, the RPMI with bacteria was used to replace the RPMI of the RAW234.7 macrophages. The following steps, starting from the centrifugation for 10 min at 250 g, room temperature, are the same as for HeLa cells infection.

### Quantification of bacterial load during infection

To quantify the amount of intracellular bacteria during infection, a colony forming units (c.f.u.) assay was performed. At the time point of interest, the cells were washed with PBS and then incubated for 5 min at room temperature with PBS containing 0.1% Triton X-100. The lysates were then diluted in PBS, plated in technical duplicates on LB plates and incubated at 37°C. The following day, the colonies were counted and the number was normalized to the number of cells in the bacterial inoculum.

*Salmonella* strains expressing green fluorescent protein (GFP) were used to quantify the amount of intracellular bacteria per single host cell. The infected cells were washed twice with PBS and collected detatching them from the plate via trypsinization. The cells were then pelleted by centrifugation for 5 min at 250 $g$ at room temperature, then washed once with ice cold PBS, and resuspended in ice cold PBS for analysis.

The fluorescence was measured with the BD Accuri C6. The analysis of changes in fluorescence during an infection time course was performed using the FlowJo software.

#### LDH assay

To evaluate the host cell death rate during infection, the levels of lactate dehydrogenase (LDH) in the supernatant were quantified with the Cytotoy96 assay according to the manufacturer's instruction.

### 6.4.4 sPepFinder

*In silico* prediction of small proteins was performed with sPepFinder (Li and Chao, 2020). These were based on training from known bacterial small proteins using features such as ribosome binding site and aminoacid composition. Cross validation in a test dataset of ten bacterial species achieved 92.8% accuracy. The predicted proteins were then filtered to exclude CDSs for which no expression could be detected in a transcriptomic dataset (Srikumar et al., 2015).

### 6.4.5 Ribo-seq

#### Experimental procedure

The ribo-seq protocol was based on the one described in Oh et al. (2011). Wild-type *Salmonella* were harvested via fast filtration after growth in SPI-1- and SPI-2-inducing conditions, as well as after growth in LB medium to an $OD_{600}$ of 0.4, and immediately frozen in liquid nitrogen. In parallel, an aliquot of cells was mixed with stop mix solution for total RNA extraction. The pellets for ribo-seq were resuspended in lysis buffer and lysed via vortexing with glass beads (10x for 30s, with cooling on ice in between each round). The samples were centrifuged at 21,000 *g* for 10 min. The supernatant was then moved to a new tube and the $A_{260}$ measured to digest 14-17 $A_{260}$ of lysate with 800 U each of micrococcal nuclease (MNase). This incubation was carried out at 25°C with 14,500 shaking for 20 min in the presence of 2 mM $CaCl_2$ and 500 U RNase inhibitor. The digestion was stopped with EGTA at a final concentration of 6 mM and immediately loaded onto 10%-55% sucrose gradients (sucrose buffer with 2 mM fresh dithiothreitol). The gradients were ultracentrifuged at 35,000 rpm, 4°C for 2 h 30 min using the SW 40 Ti rotor in a Beckman Coulter Optima L-80 XP ultracentrifuge. The 70S monosome

fractions were then collected on a Gradient Station *ip*, and the RNA extracted from the fractions using the hot phenol-chlorophorm-isoamyl alcohol (25:24:1) method, and from the total RNA sample using the hot phenol method, as already described (Sharma et al., 2007; Vasquez et al., 2014). The depletion of ribosomal RNA was performed using Dynabeads MyOne Streptavidin T1 beads and a probe set for *Salmonella* (Senterica_-riboPOOL-RP1, siTOOLs, Germany). Fragmentation of total RNA was performed with RNA Fragmentation Reagent, then fragmented RNA and monosome RNA were selected for a size between 26-34 nt on 15%PAA/7 M urea gels as already described (Ingolia et al., 2012). RNA was extracted and concentrated via precipitation in isopropanol with 15 $\mu$g GlycoBlue and dissolved in water. Preparation of libraries was performed at Vertis Biotechnologie AG (Freising, Germany) using a small RNA protocol and sequenced on a NextSeq500 (Illumina; high-output, 75 cycles) at the Core Unit SysMed of the University of Würzburg.

### Data analysis

Data analysis was performed first processing the reads with HRIBO (version 1.4.3; (Gelhausen et al., 2020)). Adapters were trimmed with cutadapt (version 2.1; (Martin, 2011)), and mapping was carried out with segemehl (version 0.3.34; (Otto et al., 2014). Removal of reads mapping to ribosomal RNA and mult-mappers was performed with SAMtools (version 1.9; (Li et al., 2009)). ORFs were called with a variation of REPARATION (Ndah et al., 2017). Subread featureCounts (version 1.6.3; (Liao et al., 2014)) was applied of quality control read count statistics for each processing step, which were analysed with FastQC (version 0.11.8; (Andrews et al., 2010)), and results were merged with MultiQC (version 1.7; (Ewels et al., 2016)). The prediction of sORFs was based on the first replicate, and a second replicate was generated and analysed to further confirm the predictions. Analysis of the predicted sORFs was carried out with the following criteria to exclude low-confidence CDSs: at least 6 RPKM in the total RNA of both replicates in at least one growth conditions, and at least 10 RPKM in one replicate of ribosome footprints in the corresponding growth conditions. These criteria were based on known small proteins (indicating translation of 356 sORFs out of 470), and then applied to STsORFs, together with visual inspection of read coverages around the putative start codons. The only STsORFs that did not pass the ribo-seq cut-offs but was still included was STsORF111, since it ranked highly in the sPepFinder predictions.

### 6.4.6 Homologue search

Identification of STsORFs homologs was performed with BLAST. The figures to show conservation of the flanking regions were generated with clinker (Gilchrist and Chooi, 2020) using the default settings. 5'000 nucleotides upstream and downstream the small protein of interest were included.

### 6.4.7 Dual RNA-seq

Dual RNA-seq data were taken from (Westermann et al., 2016) and reanalysed as described in (Westermann et al., 2016) to include STsORFs.

### 6.4.8 TraDIS

The transposon mutant library was constructed as previously described (Langridge et al., 2009) and resulted in approximately 100,000 mutants. To evaluate gene requirement during virulence, infection of RAW264.7 marcophages was carried out as described in Section 6.4.3 with few variations. The macrophages were seeded in 75 ml flasks two days before infection at a densitfy of 2 x $10^6$, and infection was carried out with an m.o.i. of 20. Genomic DNA was prepared from the inoculum and from the bacteria recovered at 20 h p.i. For this, the macrophages were washed once with PBS, then 10 ml of PBS were added and the cells scraped and collected via centrifugation for 10 min at 250 g. The pellets were pooled and resuspended in 6 ml of PBS-Triton X-100 0.1% (v/v), incubated for 10 min at room temperature with occasional vortexing, and centrifuged for 10 min at 250 g. The bacteria, now in the supernatant, were collected via centrifugation and the DNA was extracted. The phenol-chloroform method was used to extract DNA from both inoculum and bacteria after infection. The bacterial pellets were resuspended in 250 $\mu$l of a solution of 50 mM Tris-HCl and 50 mM EDTA pH8 and frozen at $-20$°C. The suspension was then defrosted, and 2.5 $\mu$g/ml lysozyme were added and incubated on ice for 45 min. After this, 2.4 $\mu$g/ml of RNaseA per OD were added and incubated for 40 min at 37°C. Approximately 333 $\mu$g/ml of proteinase K in buffer were added and incubated for 30-60 min at 50°C. The DNA was isolated by addition of 400 $\mu$l phenol/chrloroform and 300 $\mu$l water in a phase lock tube, the mix shaken and the tubes centrifuged for 15 min at 15°C. The aqueous phase was precipitated with 1.4 ml 100% ethanol with 0.1 M sodium acetate and centrifugation for 20 in at 13,000 rpm,

room temperature. The pellet was washed with 70% ethanol and resuspended in water. The sequencing of two replicates was carried out on a MiSeq sequencer (Illumina) at the Wellcome Sanger Institute, following the TraDIS dark-cycle sequencing protocol for 50 cycles (Barquist et al., 2016). Data analysis was performed with the Bio-TraDIS Toolkit (Barquist et al., 2016) and differential abundance between inoculum and after infection was quantified with the tradis_comparison.R script with edgeR (Robinson et al., 2010).

### 6.4.9 Grad-seq

Grad-seq data were taken from (Gerovac et al., 2020) and reanalysed as described in (Gerovac et al., 2020) to include STsORFs.

### 6.4.10 Coomassie staining and western blotting

Protein samples were resolved on 12% or 15% SDS-PAGE gels. The samples, dissolved in 1x protein loading buffer (0.01 OD/$\mu$l), were boiled for 5' at 95°C prior to gel loading. For each sample, 0.05 OD/sample were loaded for Coomassie staining and 0.1 OD/sample were loaded for western blotting. The run on gel was carried on for 1.5-2 h at 45 mA in 1x running buffer.

The Roti-Blue (Roth) Coomassie staining was used to stain the gels according to the manufacturer's instructions. For western blotting, the resolved samples were transferred from the gel to a PVDF membrane (Perkin Elmer) for 90 min at 0.2 mA/cm$^2$ and 4°C with a semidry blotter (Peqlab). The membrane was then blocked with 10% (w/v) milk in TBS-T by incubating for 1 h at room temperature. After this, the membrane was incubated with the primary (1 h at room temperature or overnight at 4°C) and secondary (1 h at room temperature) antibodies required, with three washes in TBS-T between the primary and secondary antibody and after the secondary antibody. The membranes were then developed with ECL chemilumiescent solution using the Fuji LAS-4000 imager.

For mass spectrometry analysis, protein samples were resolved on Bolt 4 to 12%, Bis-Tris gels and stained with SimplyBlue Coomassie according to the manufacturer's instructions.

### 6.4.11 Mass spectrometry

Tryptic peptides from the total proteome samples were separated on a liquid chromatography and electrospray electrospray ionization-based mass spectrometry, carried out by the Becher lab (University of Greifswald). For this, peptides were loaded on a self-packed analytical column (OD 360 $\mu$m, ID 100 $\mu$m, length 20 cm) filled with of Reprosil-Gold 300 C18, 5 $\mu$m material (Dr. Maisch, Ammerbuch-Entringen, Germany) and eluted by a binary nonlinear gradient of 5%–99% acetonitrile in 0.1% acetic acid over 83 min with a flow rate of 300 nl/min. LC-MS/MS analyses were performed on an LTQ Orbitrap Elite using an EASY-nLC 1200 LC system. For mass spectrometry analysis, a full scan in the Orbitrap with a resolution of 60,000 was followed by collision-induced dissociation of the twenty most abundant precursor ions. MS2 experiments were acquired in the linear ion trap.

Data analysis was performed with MaxQuant (version 1.6.2.6, (Cox and Mann, 2008)), searching against the *Salmonella* proteome (annotation as of 23/08/2018). Two peptides with at least one being unique were required for the identification of a protein, with a protein being included in downstream analysis only if detected in at least two of the three biological replicates. Two strains were compared (wild-type vs $\Delta mgrB$, $\Delta mgrB$ vs $mgrB^+$, or wild-type vs $mgrB^+$) using a student's $t$-test and a P-value $< 0.01$ based on all possible permutations.

