PhD Thesis

Investigation of the transcriptional landscape and RNA biology of *Salmonella* Typhimurium plasmids

Submitted by

Sathesh Kumar Sivasankaran

Department of Microbiology

School of Genetics and Microbiology

Trinity College Dublin

Dublin 2, Ireland.

November 2014

Dedicated to my parents,

Saraswasthy Sivasankaran and Sivasankaran Vadivelu

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

Any parts, words or ideas, of the thesis, however limited, and including data, tables, graphs, etc., which are quoted from or based on other sources have been acknowledged as such without exception.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Sathesh Kumar Sivasankaran

Date:

Summary

This study focused on the transcriptional landscape and RNA biology of three plasmids of *Salmonella enterica* serovar Typhimurium, a model Gram-negative pathogen. Plasmids are key players in terms of mobile genetic elements that increase bacterial adaptability and diversity. The research work was designed to identify transcriptional start sites, non-coding RNAs, expression profiles of both coding and non-coding genes, and the functional role of non-coding RNAs in plasmids of *Salmonella*.

The transcriptome of *Salmonella* plasmids was analysed using next-generation sequencing technology. Initially, a bioinformatics analysis pipeline was developed to analyse hundreds of RNA-Seq libraries produced during this study. The regulatory roles of a plasmid-encoded sRNA and four asRNAs were studied by transient overexpression of sRNA/asRNAs and subsequent observation of global changes in transcript abundance with microarrays. In addition, the putative targets of an sRNA were confirmed by transcriptomic analyses using RNA-seq in the sRNA deletion mutant.

The major findings were identification of 162 plasmid-associated TSS in four different growth conditions, including 59 primary and 56 antisense TSS. Of the 120 TSS expressed in ESP, 86 were bound by the transcriptional factor RpoD. A total of 39 antisense RNAs and a pSLT$^{4/74}$-encoded sRNA STncP1-1, were identified, and the expression of 9 asRNAs and STncP1-1 were validated using Northern blot.

The putative role of STncP1-1 in *Salmonella* virulence was identified using macrophage experiments and mouse models. A total of 72 transcripts were identified as candidate targets of plasmid-encoded sRNA/asRNA-mediated regulation in *S*. Typhimurium. However, confirmatory experiments were inconclusive.

The distinct expression profiles of plasmids-encoded genes across 23 environmental conditions were identified. It was found that genes with similar function had similar expression patterns, e.g., partitioning systems from pSLT$^{4/74}$ and pCol1B$^{4/74}$, the *spv* locus, clustered with SPI-2, and streptomycin resistance genes clustered with *rpsM*. The highly complex network of interactions between plasmids and chromosome of *Salmonella* at the regulatory level was identified using a series of mutants that lack known regulators and including transcription factors, global regulators, SPI-1 and SPI-2 associated regulators.

Acknowledgements

There are a number of people that I would like to thank for their help and guidance in the past four years.

First of all, I would like to thank my supervisor Prof. Jay C. D. Hinton and co-supervisor Dr. Karsten Hokamp for their great supervision, frequent and productive discussions, valuable, inspiration and lots of moral support during my PhD.

I want to thank the members of my committee, Prof. Charles J. Dorman, Dr. Stephen Smith and Dr. Adrian Bracken for their guidance throughout the course of my PhD work.

I sincerely thank Dr. Carsten Kröger, Dr. Magali Hébrard, Dr. Shabarinath Srikumar, Jane Twohig, Kristian Händler, and Aoife Colgan for their assistance in my project, especially teaching me wet lab experiments and creating a wonderful working environment. I would like to thank Dr. Disa L. Hammarlöf for her support during the Liverpool visit.

I wish to thank Dr. Jarlath E. Nally and Dr. Melissa J. Caimano for giving me the opportunity to work on the *Leptospira* project, and Dr. Carsten Kröger and Dr. Andrew D. S. Cameron for the *Acinetobactor* project.

I want to thank Prof. Josep Casadesús and his lab members for their support during the Sevilla visit.