### 6.4.12 Cell fractionation

For sub-fractionation of bacterial cells, a total of 55 OD were collected via centrifugation at 4,000 rpm, 4°C for 20 min, washed with lysis buffer and the final pellet frozen in liquid nitrogen for storage at −20°C. The pellet was then thawed, resuspended in 4 ml and the cells disrupted by sonication (20 s burst, 10 s rest, for 15 rounds). After this, the samples were centrifuged at 4,000 rpm, 4°C for 30 min to remove unbroken cells. The equivalent of 1 OD was collected from the supernatant as "whole lysate" sample. The rest of the supernatant was transferred to an ultracentrifuge tube and ultracentrifuged for 90 min at 150,000 $g$ and 4°C using an Optima XP-80 ultracentrifuge (SW 60 Ti rotor). The supernatant (cytosolic fraction) was collected and stored at −20°C, while the pellet (membrane fraction) was incubated at 4°C overnight and with gentle stirring in the same volume of lysis buffer for the isolation of the inner membrane. After solubilization, the sample was ultracentrifuged for 60 min at 150,000 $g$ and 4°C. The supernatant

(inner membrane fraction) was then collected and stored at −20°C, while the pellet was solubilized by stirring at 4°C with the same amount of lysis buffer for the isolation of the outer membrane. After solubilization, the sample was stored at −20°C.

To concentrate the samples, each fraction was treated as follows. An aliquot of 500 $\mu$l was taken and incubated for 10 min at room temperature with 1/10 v/v of sodium desoxycholate. After this, 1/10 v/v of 100% trichloroacetic acid (TCA) was added and incubated for 10 min at room temperature. The samples were then centrifuged for 10 min at full speed, 4°C. The supernatant was removed and the pellets washed three times with ice cold 80% acetone. After the last wash, the pellets were dried and resuspended in 5x protein loading buffer brought to 1x with 1 M Tris as follows: 75 $\mu$l for the cytosolic fraction, 75 $\mu$l for the inner membrane, and 100 $\mu$l for the outer membrane.

### 6.4.13 Isolation of secreted effectors

The isolation of secreted effector proteins was performed on strains grown in SPI-1-inducing conditions, that is in LB to an $OD_{600}$ of 2. At this phase, 2 ml of cells were collected via centrifugation for 10 min at full speed and 4°C. The supernatant was then collected and centrifuged again to remove remaining bacterial cells. After this, 1.6 ml of the supernatant was moved to a new tube, 0.4 ml of ice cold 25% TCA were added, and the samples were incubated on ice for 15 min. This was followed by a 30 min centrifugation at full speed, 4°C, and the pellets were then washed twice with ice cold 80% acetone. After the last wash, the pellets were dried and resuspended in 1x protein loading buffer at a final concentration of 10 OD/$\mu$l.

### 6.4.14 RNA extraction

The extraction of total RNA was performed via hot phenol. A total of 4 OD of cells were collected when grown in SPI-1 conditions, while 3 OD of cells were collected when grown in SPI-2 conditions. The cells were mixed with 0.2 volumes of stop mix solution and snap frozen in liquid nitrogen. The samples were then thawed and centrifuged at 4,000 rpm for 20 min at 4°C. The cell pellet was resuspended in 600 $\mu$l of 1x TE with 0.5 mg/ml lysozyme, and 60 $\mu$l of 10% (w/v) SDS, and incubated at 64°C for 2 min. After the incubation, the samples were put back on ice and 66 $\mu$l 3 M NaOAc pH 5.2 and 750 $\mu$l acidic phenol were added and incubated for 6 min at 64°C with occasional inversion. The tubes were then centrifuged for 15 min at 16,000 $g$ at 4°C. The aqueous phase was

moved to PLG tubes with 750 $\mu$l chroroform, shaken, and centrifuged for 15 min at 16,000 $g$ at 4°C. The aqueous phase was moved to a new 2 ml tube together with 1.4 ml of 30:1 solution and incubated for at least 1 h at −20°C. After the incubation, RNA was precipitated via centrifugation for 30 min at 16,000 $g$ and 4°C. The supernatant was discarded and the RNA pellet was washed with 350 $\mu$l 70% ethanol, after which the RNA pellet was dried and resuspended in water by shaking at 65°C and 700 rpm for 5 min.

### 6.4.15 DNA digestion from RNA samples

The removal of DNA from the RNA samples was performed by digestion with DNase I (1 U for each mg of RNA) and incubation at 37°C for 45 min. The RNA was then extracted with phenol/chlorophorm/isoamyl alcohol (P/C/I): 1 volume of acidic P/C/I was added to the DNase I-digested samples, moved to PLG tubes, shaken, and centrifuged for 15 min at 16,000 $g$ at 4°C. The aqueous phase was transferred to a new tube with 3 volumes of 30:1 solution and incubated for at least 1 h at −20°C. The RNA was collected, washed with ethanol, and resuspended in water as described for the hot phenol RNA extraction.

### 6.4.16 Northern blotting

RNA samples were resolved with denaturing polyacrylamide gel electrophoresis (PAGE) using either 4 or 6% polyacrylamide gels with 7 M urea in 1x TBE. The RNA samples (between 5 and 10 $\mu$g) were mixed with an equal volume of 2x RNA loading buffer and incubated at 95°C for 5 min prior loading. The run was performed for 2 h (6% gel) or 3 h (4% gel). After the run, the gels were transferred to Hybond membranes for 1 h at 50 V in 1x TBE via a wet blotting system. The RNA was then crosslinked to the membrane by exposure to 254 nm UV light at 0.12 J. The membranes were pre-incubated with 15 ml Roti Hybri-Quick for 30 min at 42°C, after which 2-5 pmol of [$^{32}$P]-labelled gene-specific DNA oligonucleotide was added to and hybridized rotating at 42°C overnight. The following day, the membranes were washed with 5x SSC-S, 1x SSC-S, and 0.5x SSC-S for 15 min each rotating at 42°C. After the washes, the membranes were exposed on phosphor screens.

### 6.4.17 RNA sequencing

Generation of complementary DNA (cDNA) libraries from totaal RNA samples was performed at Vertis Biotechnologie AG (Freising, Germany). The Ribo-Zero rRNA Removal Kit (Bacteria; Illumina) was used for ribosomal RNA depletion. The NextSeq 500 platform (Illumina) was used for sequencing, with approximately 20 million reads/library. Adapters were removed with FASTQ format reads using cutadapt, then fastq_quality_trimmer from the FastX suite (Version 0.3.7) was used for quality trimming. Read mapping was performed with READemption (version 0.4.5; (Förstner et al., 2014), and differential gene expression analysis was performed with edgeR (version 3.20.8; (Robinson et al., 2010)). A cut-off of at least 5 uniquely mapped reads per gene in two experiments was set.

### 6.4.18 Rifampicin assay

Rifampicin (Fluka) was added to cells grown in SPI-2 medium to an $OD_{600}$ of 0.3 to a final concentration of 500 $\mu$g/ml. At the desired time points, 10 ml of culture were taken, added to 2 ml stop mix solution, and snap frozen for RNA extraction.

### 6.4.19 Motility assay

To measure *Salmonella* motility, 6 $\mu$l of overnight cultures were spotted at the center of 0.3% SPI-2 agar plates. The swimming capacity was evaluated measuring the distance from the point of inoculation after 6 h of incubation at 37°C.

### 6.4.20 Analysis of GFP reporters

Transcriptional and translational reporters fused to GFP (pXG-1 plasmid series) were tested in response to the presence or absence of YjiS encoded on a different plasmid. The strains were grown in either SPI-1 or SPI-2 conditions and, at the appropriate cell density, the cells were collected, washed twice with ice cold PBS, and the fluorescence was measured using the BD Accuri C6. The changes in GFP levels in the presence of YjiS was calculated relative to GFP level in the strain lacking YjiS.

### 6.4.21 Formaldehyde crosslinking

For the crosslinking of SPA-tagged YjiS to its interactors, 60 OD of cells grown in SPI-2 medium were collected via centrifugation at 4,000 rpm and 4°C for 20 min. The pellet was washed once with ice cold PBS and resuspended in 1 ml of PBS, to which formaldehyde (Roth) was added to a final concentration of 1%. The cells were then incubated for 10 min at 37°C. The crosslinking was quenched by adding glycine (Roth) to a final concentration of 0.125 M. An aliquot of the solution was stored as input of the crosslinking, while the remaining sample was centrifuged for 2 min at 13,000 rpm, 4°C and resuspended in 800 $\mu$l of lysis buffer. The samples were then lysed via mechanical beating on a Mixer mill MM400 (Retsch) for 10 min at 30 Hz. The samples were centrifuged at full speed and 4°C for 30 min. Half of the supernatant was moved to a new tube and stored at −20°C, while the other half was treated with 1/s volume of RNase A/T1 mix for to min at 20°C. The treatment was stopped adding 5x protein loading buffer and boiling for 5 min at 95°C. The shift upon crosslink and treatment with RNase was evaluated via western blotting.

### 6.4.22 CoIP for protein-protein interaction

For the identification of the interactome of YjiS, a co-immunoprecipitation (coIP) was performed on a total of 60 OD of *Salmonella* wild-type or with SPA-tagged YjiS grown in SPI-2 conditions. Once at the required OD, the cells were collected via centrifugation at 4,000 rpm and 4°C for 20 min, then washed twice with ice-cold PBS. The final pellet was moved to 2 ml tubes and resuspended in 600 $\mu$l of lysis buffer for coIP, to which an equal volume of glass beads was added. The samples were then lysed via mechanical beating on a Mixer mill MM400 for 10 min at 30 Hz. After this, the samples were centrifuged at full speed, 4°C for 30 min. The supernatant was moved to a new 2 ml tube and centrifuged as before but for 10 min. The supernatant was moved again to a new tube, and an aliquot of 25 $\mu$l to which 5x protein loading buffer was added was stored at −20°C as input. 25 $\mu$l of pre-washed anti-FLAG M2 magnetic beads were added to the remaining sample. The tubes were then incubated for 16 h rotating at 4°C. After incubation, the magnetic beads were collected and washed 3 times with 500 $\mu$l of lysis buffer. To elute the proteins bound, the beads were soaked in 35 $\mu$l of 1x protein loading buffer, boiled for 5 min at 95°C, spun down and the supernatant collected. Both input and pulled-down proteins were analysed via Coomassie staining.

### 6.4.23 Data availability

The data generated in this work can be found in a USB thumb drive attached to this thesis. The data generated in the first two sections are published in Venturini et al. (2020).

# 7 Bibliography

Adams, P. P., Baniulyte, G., Esnault, C., Chegireddy, K., Singh, N., Monge, M., Dale, R. K., Storz, G., and Wade, J. T. (2021). Regulatory roles of *Escherichia coli* 5'UTR and ORF-internal RNAs detected by 3'end mapping. *Elife*, 10:e62438.

Adkins, J. N., Mottaz, H. M., Norbeck, A. D., Gustin, J. K., Rue, J., Clauss, T. R., Purvine, S. O., Rodland, K. D., Heffron, F., and Smith, R. D. (2006). Analysis of the *Salmonella* Typhimurium proteome through environmental response toward infectious conditions. *Mol. Cell. Proteom.*, 5(8):1450–1461.

Alix, E. and Blanc-Potard, A.-B. (2008). Peptide-assisted degradation of the *Salmonella* MgtC virulence factor. *EMBO J.*, 27(3):546–557.

Álvarez-Ordóñez, A., Begley, M., Prieto, M., Messens, W., López, M., Bernardo, A., and Hill, C. (2011). *Salmonella* spp. survival strategies within the host gastrointestinal tract. *Microbiology*, 157(12):3268–3281.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C., and Wingett, S. (2010). FastQC: a quality control tool for high throughput sequence data. . Babraham Institute.

Aranda, C. A., Swanson, J. A., LooMIs, W. P., and Miller, S. I. (1992). *Salmonella* Typhimurium activates virulence gene transcription within acidified macrophage phagosomes. *Proc. Natl. Acad. Sci. U.S.A.*, 89(21):10079–10083.

Aussel, L., Loiseau, L., Hajj Chehade, M., Pocachard, B., Fontecave, M., Pierrel, F., and Barras, F. (2014). ubij, a new gene required for aerobic growth and proliferation in macrophage, is involved in coenzyme Q biosynthesis in *Escherichia coli* and *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.*, 196(1):70–79.

Babu, M., Butl, G., Pogoutse, O., Li, J., Greenblatt, J. F., and Emili, A. (2009). Sequential peptide affinity purification system for the systematic isolation and identification of protein complexes from *Escherichia coli*. In *Proteomics*, pages 373–400. Springer.

Baek, C.-H., Wang, S., Roland, K. L., and Curtiss, R. (2009). Leucine-responsive reg-

ulatory protein (Lrp) acts as a virulence repressor in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.*, 191(4):1278–1292.

Baek, J., Lee, J., Yoon, K., and Lee, H. (2017). Identification of unannotated small genes in *Salmonella*. *G3 (Bethesda)*, 7(3):983–989.

Barquist, L., Langridge, G. C., Turner, D. J., Phan, M.-D., Turner, A. K., Bateman, A., Parkhill, J., Wain, J., and Gardner, P. P. (2013). A comparison of dense transposon insertion libraries in the *Salmonella* serovars Typhi and Typhimurium. *Nucleic Acids Res.*, 41(8):4549–4564.

Barquist, L., Mayho, M., Cummins, C., Cain, A. K., Boinett, C. J., Page, A. J., Langridge, G. C., Quail, M. A., Keane, J. A., and Parkhill, J. (2016). The TraDIS toolkit: sequencing and analysis for dense transposon mutant libraries. *Bioinformatics*, 32(7):1109–1111.

Bijlsma, J. J. and Groisman, E. A. (2005). The PhoP/PhoQ system controls the intramacrophage type three secretion system of *Salmonella* enterica. *Mol. Microbiol.*, 57(1):85–96.

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., et al. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science*, 277(5331):1453–1462.

Bliven, K. A. and Maurelli, A. T. (2012). Antivirulence genes: insights into pathogen evolution through gene loss. *Infect. Immun.*, 80(12):4061–4070.

Blount, Z. D. (2015). The natural history of model organisms: The unexhausted potential of *E. coli*. *Elife*, 4:e05826.

Borodovsky, M., Mclninch, J. D., Koonin, E. V., Rudd, K. E., Médigue, C., and Danchin, A. (1995). Detection of new genes in a bacterial genome using Markov models for three gene classes. *Nucleic Acids Res.*, 23(17):3554–3562.

Borodovsky, M., Rudd, K. E., and Koonin, E. V. (1994). Intrinsic and extrinsic approaches for detecting genes in a bacterial genome. *Nucleic Acids Res.*, 22(22):4756–4767.

Brenner, S., Jacob, F., and Meselson, M. (1961). An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature*, 190(4776):576–581.

Brocchieri, L. and Karlin, S. (2005). Protein length in eukaryotic and prokaryotic pro-

teomes. *Nucleic Acids Res.*, 33(10):3390–3400.

Cain, A. K., Barquist, L., Goodman, A. L., Paulsen, I. T., Parkhill, J., and van Opijnen, T. (2020). A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.*, pages 1–15.

Carsiotis, M., Weinstein, D., Karch, H., Holder, I., and O'Brien, A. (1984). Flagella of *Salmonella* typhimurium are a virulence factor in infected C57BL/6J mice. *Infect. Immun.*, 46(3):814–818.

Carvunis, A.-R., Rolland, T., Wapinski, I., Calderwood, M. A., Yildirim, M. A., Simonis, N., Charloteaux, B., Hidalgo, C. A., Barbette, J., Santhanam, B., et al. (2012). Proto-genes and de novo gene birth. *Nature*, 487(7407):370–374.

Cassidy, L., Helbig, A. O., Kaulich, P. T., Weidenbach, K., Schmitz, R. A., and Tholey, A. (2020). Multidimensional separation schemes enhance the identification and molecular characterization of low molecular weight proteomes and short open reading frame-encoded peptides in top-down proteomics. *J Proteomics*, page 103988.

Chaudhuri, R. R., Morgan, E., Peters, S. E., Pleasance, S. J., Hudson, D. L., Davies, H. M., Wang, J., van Diemen, P. M., Buckley, A. M., Bowen, A. J., et al. (2013). Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet.*, 9(4):e1003456.

Cherepanov, P. P. and Wackernagel, W. (1995). Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*, 158(1):9–14.

Cho, J., Carr, A. N., Whitworth, L., Johnson, B., and Wilson, K. S. (2017). MazEF toxin-antitoxin proteins alter *Escherichia coli* cell morphology and infrastructure during persister formation and regrowth. *Microbiology*, 163(3):308–321.

Choi, E., Lee, K.-Y., and Shin, D. (2012). The MgtR regulatory peptide negatively controls expression of the MgtA $Mg^{2+}$ transporter in *Salmonella* enterica serovar Typhimurium. *Biochem. Biophys. Res. Commun.*, 417(1):318–323.

Clark, L., Perrett, C. A., Malt, L., Harward, C., Humphrey, S., Jepson, K. A., Martinez-Argudo, I., Carney, L. J., La Ragione, R. M., Humphrey, T. J., et al. (2011). Differences in *Salmonella enterica* serovar Typhimurium strain invasiveness are associated with heterogeneity in spi-1 gene expression. *Microbiology*, 157(Pt 7):2072.

Colgan, A. M., Kröger, C., Diard, M., Hardt, W.-D., Puente, J. L., Sivasankaran, S. K.,

Hokamp, K., and Hinton, J. C. (2016). The impact of 18 ancestral and horizontally-acquired regulatory proteins upon the transcriptome and sRNA landscape of *Salmonella enterica* serovar Typhimurium. *PLoS Genet.*, 12(8):e1006258.

Cox, J. and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, 26(12):1367–1372.

Dalebroux, Z. D. and Miller, S. I. (2014). *Salmonellae* PhoPQ regulation of the outer membrane to resist innate immunity. *Curr. Opin. Microbiol.*, 17:106–113.

Darfeuille, F., Unoson, C., Vogel, J., and Wagner, E. G. H. (2007). An antisense rna inhibits translation by competing with standby ribosomes. *Mol. Cell*, 26(3):381–392.

Darwin, A. J. (2013). Stress relief during host infection: the phage shock protein response supports bacterial virulence in various ways. *PLoS Pathog.*, 9(7):e1003388.

Das, A., Urbanowski, J., Weissbach, H., Nestor, J., and Yanofsky, C. (1983). *In vitro* synthesis of the tryptophan operon leader peptides of *Escherichia coli*, *Serratia marcescens*, and *Salmonella* Typhimurium. *Proc. Natl. Acad. Sci. U.S.A.*, 80(10):2879–2883.

Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.*, 97(12):6640–6645.

Edelmann, D. and Berghoff, B. A. (2019). Type I toxin-dependent generation of superoxide affects the persister life cycle of *Escherichia coli*. *Sci. Rep.*, 9(1):1–10.

Eguchi, Y., Ishii, E., Yamane, M., and Utsumi, R. (2012). The connector SafA interacts with the multi-sensing domain of PhoQ in *Escherichia coli*. *Mol. Microbiol.*, 85(2):299–313.

Ekman, D. and Elofsson, A. (2010). Identifying and quantifying orphan protein sequences in fungi. *J. Mol. Biol.*, 396(2):396–405.

Erickson, H. P. (2009). Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol. Proced. Online*, 11(1):32–51.

Eswarappa, S. M., Karnam, G., Nagarajan, A. G., Chakraborty, S., and Chakravortty, D. (2009). lac repressor is an antivirulence factor of *Salmonella* enterica: its role in the evolution of virulence in *Salmonella*. *PLoS One*, 4(6):e5789.

Ewels, P., Magnusson, M., Lundin, S., and Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047–3048.

Fàbrega, A. and Vila, J. (2013). *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin. Microbiol. Rev.*, 26(2):308–341.

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

Fisher, R. A., Gollan, B., and Helaine, S. (2017). Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.*, 15(8):453.

Flieger, A., Frischknecht, F., Häcker, G., Hornef, M. W., and Pradel, G. (2018). Pathways of host cell exit by intracellular pathogens. *Microb. Cell*, 5(12):525.

Flores-Kim, J. and Darwin, A. J. (2016). The phage shock protein response. *Annu. Rev. Microbiol.*, 70:83–101.

Fontaine, F., Fuchs, R. T., and Storz, G. (2011). Membrane localization of small proteins in *Escherichia coli*. *J. Biol. Chem.*, 286(37):32464–32474.

Förstner, K. U., Vogel, J., and Sharma, C. M. (2014). READemption—a tool for the computational analysis of deep-sequencing–based transcriptome data. *Bioinformatics*, 30(23):3421–3423.

Fröhlich, K. S., Haneke, K., Papenfort, K., and Vogel, J. (2016). The target spectrum of SdsR small RNA in *Salmonella*. *Nucleic Acids Res.*, 44(21):10406–10422.

Gagarinova, A., Stewart, G., Samanfar, B., Phanse, S., White, C. A., Aoki, H., Deineko, V., Beloglazova, N., Yakunin, A. F., Golshani, A., et al. (2016). Systematic genetic screens reveal the dynamic global functional organization of the bacterial translation machinery. *Cell Rep.*, 17(3):904–916.

Gao, R. and Stock, A. M. (2018). Overcoming the cost of positive autoregulation by accelerating the response with a coupled negative feedback. *Cell Rep.*, 24(11):3061–3071.

Garcia Vescovi, E. (1996). $Mg^{2+}$ as an extracellular signal: Environmental regulation of *Salmonella* virulence. *Cell*, 84:165–174.

Gelhausen, R., Heyl, F., Svensson, S. L., Froschauer, K., Hadjeras, L., Sharma, C. M., Eggenhofer, F., and Backofen, R. (2020). HRIBO-High-throughput analysis of bacte-

rial ribosome profiling data. *BioRxiv*.

Gerdes, K., Gultyaev, A. P., Franch, T., Pedersen, K., and Mikkelsen, N. D. (1997). Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.*, 31(1):1–31.

Gerdes, K., Larsen, J., and Molin, S. (1985). Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.*, 161(1):292–298.