I wish to thank Mr. Mithileash Mohan, Mr. Servesh Muralidharan, Mr. Aravind Vasudevan, Ms. Rukmani Sridharan, Dr. Atul Nautiyal, Ms. Arunima Gulati, Mr. Sandeep Pathivada and Mr. Mahendiran Venkatachalapathy for their support during my PhD.

I would like to express my deepest thanks to my sister Mrs. Sharmeela Devi Balamurugan and brother Mr. Saravanan for their belief in me and care during every stage of my life.

Special thanks go to my wife Mrs. Ramyaa Manoharan for her enormous support during my PhD.

Finally, I want to thank the Science Foundation Ireland for providing the financial support that made this PhD possible.

Publications and Presentations arising from PhD research

Peer reviewed journal articles co-authored during PhD

2013 An infection-relevant transcriptomic compendium for *Salmonella enterica* serovar Typhimurium. ***Cell Host Microbe***. 14:683-95.

2012 The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. ***Proc Natl Acad Sci U S A***. 109:E1277-86.

Peer reviewed review co-authored during PhD project

2011 The challenge of relating gene expression to the virulence of *Salmonella enterica* serovar Typhimurium. ***Curr Opin Biotechnol*.** 22: 200-10.

Peer reviewed journal article arising from collaboration during PhD

2014 A model system for studying the transcriptomic and physiological changes associated with mammalian host-adaptation by *Leptospira interrogans* serovar Copenhageni. ***PLoS Pathog***. 10: e1004004.

Poster presentations

2013 Dublin academy of pathogenomics and infection biology (DAPI), Dublin

2013 3rd international conference on Regulating with RNA in Bacteria, Würzburg (Germany)

2012 Dublin academy of pathogenomics and infection biology (DAPI), Dublin

2012 Young microbiologist symposium (YMS), Cork

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

(v/v)	(volume/volume)
(w/v)	(weight/volume in g/ml)
4/74	*Salmonella enterica* sv Typhimurium 4/74
aa	amino acid(s)
Amp	ampicillin
APS	ammonium persulfate
asRNA	antisense RNA
ATP	adenosine triphosphate
BHR	broad host range
BLAST	basic local alignment search tool
bp	base pair
cDNA	complementary DNA
CDS	coding sequence
Cm	chloramphenicol
D23580	*Salmonella enterica* sv Typhimurium D23580
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
dRNA-Seq	differential RNA sequencing
E. coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetate
EEP	early exponential phase
ESP	early stationary phase
Hfq	host factor for Qβ replication
IGR	intergenic region
Kan	kanamycin
kb	kilo base
kcal	kilocalorie
LB	lennox broth
LEP	late exponential phase
LSP	late stationary phase
MB	megabyte
MEP	mid exponential phase
MES	2-(N-Morpholino)ethanesulfonic acid
NCBI	national center for Biotechnology Information
ncRNA	noncoding RNA

NGS	next-generation sequencing
NHR	narrow host rage
nt	nucleotide(s)
OD	optical density
ORF	open reading frame
PAA	polyacrylamide
PBS	phosphate buffered saline
PCR	polymerase chain reaction
Perl	practical extraction and report language
RACE	rapid amplification of cDNA ends
RBS	ribosome binding site
RNA	ribonucleic acid
RNase	ribonuclease
RNA-Seq	RNA sequencing
rRNA	ribosomal RNA
SD	Shine-Dalgarno
SDS	sodiumdodecylsulfate
SL1344	*Salmonella enterica* sv Typhimurium 4/74
sRNA	small regulatory RNA
SRP	signal recognition particle
sv	serovar
T/A	toxin and antitoxin
TBE	Tris/Borate/EDTA
TEMED	tetramethylethylendiamine
TIR	translation initiation region
TraDIS	transposon-directed insertion-site sequencing
Tris	tris-(hydroxymethyl)-aminomethan
TSS	transcriptional start sites
tRNA	transfer RNA
U	uracil
UTP	uridine triphosphate
UTR	untranslated region