Gerdes, K. and Maisonneuve, E. (2012). Bacterial persistence and toxin-antitoxin loci. *Annu. Rev. Microbiol.*, 66:103–123.

Gerovac, M., El Mouali, Y., Kuper, J., Kisker, C., Barquist, L., and Vogel, J. (2020). Global discovery of bacterial RNA-binding proteins by RNase-sensitive gradient profiles reports a new FinO domain protein. *RNA*, 26(10):1448–1463.

Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G. R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.*, 21(5):566–569.

Gilchrist, C. L. M. and Chooi, Y.-H. H. (2020). clinker & clustermap. js: Automatic generation of gene cluster comparison figures. *Bioinformatics*.

Gimpel, M. and Brantl, S. (2017). Dual-function small regulatory RNAs in bacteria. *Mol. Microbiol.*, 103(3):387–397.

Groisman, E. A. (2001). The pleiotropic two-component regulatory system PhoP-PhoQ. *J. Bacteriol.*, 183(6):1835–1842.

Grützner, J., Billenkamp, F., Spanka, D.-T., Rick, T., Monzon, V., Förstner, K. U., and Klug, G. (2021). The small DUF1127 protein CcaF1 from *Rhodobacter sphaeroides* is an RNA-binding protein involved in sRNA maturation and RNA turnover. *Nucleic Acids Res.*

Guerreiro, D. N., Arcari, T., and O'Byrne, C. P. (2020). The $\sigma$b-mediated general stress response of *Listeria monocytogenes*: Life and death decision making in a pathogen. *Front. microbiol.*, 11:1505.

Gupta, R. K., Luong, T. T., and Lee, C. Y. (2015). RNAIII of the *Staphylococcus aureus* agr system activates global regulator MgrA by stabilizing mRNA. *Proc. Natl. Acad. Sci. U.S.A.*, 112(45):14036–14041.

Hassani, A. S., Amirmozafari, N., and Ghaemi, A. (2009). Virulence increasing of *Salmonella* typhimurium in Balb/c mice after heat-stress induction of phage shock protein

A. *Curr. Microbiol.*, 59(4):446–450.

Hautefort, I., Thompson, A., Eriksson-Ygberg, S., Parker, M., Lucchini, S., Danino, V., Bongaerts, R., Ahmad, N., Rhen, M., and Hinton, J. (2008). During infection of epithelial cells *Salmonella enterica* serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cell. Microbiol.*, 10(4):958–984.

Hébrard, M., Kröger, C., Sivasankaran, S. K., Händler, K., and Hinton, J. C. (2011). The challenge of relating gene expression to the virulence of *Salmonella enterica* serovar Typhimurium. *Curr. Opin. Biotechnol.*, 22(2):200–210.

Hemm, M. R., Paul, B. J., Miranda-Ríos, J., Zhang, A., Soltanzad, N., and Storz, G. (2010). Small stress response proteins in *Escherichia coli*: proteins missed by classical proteomic studies. *J. Bacteriol.*, 192(1):46–58.

Hemm, M. R., Paul, B. J., Schneider, T. D., Storz, G., and Rudd, K. E. (2008). Small membrane proteins found by comparative genomics and ribosome binding site models. *Mol. Microbiol.*, 70(6):1487–1501.

Hemm, M. R., Weaver, J., and Storz, G. (2020). *Escherichia coli* small proteome. *EcoSal Plus*, 9(1).

Henry, T., Garcia-Del Portillo, F., and Gorvel, J. (2005). Identification of *Salmonella* functions critical for bacterial cell division within eukaryotic cells. *Mol. Microbiol.*, 56(1):252–267.

Holmqvist, E., Li, L., Bischler, T., Barquist, L., and Vogel, J. (2018). Global maps of ProQ binding in vivo reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol. Cell*, 70(5):971–982.

Hör, J., Di Giorgio, S., Gerovac, M., Venturini, E., Förstner, K. U., and Vogel, J. (2020a). Grad-seq shines light on unrecognized RNA and protein complexes in the model bacterium *Escherichia coli*. *Nucleic Acids Res.*, 48(16):9301–9319.

Hör, J., Matera, G., Vogel, J., Gottesman, S., and Storz, G. (2020b). Trans-acting small RNAs and their effects on gene expression in *Escherichia coli* and *Salmonella enterica*. *EcoSal Plus*, 9(1).

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., Lina, G., Etienne, J., Ehresmann, B., Ehresmann, C., et al. (2005). *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate spa gene expression.

*EMBO J.*, 24(4):824–835.

Ibarra, J. A. and Steele-Mortimer, O. (2009). *Salmonella*–the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. *Cell. Microbiol.*, 11(11):1579–1586.

Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A.-E., and Vogel, J. (2020). Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.*, 5(10):1202–1206.

Impens, F., Rolhion, N., Radoshevich, L., Bécavin, C., Duval, M., Mellin, J., Del Portillo, F. G., Pucciarelli, M. G., Williams, A. H., and Cossart, P. (2017). N-terminomics identifies Prli42 as a membrane miniprotein conserved in Firmicutes and critical for stressosome activation in *Listeria monocytogenes*. *Nat. Microbiol.*, 2(5):1–12.

Ingolia, N. T., Brar, G. A., Rouskin, S., McGeachy, A. M., and Weissman, J. S. (2012). The ribosome profiling strategy for monitoring translation *in vivo* by deep sequencing of ribosome-protected mRNA fragments. *Nat. Protoc.*, 7(8):1534–1550.

Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., and Weissman, J. S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 324(5924):218–223.

Ishii, E., Eguchi, Y., and Utsumi, R. (2013). Mechanism of activation of PhoQ/PhoP two-component signal transduction by SafA, an auxiliary protein of PhoQ histidine kinase in *Escherichia coli*. *Biosci. Biotechnol. Biochem.*, 77(4):814–819.

Janzon, L., Löfdahl, S., and Arvidson, S. (1989). Identification and nucleotide sequence of the delta-lysin gene, hld, adjacent to the accessory gene regulator (agr) of *Staphylococcus aureus*. *MGG*, 219(3):480–485.

Karlinsey, J. E., Maguire, M. E., Becker, L. A., Crouch, M.-L. V., and Fang, F. C. (2010). The phage shock protein PspA facilitates divalent metal transport and is required for virulence of *Salmonella enterica* sv. Typhimurium. *Mol. Microbiol.*, 78(3):669–685.

Kato, A., Tanabe, H., and Utsumi, R. (1999). Molecular characterization of the PhoP-PhoQ two-component system in *Escherichia coli* K-12: identification of extracellular $Mg^{2+}$-responsive promoters. *J. Bacteriol.*, 181(17):5516–5520.

Kraus, A., Weskamp, M., Zierles, J., Balzer, M., Busch, R., Eisfeld, J., Lambertz, J., Nowaczyk, M. M., and Narberhaus, F. (2020). Arginine-rich small proteins with a domain of unknown function, DUF1127, play a role in phosphate and carbon metabolism

of *Agrobacterium tumefaciens*. *J. Bacteriol.*, 202(22).

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S. K., Hammarlöf, D. L., Canals, R., Grissom, J. E., Conway, T., Hokamp, K., et al. (2013). An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. *Cell Host Microbe*, 14(6):683–695.

Langridge, G. C., Phan, M.-D., Turner, D. J., Perkins, T. T., Parts, L., Haase, J., Charles, I., Maskell, D. J., Peters, S. E., Dougan, G., et al. (2009). Simultaneous assay of every *Salmonella* typhi gene using one million transposon mutants. *Genome res.*, 19(12):2308–2316.

LaRock, D. L., Chaudhary, A., and Miller, S. I. (2015). *Salmonellae* interactions with host processes. *Nat. Rev. Microbiol.*, 13(4):191–205.

Laub, M. T. and Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annu. Rev. Genet.*, 41:121–145.

Lee, E.-J. and Groisman, E. A. (2012a). Control of a *Salmonella* virulence locus by an ATP-sensing leader messenger RNA. *Nature*, 486(7402):271–275.

Lee, E.-J. and Groisman, E. A. (2012b). Tandem attenuators control expression of the *Salmonella mgtCBR* virulence operon. *Mol. Microbiol.*, 86(1):212–224.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16):2078–2079.

Li, L. and Chao, Y. (2020). sPepFinder expedites genome-wide identification of small proteins in bacteria. *bioRxiv*.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930.

Lippa, A. M. and Goulian, M. (2009). Feedback inhibition in the PhoQ/PhoP signaling system by a membrane peptide. *PLoS Genet.*, 5(12):e1000788.

Lloyd, C. R., Park, S., Fei, J., and Vanderpool, C. K. (2017). The small protein SgrT controls transport activity of the glucose-specific phosphotransferase system. *J. Bacteriol.*, 199(11).

Mangan, M. W., Lucchini, S., Danino, V., Cróinín, T. Ó., Hinton, J. C., and Dorman, C. J. (2006). The integration host factor (IHF) integrates stationary-phase and viru-

lence gene expression in *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.*, 59(6):1831–1847.

Mann, B., van Opijnen, T., Wang, J., Obert, C., Wang, Y.-D., Carter, R., McGoldrick, D. J., Ridout, G., Camilli, A., Tuomanen, E. I., et al. (2012). Control of virulence by small RNAs in *Streptococcus pneumoniae*. *PLoS Pathog.*, 8(7):e1002788.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1):10–12.

Meganathan, R. (2001). Ubiquinone biosynthesis in microorganisms. *FEMS Microbiol. Lett.*, 203(2):131–139.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., Altuvia, Y., Argaman, L., and Margalit, H. (2016). Global mapping of small RNA-target interactions in bacteria. *Mol. Cell*, 63(5):884–897.

Melior, H., Li, S., Madhugiri, R., Stötzel, M., Azarderakhsh, S., Barth-Weber, S., Baumgardt, K., Ziebuhr, J., and Evguenieva-Hackenberg, E. (2019). Transcription attenuation-derived small RNA rnTrpL regulates tryptophan biosynthesis gene expression in trans. *Nucleic Acids Res.*, 47(12):6396–6410.

Melior, H., Maaß, S., Li, S., Förstner, K. U., Azarderakhsh, S., Varadarajan, A. R., Stötzel, M., Elhossary, M., Barth-Weber, S., Ahrens, C. H., et al. (2020). The leader peptide peTrpL forms sntibiotic-containing ribonucleoprotein complexes for posttranscriptional regulation of multiresistance genes. *MBio*, 11(3).

Meselson, M. and Stahl, F. W. (1958). The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 44(7):671–682.

Meydan, S., Marks, J., Klepacki, D., Sharma, V., Baranov, P. V., Firth, A. E., Margus, T., Kefi, A., Vázquez-Laslop, N., and Mankin, A. S. (2019). Retapamulin-assisted ribosome profiling reveals the alternative bacterial proteome. *Mol. Cell*, 74(3):481–493.

Miller, S., Pulkkinen, W., Selsted, M., and Mekalanos, J. (1990). Characterization of defensin resistance phenotypes associated with mutations in the *phoP* virulence regulon of *Salmonella* Typhimurium. *Infect. Immun.*, 58(11):3706–3710.

Miozzari, G. and Yanofsky, C. (1978). Translation of the leader region of the *Escherichia coli* tryptophan operon. *J. Bacteriol.*, 133(3):1457–1466.