UNITS

°C	degree Celsius
kDa	kilo Dalton
g	gram
h	hour
l	litre
M	molar
min	minute
molar	gram molecule
rpm	rounds per minute
sec	second
msec	milliseconds
u	unit
V	Volt
W	Watt

MULTIPLES

M	mega (10^6)
k	kilo (10^3)
m	milli (10^{-3})
µ	micro(10^{-6})
n	nano (10^{-9})
p	pico (10^{-12})

CONTENTS OF ACCOMPANYING CD

Supplementary Datasets (Excel sheets)

3.1 TPM based expression values of all coding and sRNAs for ESP and InSPI2 growth condition (FRT-Seq).

3.2 List of *S*. Typhimurium 4/74 RNA-Seq and dRNA-Seq and sRNA-Seq libraries analysed with my RNA-Seq analysis pipeline.

4.1 Binding region of RNAP, RNAP+rif and RpoD identified by ChIP-chip.

5.1 Absolute and relative expression values of plasmid-encoded genes of *S*. Typhimurium 4/74 in 23 environmental conditions.

5.2 Absolute and relative expression values of plasmid-encoded genes of *S*. Typhimurium 4/74 in different regulatory mutants.

CHAPTER 1

INTRODUCTION

1.1 *Salmonella*

Salmonella is a food/water-borne, Gram-negative, facultative and rod-shaped bacterial pathogen from the ***Enterobacteriaceae*** family**,** commonly known as "enteric" bacteria. It was first identified by Dr. Daniel E. Salmon and Dr. Theobald Smith in their hunt for the causative agent of swine plague (Salmon and Smith 1886). Members of this genus are generally motile with peritrichous flagella. Initially, each *Salmonella* isolate was considered as a species and named according to clinical considerations (Kauffmann 1941, Kauffmann 1966), example is *Salmonella enterica* serovar Typhimurium (mouse typhoid fever). After it was recognized that host specificity did not exist for many species, new isolate received species name according to the location at which was isolated (Le-Minor and Popoff 1987), for example *S.* Dublin. Subsequently, the high genetic similarity between *Salmonella* isolates suggested that *Salmonella* consisted of only two species *S. enterica* and *S. bongori* (Reeves et al. 1989). One other species, *S. subterranean* was subsequently proposed by the judicial commission of the international committee of systemic bacteriology (Shelobolina et al. 2004), but the proposed species was later reported not to belong to the genus *Salmonella* by others (Grimont and Weil 2007).

The expansion and diversification within the *S*. *enterica* species has led to the classification of six subspecies: I – *enterica*, II – *salamae*, IIIa – *arizonae*, IIIb – *diarizonae*, IV – *houtenae* and VI – *indica* (Boyd et al. 1996; Agbaje et al. 2011). *S. bongori* was previously categorised as subspecies V, but its genome sequence confirmed that it was a distinct species (Fookes et al. 2010; Su and Chiu 2007). Subspecies II, IIIa, IIIb, IV and VI are commonly isolated from cold-blooded animals and the environment (Murray et al. 2007), whereas *S. enterica* subspecies *enterica* has acquired the additional ability to colonise, and is more frequently isolated from warm-blooded animals (Porwollik et al. 2004). Further diversification of subspecies of *Salmonella* has generated more than 2,500 serovars, and most belong to subspecies *enterica*. The serotypes are characterised on the basis of their binding of antiserum

that recognises two highly variable surface structures: the 'O' antigen of lipopolysaccharide (LSP), and flagellar 'H' antigen.

1.1.1 *Salmonella disease and virulence*

Salmonella bacteria cause a substantial burden of disease and mortality on a global scale, with incidences of over 100 million cases and 400,000 deaths per year (Crump et al. 2004; Majowicz et al. 2010). In the European Union (EU), over 100,000 human cases of *Salmonella* infection are reported each year (EFSA and ECDPC 2013). The European Food Safety Authority (EFSA) has estimated that the overall economic burden of human *Salmonella* infection could be as high as EUR 3 billion each your (EFSA and ECDPC 2013). *Salmonella* causes clinical conditions ranging from mild gastrointestinal infection to systemic infections such as typhoid fever. Typhoid remains a big killer in the developing world, and *Salmonella* serovars Typhi and Paratyphi are responsible for about 216,000 deaths each year (Crump et al. 2004). Typhoid fever is a waterborne disease predominantly transmitted via polluted drinking water (Scherer and Miller 2001). Enteric fever is rarely a cause of death in the developed world, and is not common in Europe.

Salmonellosis is generally a self-limiting gastroenteritis that is the most common food-borne disease caused by non-typhoidal *Salmonella* serovars (NTS), including *S*. Typhimurium and *S*. Enteritidis. Worldwide, NTS are responsible for 1 billion cases of gastroenteritis cases and approximately 155,000 deaths each year (Majowicz et al. 2010). The major sources of salmonellosis are contaminated farm animals/poultry products or food contaminated with faecal matter. Symptoms of *Salmonella* infection usually appear 12–72 hours after infection, and include fever, abdominal pain, diarrhoea, dehydration, nausea and occasionally vomiting. The illness usually lasts 4–7 days, and most people recover without treatment. However, in some people, especially the elderly, infants, and those with impaired immune systems, the illness may be so severe that the patient requires hospitalisation. On the island of Ireland, salmonellosis is primarily linked to two serotypes, *S*. Enteritidis and *S*. Typhimurium (Foley et al. 2007). Since 2004, the number of salmonellosis cases in Ireland, ranged from 415 to 314 per year (Figure 1.1; HPCS 2012). Of the 314 human salmonellosis cases report in 2012, 309 were laboratory confirmed, with 5 probable cases (HPSC 2012).

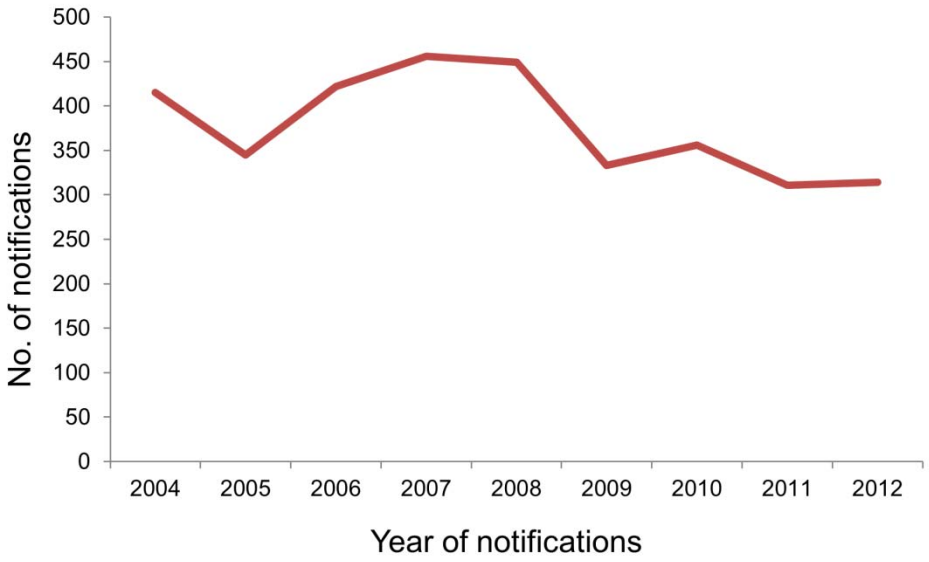


Figure 1.1: Incidence of salmonellosis in Ireland 2004-2012.
The number of cases of salmonellosis reported per annum in Ireland. Figure reproduced from HPSC salmonellosis annual report, Ireland 2012.