Miravet-Verde, S., Ferrar, T., Espadas-García, G., Mazzolini, R., Gharrab, A., Sabido,

E., Serrano, L., and Lluch-Senar, M. (2019). Unraveling the hidden universe of small proteins in bacterial genomes. *Mol. Syst. Biol.*, 15(2):e8290.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L., Tosatto, S. C., Paladin, L., Raj, S., Richardson, L. J., et al. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Res.*, 49(D1):D412–D419.

Mohammad, F., Green, R., and Buskirk, A. R. (2019). A systematically-revised ribosome profiling method for bacteria reveals pauses at single-codon resolution. *Elife*, 8:e42591.

Monsieurs, P., De Keersmaecker, S., Navarre, W. W., Bader, M. W., De Smet, F., McClelland, M., Fang, F. C., De Moor, B., Vanderleyden, J., and Marchal, K. (2005). Comparison of the PhoPQ regulon in *Escherichia coli* and *Salmonella* typhimurium. *J. Mol. Evol.*, 60(4):462–474.

Moon, K. and Gottesman, S. (2009). A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides. *Mol. Microbiol.*, 74(6):1314–1330.

Ndah, E., Jonckheere, V., Giess, A., Valen, E., Menschaert, G., and Van Damme, P. (2017). Reparation: ribosome profiling assisted (re-) annotation of bacterial genomes. *Nucleic Acids Res.*, 45(20):e168–e168.

Neuhaus, K., Landstorfer, R., Simon, S., Schober, S., Wright, P. R., Smith, C., Backofen, R., Wecko, R., Keim, D. A., and Scherer, S. (2017). Differentiation of ncRNAs from small mRNAs in *Escherichia coli* O157: H7 EDL933 (EHEC) by combined RNAseq and RIBOseq–*ryhB* encodes the regulatory RNA RyhB and a peptide, RyhP. *BMC Genomics*, 18(1):216.

Novick, R. P., Ross, H., Projan, S., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993). Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J.*, 12(10):3967–3975.

Oh, E., Becker, A. H., Sandikci, A., Huber, D., Chaba, R., Gloge, F., Nichols, R. J., Typas, A., Gross, C. A., Kramer, G., et al. (2011). Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor *in vivo*. *Cell*, 147(6):1295–1308.

Orr, M. W., Mao, Y., Storz, G., and Qian, S.-B. (2020). Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res.*, 48(3):1029–1042.

Otto, C., Stadler, P. F., and Hoffmann, S. (2014). Lacking alignments? The next-generation sequencing mapper segemehl revisited. *Bioinformatics*, 30(13):1837–1843.

Papenfort, K., Podkaminski, D., Hinton, J. C., and Vogel, J. (2012). The ancestral

SgrS RNA discriminates horizontally acquired *Salmonella* mrnas through a single GU wobble pair. *Proc. Natl. Acad. Sci. U.S.A.*, 109(13):E757–E764.

Papenfort, K., Said, N., Welsink, T., Lucchini, S., Hinton, J. C., and Vogel, J. (2009). Specific and pleiotropic patterns of mRNA regulation by ArcZ, a conserved, Hfq-dependent small RNA. *Mol. Microbiol.*, 74(1):139–158.

Papenfort, K., Sun, Y., Miyakoshi, M., Vanderpool, C. K., and Vogel, J. (2013). Small RNA-mediated activation of sugar phosphatase mrna regulates glucose homeostasis. *Cell*, 153(2):426–437.

Park, S.-Y., Cromie, M. J., Lee, E.-J., and Groisman, E. A. (2010). A bacterial mRNA leader that employs different mechanisms to sense disparate intracellular signals. *Cell*, 142(5):737–748.

Parkhill, J., Dougan, G., James, K., Thomson, N., Pickard, D., Wain, J., Churcher, C., Mungall, K., Bentley, S., Holden, M., et al. (2001). Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature*, 413(6858):848–852.

Pérez-Morales, D., Banda, M. M., Chau, N. E., Salgado, H., Martínez-Flores, I., Ibarra, J. A., Ilyas, B., Coombes, B. K., and Bustamante, V. H. (2017). The transcriptional regulator SsrB is involved in a molecular switch controlling virulence lifestyles of salmonella. *PLoS Pathog.*, 13(7):e1006497.

Poptsova, M. S. and Gogarten, J. P. (2010). Using comparative genome analysis to identify problems in annotated microbial genomes. *Microbiology*, 156(7):1909–1917.

Rice, J. B., Balasubramanian, D., and Vanderpool, C. K. (2012). Small RNA binding-site multiplicity involved in translational regulation of a polycistronic mrna. *Proc. Natl. Acad. Sci. U.S.A.*, 109(40):E2691–E2698.

Rice, J. B. and Vanderpool, C. K. (2011). The small RNA SgrS controls sugar–phosphate accumulation by regulating multiple PTS genes. *Nucleic Acids Res.*, 39(9):3806–3819.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140.

Roesser, J. R. and Yanofsky, C. (1991). The effects of leader peptide sequence and length on attenuation control of the trp operon of *E. coli*. *Nucleic Acids Res.*, 19(4):795–800.

Salazar, M. E., Podgornaia, A. I., and Laub, M. T. (2016). The small membrane protein

mgrb regulates PhoQ bifunctionality to control PhoP target gene expression dynamics. *Mol. Microbiol.*, 102(3):430–445.

Samayoa, J., Yildiz, F. H., and Karplus, K. (2011). Identification of prokaryotic small proteins using a comparative genomic approach. *Bioinformatics*, 27(13):1765–1771.

Sano, G.-i., Takada, Y., Goto, S., Maruyama, K., Shindo, Y., Oka, K., Matsui, H., and Matsuo, K. (2007). Flagella facilitate escape of *Salmonella* from oncotic macrophages. *J. Bacteriol.*, 189(22):8224–8232.

Santos, S. C., Bischler, T., Westermann, A. J., and Vogel, J. (2021). MAPS integrates regulation of actin-targeting effector SteC into the virulence control network of *Salmonella* small RNA PinT. *Cell Rep.*, 34(5):108722.

Sberro, H., Fremin, B. J., Zlitni, S., Edfors, F., Greenfield, N., Snyder, M. P., Pavlopoulos, G. A., Kyrpides, N. C., and Bhatt, A. S. (2019). Large-scale analyses of human microbiomes reveal thousands of small, novel genes. *Cell*, 178(5):1245–1259.

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, 9(7):671–675.

Sharma, C. M., Darfeuille, F., Plantinga, T. H., and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C/A-rich elements inside and upstream of ribosome-binding sites. *Genes & development*, 21(21):2804–2817.

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

Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007). The RNA chaperone Hfq is essential for the virulence of *Salmonella* typhimurium. *Mol. Microbiol.*, 63(1):193–217.

Smirnov, A., Förstner, K. U., Holmqvist, E., Otto, A., Günster, R., Becher, D., Reinhardt, R., and Vogel, J. (2016). Grad-seq guides the discovery of ProQ as a major small RNA-binding protein. *Proc. Natl. Acad. Sci. USA*, 113(41):11591–11596.

Srikumar, S., Kröger, C., Hebrard, M., Colgan, A., Owen, S. V., Sivasankaran, S. K., Cameron, A. D., Hokamp, K., and Hinton, J. C. (2015). RNA-seq brings new insights to the intra-macrophage transcriptome of *Salmonella* Typhimurium. *PLoS Pathog.*, 11(11):e1005262.

Stapels, D. A., Hill, P. W., Westermann, A. J., Fisher, R. A., Thurston, T. L., Saliba, A.-

E., Blommestein, I., Vogel, J., and Helaine, S. (2018). *Salmonella* persisters undermine host immune defenses during antibiotic treatment. *Science*, 362(6419):1156–1160.

Storz, G., Wolf, Y. I., and Ramamurthi, K. S. (2014). Small proteins can no longer be ignored. *Annu. Rev. Biochem.*, 83:753–777.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.*, 14(2):178–192.

Unoson, C. and Wagner, E. G. H. (2008). A small SOS-induced toxin is targeted against the inner membrane in *Escherichia coli*. *Mol. Microbiol.*, 70(1):258–270.

Urban, J. H. and Vogel, J. (2007). Translational control and target recognition by *Escherichia coli* small RNAs *in vivo*. *Nucleic Acids Res.*, 35(3):1018–1037.

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

Vakulskas, C. A., Potts, A. H., Babitzke, P., Ahmer, B. M., and Romeo, T. (2015). Regulation of bacterial virulence by Csr (Rsm) systems. *Microbiol. Mol. Biol. Rev.*, 79(2):193–224.

Valdivia, R. H. and Falkow, S. (1997). Fluorescence-based isolation of bacterial genes expressed within host cells. *Science*, 277(5334):2007–2011.

Vanderpool, C. K. and Gottesman, S. (2004). Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Mol. Microbiol.*, 54(4):1076–1089.

Vasquez, J.-J., Hon, C.-C., Vanselow, J. T., Schlosser, A., and Siegel, T. N. (2014). Comparative ribosome profiling reveals extensive translational complexity in different Trypanosoma brucei life cycle stages. *Nucleic Acids Res.*, 42(6):3623–3637.

Venturini, E., Svensson, S. L., Maaß, S., Gelhausen, R., Eggenhofer, F., Li, L., Cain, A. K., Parkhill, J., Becher, D., Backofen, R., et al. (2020). A global data-driven census of *Salmonella* small proteins and their potential functions in bacterial virulence. *microLife*, 1(1):uqaa002.

Verdon, J., Girardin, N., Lacombe, C., Berjeaud, J.-M., and Héchard, Y. (2009). $\delta$-

hemolysin, an update on a membrane-interacting peptide. *Peptides*, 30(4):817–823.

Vogel, J. (2009). A rough guide to the non-coding rna world of salmonella. *Mol. Microbiol.*, 71(1):1–11.

Vogel, J., Argaman, L., Wagner, E. G. H., and Altuvia, S. (2004). The small RNA IstR inhibits synthesis of an SOS-induced toxic peptide. *Curr. Biol.*, 14(24):2271–2276.

Wada, A. (1986). Analysis of *Escherichia coli* ribosomal proteins by an improved two dimensional gel electrophoresis. *J. Biochem.*, 100(6):1595–1605.

Weaver, J., Mohammad, F., Buskirk, A. R., and Storz, G. (2019). Identifying small proteins by ribosome profiling with stalled initiation complexes. *MBio*, 10(2).

Weinstein, D., Carsiotis, M., Lissner, C., and O'brien, A. (1984). Flagella help *Salmonella* typhimurium survive within murine macrophages. *Salmonella*, 46(3):819–825.

Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., Müller, L., Reinhardt, R., Stadler, P. F., and Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature*, 529(7587):496–501.

Westermann, A. J., Gorski, S. A., and Vogel, J. (2012). Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.*, 10(9):618–630.

Westermann, A. J., Venturini, E., Sellin, M. E., Förstner, K. U., Hardt, W.-D., and Vogel, J. (2019). The major RNA-binding protein ProQ impacts virulence gene expression in *Salmonella enterica* serovar Typhimurium. *MBio*, 10(1).

Westermann, A. J. and Vogel, J. (2021). Cross-species RNA-seq for deciphering host–microbe interactions. *Nat. Rev. Genet.*, pages 1–18.

Williams, A. H., Redzej, A., Rolhion, N., Costa, T. R., Rifflet, A., Waksman, G., and Cossart, P. (2019). The cryo-electron microscopy supramolecular structure of the bacterial stressosome unveils its mechanism of activation. *Nat. Commun.*, 10(1):1–10.

Wilmaerts, D., Bayoumi, M., Dewachter, L., Knapen, W., Mika, J. T., Hofkens, J., Dedecker, P., Maglia, G., Verstraeten, N., and Michiels, J. (2018). The persistence-inducing toxin HokB forms dynamic pores that cause ATP leakage. *MBio*, 9(4).

Wilmaerts, D., Dewachter, L., De Loose, P.-J., Bollen, C., Verstraeten, N., and Michiels, J. (2019). HokB monomerization and membrane repolarization control persister awakening. *Mol. Cell*, 75(5):1031–1042.

Wolf, Y. I., Novichkov, P. S., Karev, G. P., Koonin, E. V., and Lipman, D. J. (2009). The universal distribution of evolutionary rates of genes and distinct characteristics of eu-

karyotic genes of different apparent ages. *Proc. Natl. Acad. Sci. U.S.A.*, 106(18):7273.

Yadavalli, S. S., Goh, T., Carey, J. N., Malengo, G., Vellappan, S., Nickels, B. E., Sourjik, V., Goulian, M., and Yuan, J. (2020). Functional determinants of a small protein controlling a broadly conserved bacterial sensor kinase. *J. Bacteriol.*

Yeom, J., Shao, Y., and Groisman, E. A. (2020). Small proteins regulate *Salmonella* survival inside macrophages by controlling degradation of a magnesium transporter. *Proc. Natl. Acad. Sci. U.S.A.*, 117(33):20235–20243.

Yim, H. H. and Villarejo, M. (1992). *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. *J. Bacteriol.*, 174(11):3637–3644.

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

Yoshitani, K., Ishii, E., Taniguchi, K., Sugimoto, H., Shiro, Y., Akiyama, Y., Kato, A., Utsumi, R., and Eguchi, Y. (2019). Identification of an internal cavity in the PhoQ sensor domain for PhoQ activity and SafA-mediated control. *Biosci. Biotechnol. Biochem.*, 83(4):684–694.

Yuan, P., D'Lima, N. G., and Slavoff, S. A. (2017). Comparative membrane proteomics reveals a nonannotated *E. coli* heat shock protein. *Biochemistry*, 57(1):56–60.

Zenk, S. F., Jantsch, J., and Hensel, M. (2009). Role of *Salmonella enterica* lipopolysaccharide in activation of dendritic cell functions and bacterial containment. *J. Immunol.*, 183(4):2697–2707.

# 8 Appendix

**Table 8.1: List of STsORFs annotated in this study.** For each protein is reporter the tag, the length (aa), the detection with Ribo-seq or sPepFinder (indicated with y), and the aa sequence in the line below.