Approximately 70% of clinical cases worldwide are caused only by twelve virulent serotypes, including the most well-known Typhi, Paratyphi, Enteritidis, Choleraesuis and Typhimurium (Anjum et al. 2005). Such virulence has been gained by the expression of several horizontally acquired genetic elements that integrated into the *Salmonella* core genome and therefore referred as "*Salmonella* Pathogenicity Islands" (SPI) (Boyd and Hartl 1998). At present, 12 different SPIs have been described, which carry important virulence information for host cell invasion and intracellular replication (Hensel 2004; Sabbagh et al. 2010).

Among all SPIs, the most studied islands are SPI-1 and SPI-2, each encoding a specialised apparatus, termed the type III secretion system (T3SS), which mediates the translocation of virulence proteins from the bacterial cell to the cytoplasm of the host cell (Hansen-Wester and Hensel 2001). The virulence factors encoded by SPI-1 permit the invasion of non-phagocytic cells (Lara-Tejero and Galán 2009) and are controlled by the transcriptional activator HilA (Bajaj et al. 1995). Once inside the phagocytic cell, *Salmonella* replicates within a unique phagosome-like compartment, termed the '*Salmonella* containing vacuole' (SCV) (Haraga et al. 2008). The acidic pH milieu of the SCV triggers the expression of SPI-2 effectors, which translocate effector proteins across the bacterial membrane and subsequently turn the SCV into a protected niche for survival of *Salmonella* (Bakowski et al. 2008). The expression of SPI-2 encoded

genes is activated by the SsrA/B two-component system (Garmendia et al. 2003). Apart from chromosomal-encoded virulence factors, the *Salmonella* plasmid (pSLT) also encodes certain virulence proteins (see below).

1.1.2 *Salmonella infection – a series of stressful events*

Salmonella is a well-studied bacterium and the mechanisms of *Salmonella* virulence have been studied extensively with animal models including mice, pigs and chickens in recent decades. The majority of infection-relevant studies have been done with *S*. Typhimurium, which causes a typhoid-like fever in mouse, but still not all aspects of infection are fully understood. The *Salmonella* bacteria experience an extremely stressful series of events and changeable conditions from the anaerobic nutrient-rich surroundings of the GI tract of an animal, to a completely different environment on the outside, followed by entry into a human host with further changes in environmental conditions. The stresses encountered by *Salmonella* during the infection have been reported by Hébrard et al. (2011).

During the infection process, *Salmonellae* have to cross the low pH milieu of the stomach, which requires an adaptive acid-tolerance that is essential for efficient colonization of the host (Garcia-del Portillo et al. 1993). Entering the small intestine, *Salmonella* encounters deleterious agents like bile or antimicrobial peptides and stresses including temperature, osmolarity, low oxygen and low nutrient content (Figure 1.2A). In addition, *Salmonella* is confronted with a number of challenges imposed by host cell such as phagocytic cells harbouring multiple stress factors including low iron, magnesium and phosphate, and a multitude of cytotoxic compounds (Figure 1.2B). To adapt to these series of stressful environmental events, *Salmonella* must sense and respond immediately by governing the changes in gene expression and activating specific sets of virulence genes that allow *Salmonella* to resist, evade, or even systematically manipulate the effectors of host immune responses (Haraga et al. 2008).