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| **STsORF1** | 16 | y | |
| MLLALRLLGLYHHVMD | | | |
| **STsORF2** | 32 | | y |
| VLSFYATRHHPCTKQVHSRPHTIILIKTKELK | | | |
| **STsORF3** | 35 | | y |
| MSFNYCSIRKNPLHFSLFFCKALKVIIPSVNEHSD | | | |
| **STsORF5** | 41 | | y |
| MLCANFCVFIFMALVSNIPCSSGKYLNITGWLHYKVRGTVS | | | |
| **STsORF6** | 26 | | y |
| MYRGPAGVVVVAPELLPCADIFPWRQ | | | |
| **STsORF7** | 31 | | y |
| MKILNNNTIIVGMEIHLLAKLPHVRFHFYVM | | | |
| **STsORF8** | 24 | | y |
| MERCCKKKMKMNLYPTWIKPVIPE | | | |
| **STsORF9** | 24 | | y |
| MALLQNRDGYERSGNLKRKIKHSP | | | |
| **STsORF10** | 26 | | y |
| VRGVCRVNRGLISKYYVFETVELKKR | | | |
| **STsORF11** | 25 | | y |
| MKNDKTEMESFASAWAFFRIGMPGR | | | |
| **STsORF12** | 16 | | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MLRDNWYNQCKNKTIT | | | |
| **STsORF13** | 32 | | y |
| MMWVRVLLLVILACLLIFMWRIPVIDEKQQKR | | | |
| **STsORF14** | 17 | | y |
| MNKFTSLLAVVFLFQLA | | | |
| **STsORF15** | 14 | | y |
| MANGVPVLNVIYFQ | | | |
| **STsORF16** | 14 | | y |
| VVNFDLFHKALDGF | | | |
| **STsORF17** | 17 | | y |
| MMFAHRNSFDFHFFNAR | | | |
| **STsORF18** | 16 | | y |
| MCIPDSQRRCSPLINC | | | |
| **STsORF19** | 41 | | y |
| MYKIISIDNLQKKSVSYLAIIEHQCGMFFTGCGAVKNAQNV | | | |
| **STsORF20** | 33 | | y |
| MFVTFHHYCFLPEGIVRWKAANKFVLGHACDPF | | | |
| **STsORF21** | 46 | y | |
| MIHIVAMQYTKSFKLHRGGKAANPQELTSVSDWGVQPQPTKRQLRG | | | |
| **STsORF22** | 24 | | y |
| MNYCKILCCVNPRRRTLFIPYLRI | | | |
| **STsORF23** | 42 | y | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MNEFKRCIRVFSHSPFKVRLMLLSMLCDMINGKPEQDNPSTK | | | |
| **STsORF24** | 46 | | y |
| MPARSLCQKSFKNILEPLHLYRRQSLIEATSVVINGASLTLIQYWA | | | |
| **STsORF25** | 63 | | y |
| MKQLPLSGVIIVSIEHVIATPFCTRQLADYFSRVTIIECQENRDFSCNYDERLKNDLCIFFFI | | | |
| **STsORF26** | 17 | | y |
| VLFLFGCCMQYDCTLNS | | | |
| **STsORF27** | 25 | | y |
| MNQKFEAVNAIAGNVTDVADGTDRC | | | |
| **STsORF28** | 30 | | y |
| VLCMCCIIMHVNTMFTGLVKSTHGVDRRHS | | | |
| **STsORF29** | 23 | | y |
| MFLLYCRIFLKLNPDFGEVAKCK | | | |
| **STsORF30** | 38 | | y |
| MIQTKPTSNTLKMLFIMKLRGGVYLIKQGGLCLNTSCL | | | |
| **STsORF31** | 34 | y | |
| MPFTRDKTRIFGGNVRGKRSSLSFCSVCGSIFSV | | | |
| **STsORF32** | 22 | | y |
| MLLFCLHSITFLYFRRGVNIGR | | | |
| **STsORF33** | 50 | | y |
| VSSLPSLKRHYLCKILCGILTPVTLTCFNFCHYSIASRLLVPLTGDFIYG | | | |
| **STsORF34** | 98 | y | |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MVMSAPGHIVYSSYNTLYGHSLSGGGLVILKALIISLTDHTHDVICGARSRVWRRFKKQAKAYKEANPQMCVRIIAFKRTRVMYTYNSRCYPWEDKKQ | | | |
| **STsORF35** | 69 | | y |
| MNTLTIDIYASENNPTLIMLSANTRPRPEACLNAISDICYFRYPAGHTSGHHHWWSINALPFTSHDYRH | | | |
| **STsORF36** | 11 | y | |
| MRRNIRAKEKR | | | |
| **STsORF37** | 18 | | y |
| MTIGVKDELLLRSELREF | | | |
| **STsORF38** | 30 | y | |
| VRTVKHNAICSDYMASLLNDMFLAERCQVS | | | |
| **STsORF39** | 24 | | y |
| VNYSWFCRFGIVAITECRIGTAGV | | | |
| **STsORF40** | 41 | y | y |
| VKKWLIGGILIASFLTGCLMWHNIDKWFNKDIEFFYVGDDS | | | |
| **STsORF41** | 25 | y | |
| MTGSGYLCLRYNRKNISAQTRTPIS | | | |
| **STsORF42** | 35 | y | |
| MYDPPFLEALMITASFFAIFIIIVVSVLLLEGGGD | | | |
| **STsORF43** | 13 | y | |
| VVRIIMYFRSEAL | | | |
| **STsORF44** | 65 | y | y |
| MKRFKEMATIFLFALLAAGFTSSAMAASGSDSPWDFDFSGPWLFCKLQPPDSLQLPPPPGEYCWH | | | |
| **STsORF45** | 71 | | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MIWDTVSCYYGLFLENCFINKRCCVIVLHNFCYIMLLYISLSTLRNLFFFSSQLAKKCFLYFHCFTVVMTN | | | |
| **STsORF46** | 20 | | y |
| VKNETKNSAKKPPDGGFFCI | | | |
| **STsORF47** | 51 | | y |
| MDKSKQMSSIMNRLIELTGWIVLVISVILLGIANHIDNYQPPEPTASVQKK | | | |
| **STsORF48** | 16 | | y |
| MFDFSQEVDRRGTWCT | | | |
| **STsORF49** | 31 | y | y |
| MLGSINLFIVVLGIILFSGFLAAWFSHKWDD | | | |
| **STsORF50** | 16 | | y |
| MFINSGCAAVIALIST | | | |
| **STsORF51** | 31 | | y |
| MNVSSKTVVLINVFAAAGLLGLISLRFGWFA | | | |
| **STsORF52** | 27 | | y |
| MIIRDMKGTVTEAFICLSCKYVCVNGA | | | |
| **STsORF53** | 17 | | y |
| MCNNRFMCLNKYCSSCI | | | |
| **STsORF54** | 60 | y | y |
| MSCLLTLLYLHSSLSDPIPTDPVPIPEPLPRPQPMPDPPPDEEPIKMSHQTPGSARIRAC | | | |
| **STsORF55** | 24 | | y |
| MSLFLTSFIFEKKEYDEEFYQLDG | | | |
| **STsORF56** | 57 | y | |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MKRSRTEVGRWRMLRQASRRKARWLEGQSRRNMRIHTIRKCIFNRQRNSLLFAIHGV | | | |
| **STsORF57** | 32 | | y |
| MNFNGIICEGMCIYPSYFKLLMCWLRSLTPVT | | | |
| **STsORF58** | 71 | | y |
| MFRFRRCKKPPGTDGASGGWLIAQDAIYEDLQPHSGYPGREDRLPLVYPEYFHQEAAHPLVMDRLNQDVQT | | | |
| **STsORF59** | 32 | y | y |
| MNNQPLRVSKDQPRKPDKTPEDEGKNPKKNQK | | | |
| **STsORF60** | 30 | | y |
| VRIYLSQFCCINVLSICEGSRNLCDGSRHV | | | |
| **STsORF61** | 91 | | y |
| MPFSIKNICSGPKGHCPEISSPIQDKPVPRNCTLTSTTCDIQSYTVFSRWSCSYEMRPPGAEERTPRLKFSATELSWLSKTIETERRNTKE | | | |
| **STsORF62** | 47 | | y |
| VNIRLLIKQSKSHADGFIVSIPLKYVICITYFLKSPSRSVVCHEADE | | | |
| **STsORF63** | 23 | | y |
| MPRRCYGACCCGKSIKEMVLFKK | | | |
| **STsORF64** | 23 | | y |
| (ML)MRITTGCAARPLVMSGARIKR | | | |
| **STsORF65** | 25 | | y |
| VPFRCPAGNNLMSSIFCIQRGRIQG | | | |
| **STsORF66** | 63 | y | |
| MLSAENGAGGIACECGTETEQGFRLLLWGGVRQVISTEDVLVVCRAEFAFRKQGVKNGGQTRH | | | |
| **STsORF67** | 18 | | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| VSLMIILVFCTHHRMYEH | | | |
| **STsORF68** | 16 | | y |
| MTRVQFKHHHHHHHHPD | | | |
| **STsORF69** | 27 | | y |
| MKILLWAILIIFLIGLLVVTGVFKMIF | | | |
| **STsORF70** | 21 | | y |
| VKKRSAELTKSCCCHLKSRKC | | | |
| **STsORF71** | 18 | | y |
| MMIFPQPIIFRSYYSSAK | | | |
| **STsORF72** | 17 | | y |
| MSKYRLIDNMNVLICCQ | | | |
| **STsORF73** | 76 | y | |
| VCNVFCQQKTETRYPAKPANSARVAALNAPSPEINASGCAIDGSRIMANPKLNWAREAIATLISPHITPIMPHSPE | | | |
| **STsORF74** | 21 | | y |
| MQQEENVLGINGVPCRNRTCN | | | |
| **STsORF75** | 31 | y | |
| MKFSSCQIVGQEYIYYARELDAFLPYIEVKK | | | |
| **STsORF76** | 23 | | y |
| MTINEMPIKKYFTFHLSAYLSYA | | | |
| **STsORF77** | 31 | | y |
| MIITPFCFTHLFISIYEVIKCYSRYKYLLNG | | | |
| **STsORF78** | 25 | | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| VKNLHHKAEKKSVEIRQTLVQETLI | | | |
| **STsORF79** | 61 | y | y |
| MPIIAPIPRTERRLMQKTIHKTRDKNHARRLTAMLMLQPGRQYRPRCQNALLRPFTHWTLD | | | |
| **STsORF80** | 62 | y | y |
| MKEGFYWIQHNGRVQVAYYTHGVTEDLETGQTIIGVWHLTQGDDICHNGEAEILAGPLEPPI | | | |
| **STsORF81** | 41 | y | |
| MGEPKGSPVGDPVTPTLYVSPPNPIGVGGGDKQNNHRRASL | | | |
| **STsORF82** | 56 | | y |
| MHNQKTCAYHLCGKTIEQGKEVKNELTLIRGAQLTHEERDYCSVRCASYDQMAHES | | | |
| **STsORF83** | 17 | | y |
| VLNNNVFYSMSTIVASF | | | |
| **STsORF84** | 15 | | y |
| MKLTRFFFAFFFIFP | | | |
| **STsORF85** | 16 | | y |
| MAAVVRLTRVNNLCCY | | | |
| **STsORF86** | 49 | | y |
| (MVCRNFGRGFNSRQLHQIMIRIRP)VKYRKPAQHKPCGLFCVCRCPRTSG | | | |
| **STsORF87** | 86 | y | |
| MIASLLANEYVSAFGYIVGVISGCIAIYQTTQVTKKNNEIKQLNVTIKDLNSQIINITTNRNNINQGERSQYFQDNNGPVNIDNRG | | | |
| **STsORF88** | 20 | | y |
| MKNNALTGKNILSGDKMLRQ | | | |
| **STsORF89** | 59 | y | |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MFTPGDIVQPRMGGPKLKVIEVNEDHIVAVQVGNEPGEKLILKAADVTPYCEEGDFGVC | | | |
| **STsORF90** | 29 | | y |
| VYICMNFRIWKYNQQSHTIPPVIVFCFLR | | | |
| **STsORF91** | 18 | | y |
| VNNVTRIECRTCCNICAI | | | |
| **STsORF92** | 81 | | y |
| MSLKKTESTQHKLKQTCQLRLWLIVGGVCSLGMGVLNILSSGYFEPYDGGQIILGLGCLGYSLALSKKISHLQAENQPKKP | | | |
| **STsORF93** | 20 | | y |
| MKIILICEMATICSDSAKEG | | | |
| **STsORF94** | 19 | | y |
| MNFVNRLKLKVVNFIYFSN | | | |
| **STsORF95** | 16 | | y |
| (MI)LCIDPLCLLEHNDA | | | |
| **STsORF96** | 17 | | y |
| VVFCADGVLAFFSHKLM | | | |
| **STsORF97** | 29 | | y |
| MDIYADNFRGKVCVCCVRDQGIPDDSGVQ | | | |
| **STsORF98** | 64 | | y |
| MGMIIIPCFIVIAIIAVVIFIMAIKACNMEKREDRQMTLAGIILMGVIAAIPGWFLYEIFSHAP | | | |
| **STsORF99** | 58 | y | y |
| MFIAWYWIILIVLVVVGYICHMKRYCRAFRQDRDALLEARTKRFRQTSEGDSVMNEQK | | | |
| **STsORF100** | 79 | y | |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| VDAINQNCVTSYGVKALMKVINSVIVNPLILRAGRNSPPAVNQRCLTHFSSRRAVTERGEQVESPRALFAIAKRSADPV | | | |
| **STsORF101** | 45 | y | |
| MAITLLPTSAQERRLNTVFQPRSYAPVVSSQSRRLADRSSRCCWI | | | |
| **STsORF102** | 54 | y | y |
| MSKKSAKKRQPVVKPAVQEAMSAAVPLGYEEMLTELEAIVADAEARLAEEEAAA | | | |
| **STsORF103** | 22 | | y |
| MIIIKDDKDIRTVYCATGCVPH | | | |
| **STsORF104** | 50 | | y |
| MKNQQTEKKIQKTLVQKNGIPIMRLHRHGGCESLHTNSRSVEEKNPEIQG | | | |
| **STsORF105** | 19 | | y |
| MQGARNTPQKSREPLHQPN | | | |
| **STsORF106** | 20 | | y |
| MICLHWDALYNAQVQLHACR | | | |
| **STsORF107** | 15 | | y |
| MFHFYQELSRALAIL | | | |
| **STsORF108** | 21 | | y |
| MLSLFCVIYQVINFLTTIITV | | | |
| **STsORF109** | 25 | y | |
| VGGSPGRRYDKHRLTILGFMCRLHC | | | |
| **STsORF110** | 18 | | y |
| MCLLPTNLNNKTDQQTRR | | | |
| **STsORF111** | 57 | y | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MSGKRYPEEFIIKAVKQVIERGHSVSSVATRLDITTHSLYAWIKPPYSRRYHAITGV | | | |
| **STsORF112** | 28 | | y |
| MQKRGLDKIFYQVVLIAIILILLTIWIR | | | |
| **STsORF113** | 37 | | y |
| MHKKGYCLLTILSGFYFIDSGISIIPSLYAWRNVAIE | | | |
| **STsORF114** | 16 | | y |
| MFMFKHALFPPPLTRR | | | |
| **STsORF115** | 31 | y | |
| MRCGPAIPARADGSSRIPSSCASQPSMSYIV | | | |
| **STsORF116** | 15 | | y |
| MLYICLNGSVIIPIN | | | |
| **STsORF117** | 73 | y | |
| VKTADSGSCVTLHAGSYRRAPTIINVSAFVIYRLLFEWSPPCHFRFGSSVLPDCRMSYQPQNSVCGYARCRRN | | | |
| **STsORF118** | 26 | | y |
| MGTIQDTKIADEPARLWTSMESKNPR | | | |
| **STsORF119** | 12 | y | |
| VKGDVYKRKMSL | | | |
| **STsORF120** | 14 | | y |
| MLLTRALSNEKEIR | | | |
| **STsORF121** | 52 | | y |
| MNIMKNAEFPDDWLLGLRQVVRDPVLYELLKYCPLEIMQYRAFMRCRLSMFC | | | |
| **STsORF122** | 25 | | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MRYRVTPCTIHIVCGKKCNVLTGLF | | | |
| **STsORF123** | 27 | | y |
| (VQC)LAILIIYCDEHHFLMVSECNCFTS | | | |
| **STsORF124** | 25 | | y |
| MGIRIPYVKNEKKLRKGLSVLLKKE | | | |
| **STsORF125** | 17 | | y |
| VILFCSAINPCTITRYF | | | |
| **STsORF126** | 17 | | y |
| MIQFTLFCCLKMRFMAR | | | |
| **STsORF127** | 24 | | y |
| VAIGACMQESAFLAFPSHSMLSKR | | | |
| **STsORF128** | 29 | y | y |
| MKSYIYKSLTTLCSVLIVSSFIYVWVTTY | | | |
| **STsORF129** | 32 | y | |
| VENAFGFPDSHEFNAPFPRLQSVQTVMTSLSG | | | |
| **STsORF130** | 34 | y | |
| MVSVFPAMMPCVDYATRACIDLASDALLYRFQAF | | | |
| **STsORF131** | 14 | | y |
| MSSFSSQCCLCIAT | | | |
| **STsORF132** | 65 | y | y |
| MNSTIWLALALVLVLEGLGPMLYPGAWKKMVSALAQLPENVLRRFGGGLVVAGVVVYYMLRKTIG | | | |
| **STsORF133** | 17 | y | y |

**Table 8.1** *continued from previous page*

| STsORF tag | length (aa) | Ribo-seq | sPepFinder |
|---|---|---|---|
| MDPEPTPLPRWRIFLFR | | | |
| **STsORF134** | 32 | | y |
| VLLICLLFYIPFLLQVAYVLATDYPAHPWASP | | | |
| **STsORF135** | 11 | y | |
| VLIVICNRHTF | | | |
| **STsORF136** | 73 | y | |
| MTLIERIRAADRRKGERQGRQVGLEEGLEKGLEKGQRVAALRIAKQMLADGLDRETVQRFTGLTAEELQDVSH | | | |
| **STsORF137** | 16 | | y |
| MKRQRKDKIIGARGGT | | | |
| **STsORF138** | 17 | y | y |
| MCTTLKISRMSSLSAHI | | | |
| **STsORF139** | 53 | y | y |
| MFRWGIIFLVIALIAAALGFGGLAGTAAGAAKIVFVVGIVLFLVSLFMGRKRP | | | |

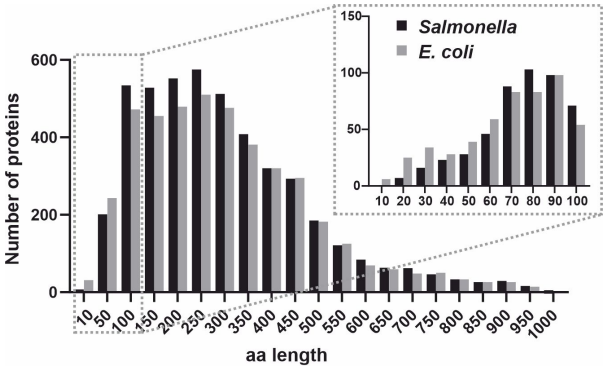



**Figure 8.1: Length distribution of CDSs annotated in *E. coli* and *Salmonella*.** The length of the CDSs annotated in *Salmonella* (black) and *E. coli* (grey) are shown in the frequency plot. The limit of the x-axes was set to a maximum of 1000 aa.