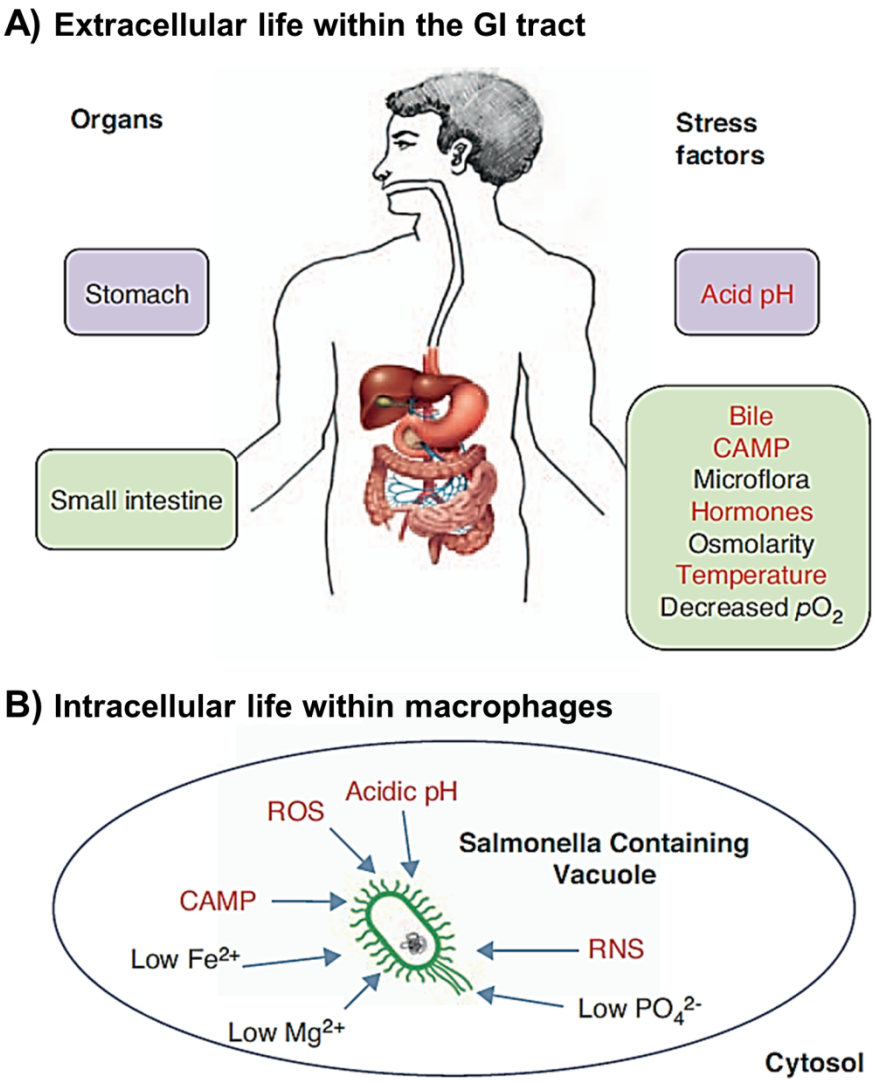


Figure 1.2: Series of stresses encountered by *Salmonella* during the infection.
A) Stresses faced by *Salmonella* upon entry into the host and within the gastrointestinal (GI) tract. Following ingestion, *S*. Typhimurium is usually faced with the temperature shift and acidic pH environment of the stomach, followed by bile, cationic antimicrobial peptides (CAMP), microflora, hormones, increased osmolarity, and reduced oxygen levels within the intestines and colon. B) Stresses encountered by *Salmonella* during survival and proliferation inside the macrophage. In the phagocytic cells, *S*. Typhimurium are thought to be exposed to reactive oxygen species (ROS), reactive nitrogen species (RNS), CAMP, acidic pH, reduced concentration of iron, magnesium and phosphate. Figure source Hébrard et al. (2011)

1.1.3 *Salmonella genomic sequences*

The *Salmonella* plasmid transcriptome analysis performed in this thesis benefits from the sequencing of three *Salmonella* genomes, namely *S*. Typhimurium 4/74, *S*. Typhimurium D23580 and *S*. Typhimurium SL1344. The first *Salmonella* genomes to be sequenced were *S*. Typhimurium LT2 (McClelland et al. 2001) and *S*. Typhi CT18 (Parkhill et al. 2001). Subsequently, the genome of many *Salmonella* serovars have been published, including *S*. Paratyphi A ATCC 9150 (McClelland et al. 2004), *S*. Choleraesuis SC-B76 (Chiu et al. 2005) and *S*. Typhimurium 14028 (Jarvik et al. 2010) etc. The genome of *S*. Enteritidis was published by Thomson et al. (2008) and *S. bongori* was published by Fookes et al. (2011). The genome of the African serovar *S*. Typhimurium D23580 was published by Kingsley et al. (2009). More recently, the genome of *S*. Typhimurium 4/74 (Richardson et al. 2011) and *S*. Typhimurium SL1344 (Kröger et al. 2012) have been published.