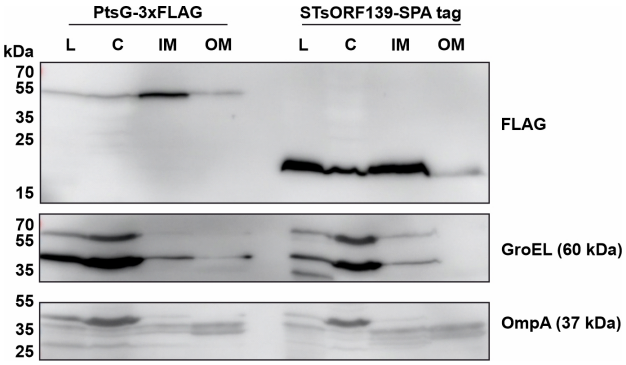



**Figure 8.2: STsORF139 is an inner membrane protein.** STsORF139-SPA tag or PtsG-3xFLAG *Salmonella* (the PtsG-3xFLAG strain was taken from (Papenfort et al., 2012)) were grown in LB medium to an $OD_{600}$ of 0.4 (PtsG) or SPI-2-inducing conditions (STsORF139), collected, and sub-fractionated. The content of each fraction (lysate (L), cytosol (C), inner membrane (IM), outer membrane (OM)) was analysed via western blotting. PtsG and STsORF139 could be detected with an $\alpha$-FLAG antibody, with PtsG serving as a control for the IM fraction. GroEL and OmpA were probed for as controls of the C and IM fractions, respectively.

# 9 Abbreviations

**Table 9.1:** List of abbreviations.

| Abbreviation | Full expansion |
|---|---|
| aa | amino acid |
| a.k.a. | also known as |
| APS | ammonium persulfate |
| ATP | adenosine triphosphate |
| bp | base pair |
| c.f.u. | colony forming unit |
| C | cytosol |
| cDNA | complementary DNA |
| CDS | coding sequence |
| Co-IP | co-immunoprecipitation |
| D | detergent |
| DNA | deoxyribonuleic acid |
| DNase | deoxyribonuclease |
| DTT | dithiothreitol |
| DUF | domain of unknown function |
| EDTA | ethylene diamine tetraacetic acid |
| EEP | early exponential phase |
| ESP | early stationary phase |
| FACS | fluorescence-activated cell sorting |
| FC | fold-change |
| FDR | false discovery rate |
| gDNA | genomic DNA |
| GFP | green fluorescent protein |
| Grad-seq | gradient profiling by sequencing |
| HS | high salt |

**Table 9.1** *continued from previous page*

| Abbreviation | Full expansion |
|---|---|
| IM | inner membrane |
| L | lysate |
| LB | Lennox broth |
| LC-MS/MS | liquid chromatography-tandem mass spectrometry |
| LDH | lactate dehydrogenase |
| LEP | late exponential phase |
| LSP | late stationary phase |
| MAC | macrophages |
| MEP | mid exponential phase |
| m.o.i. | multiplicity of infection |
| mRNA | messenger RNA |
| ncRNA | non-coding RNA |
| nt | nucleotide |
| $OD_{600}$ | optical density at 600nm |
| OM | outer membrane |
| ORF | open reading frame |
| P/C/I | phenol/chloroform/isoamyl alcohol |
| p.i. | post-infection |
| PAA | polyacrylamide |
| PAGE | polyacrylamide gel electrophoresis |
| PCR | polymerase chain reaction |
| RBP | RNA-binding protein |
| RBS | ribosome binding site |
| Ribo-seq | ribosome profiling by sequencing |
| RIL-seq | RNA interaction by ligation and sequencing |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| RNA-seq | RNA sequencing |
| rRNA | ribosomal RNA |
| SD | Shine-Dalgarno |
| SDS | sodium dodecyl sulfate |
| SE | *Salmonella* enteritidis |
| SG | *Salmonella* gallinarum |

**Table 9.1** *continued from previous page*

| Abbreviation | Full expansion |
|---|---|
| SL | *Salmonella* Typhimurium |
| sORF | small open reading frame |
| SPA | sequential peptide affinity |
| SPI | *Salmonella* pathogenicity island |
| sRNA | small regulatory RNA |
| ST | *Salmonella* Typhi |
| STsORF | *Salmonella* Typhimurium sORF |
| T3SS | type 3 secretion system |
| TA | toxin-antitoxin |
| TCS | two-component system |
| TL | translational |
| TraDIS | transposon-directed insertion sequencing |
| TSS | transcription start site |
| TX | transcriptional |
| UTR | untranslated region |
| v/v | volume/volume |
| w/v | weight/volume |

# 10 Curriculum Vitae

## Education

**04/2017 – to date: PhD student**, Institute for Molecular Infection Biology, University of Würzburg, Germany.
Project: Characterization of small proteins involved in *Salmonella* virulence

**10/2015 – 03/2017: MSc FOKUS Life Sciences**, University of Würzburg, Germany. Interdisciplinary Master in Molecular Life Sciences. Admitted to MSc/PhD fast track in April 2017.
Thesis title: Characterization of small proteins in *Escherichia coli* and *Salmonella* Typhimurium

**09/2011 – 10/2014: BSc in Biotechnology**, University of Udine, Italy.
Thesis title: Sanger sequencing of *ape1* gene in a cohort of Huntington's disease patients in search of single nucleotide polymorphisms

**11/2007 – 09/2015: Piano diploma**, Conservatorio "Jacopo Tomadini", Udine, Italy

**09/2006 – 07/2011: High school diploma**, Liceo "Niccoló Copernico", Udine, Italy

## Prior laboratory experience

**12/2015 – 01/2016: Internship**, Rudolf-Virchow-Zentrum, University of Würzburg, Germany.
Supervisor: Dr. Ingrid Tessmer. Project: Application of atomic force microscopy for the preliminary characterization of alkyl transferase-like 1 DNA-binding activity

**11/2014 – 09/2015: Internship**, Department of Medical and Biological Sciences, University of Udine, Italy.

Supervisor: Prof. Gianluca Tell. Project: Cloning of *ape1* polymorphic genes in expression vectors and characterization of SNPs role in HCT116 WT strain

**11/2014 – 09/2015: Internship**, Department of Medical and Biological Sciences, University of Udine, Italy.
Supervisor: Dr. Carlo Vascotto. Project: Preliminary characterization of APE1 – Mia40 interaction

# 11 List of publications

## Published manuscripts

**Venturini E.**, Svensson S.L., Maaß S., Gelhausen R., Eggenhofer F., Li L., Cain A.K., Parkhill J., Becher D., Backofen R., Barquist L., Sharma C.M., Westermann A.J., and Vogel J., 2020. A global data-driven census of *Salmonella* small proteins and their potential functions in bacterial virulence. *microLife* 1, no. 1, uqaa002.

Hör J., Di Giorgio S., Gerovac M., **Venturini E.**, Förstner K.U. and Vogel J., 2020. Grad-seq shines light on unrecognized RNA and protein complexes in the model bacterium *Escherichia coli*. *Nucleic acids research*, 48(16), pp.9301-9319.

Westermann A.J., **Venturini E.**, Sellin M.E., Förstner K.U., Hardt W.D., and Vogel, J., 2019. The major RNA-binding protein ProQ impacts virulence gene expression in *Salmonella enterica* Serovar Typhimurium. *MBio*, 10(1).

## Manuscripts in preparation

**Venturini E.**, Westermann A.J., Vogel J.. A collection of observations regarding the small proteins YjiS and its involvement in *Salmonella* virulence.

Matera G., **Venturini E.**, Vogel J.. RIL-seq reveals changes in the interactome of Hfq during a *Salmonella* infection time course.

# 12 Attended conferences and courses

1. **Small proteins, big questions**
   January 12$^{th}$ - 14$^{th}$, 2021. Virtual
   Poster presentation

2. **The 24$^{th}$ annual meeting of the RNA society**
   June 11$^{th}$ - 16$^{th}$, 2019. Krakow, Poland
   Poster presentation

3. **Salmonella biology and pathogenesis - Gordon Research Conference**
   June 2$^{nd}$ - 7$^{th}$, 2019. Easton, MA, USA
   Poster presentation

4. **SPP2002 annual meeting**
   March 4$^{th}$ - 5$^{th}$, 2019. Sylt, Germany
   Talk and poster presentation

5. **SPP2002 kick-off meeting**
   February 14$^{th}$ - 15$^{th}$, 2018. Kiel, Germany
   Poster presentation

6. **The 22$^{nd}$ annual meeting of the RNA society**
   May 30$^{th}$ - June 3$^{rd}$, 2017. Prague, Czech Republic
   Poster presentation

# 13 Contributions

The people who have contributed to different parts of the work presented here are listed below.

- Ribo-seq was performed by Dr. Sarah Svensson (Sharma group, IMIB, Würzburg) and data analysis was carried out by Dr. Florian Eggenhofer and Rick Gelhausen (Backofen group, University of Freiburg).
- Small protein predictions by sPepFinder were provided by Dr. Lei Li (Shenzhen Bay Laboratory, China).
- TraDIS data were provided by Prof. Dr. Lars Barquist (HIRI, Würzburg).
- Mass spectrometry analysis was performed by Dr. Sandra Maaß (Becher lab, University of Greifswald).
- RNA-seq analysis was carried out with the help of Prof. Dr. Lars Barquist (HIRI, Würzburg).
- cDNA libraries generation and sequencing were carried out by Vertis Biotechnologie AG in Freising-Weihenstephan, Germany.

# 14 Acknowledgments

I am grateful to Jörg Vogel for giving me this project. I loved every bit of the science in it, and loved to be part of the small proteins community. I also thank the members of my committee, Cynthia Sharma, Wilma Ziebuhr, and Kai Papenfort, for their time and their supervision.

A special thanks to Alex Westermann, for his unconditional support and guidance.

This work would not have been possible without all the people I have collaborated with, in particular I want to thank Lars Barquist and Sarah Svensson.

I want to thank my colleagues in the Vogel lab, past and present, who have taught me so much!

To all my friends from work: Austin Mottola (here from the start - what a journey it has been!), Manuela Fuchs, Hilde Merkert, Svetlana Durica-Mitic, Daniel Ryan, Laura Wicke, Gohar Mädler, Jakob Jung, Wael Bazzi, Yan Zhu, and many others... Thank you for being such a great community!

My italian family at work: Elisabetta Fiore and Silvia Di Giorgio, I was extremely lucky to have you! You made it all immensely better. Thanks also to Gianluca Matera who asked, right before I printed this thesis, if I ever found out what does YjiS do.

I am lucky to have moved to a covid-time office with people that really made the last moths of my PhD so much better - Sahil Sharma, Falk Ponath, and Kotaro Chihara, along with others already mentioned. The science talks (and the Wednesdays) were great!

I owe the biggest thank you to my family and to Gianluca.

I dedicate this work to LV.