The strain used throughout this study (*S*. Typhimurium 4/74), contains a 4.87 Mb chromosome with 4527 coding genes and three plasmids (see below for plasmids).

1.2 *Plasmids*

Plasmids are extra-chromosomal, self-replicating genetic elements that usually exist in a supercoiled form. Plasmids were first named by the American geneticist Joshua Lederberg in *Enterobacteriacae* (Lederberg 1952) and subsequently have been discovered in other bacteria, *Archaea* and *Eukarya* (Couturier et al. 1988). In 1952, Lederberg suggested the term "Plasmid" to refer to 'extra-nuclear' chromosomes, specifically for the R-factor and F-factor (Lederberg 1952). The structure of plasmids is generally a double stranded, circular (sometimes linear) DNA molecule, and their lengths vary from few to several hundred kilobases (Selimovic et al. 2007). The copy number of plasmids varies from one copy, generally for large plasmids, to hundreds of copies in a single cell, and cells can harbour more than one type of plasmid. Plasmids can promote their horizontal transfer to other cells through bacterial conjugation or mobilization (transformation), however not all plasmids are transferable (Francia et al. 2004). Some plasmids can integrate into the host chromosome (Chiu and Thomas 2004). Certain plasmids are able to replicate in a variety of bacterial hosts (broad host range – BHR); for instance pAKD plasmids (Sen et al. 2011), and others have a narrow host range (NHR); for example, pColE1 is limited to enteric bacteria by its inability to replicate in other bacterial families.

Plasmids differ from the chromosome in being small and generally do not encode functions essential to bacterial growth, such as RNA polymerase, ribosomal subunits, or enzymes of the tricarboxylic acid cycle (Baron *et al*. 1996). They typically consist of two distinct regions: a backbone region that carries genes essential for plasmid maintenance including replication, stability and inheritance: and one or more accessory regions that encode host-beneficial traits (Thomas 2000; Jackson et al. 2011). The genes encoded by the accessory region may aid the host cell to adapt to various environments (see below). Figure 1.3 shows the organisation of a plasmid, consisting of backbone and accessory regions.

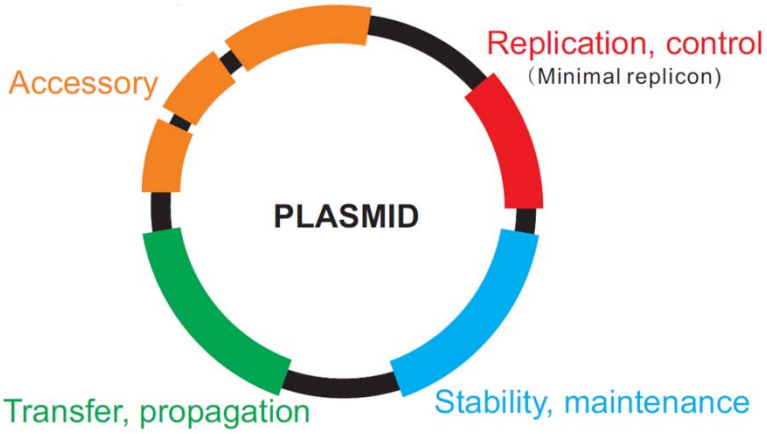


Figure 1.3: The organization of a plasmid consisting of backbone and accessory regions.
The plasmids can be divided into two parts: backbone and accessory region. The backbone can be further divided into three parts: replication and control (also defined as a 'minimal replicon'), stability and maintenance, transfer and propagation. The accessory region encodes genes for host beneficial traits. Figure source Norman et al. (2009).

1.2.1 *Plasmid backbone region*

1.2.1.1 *Replication and transfer systems*

A replicon is a region of DNA that is replicated from an origin of replication. It could potentially be a whole bacterial chromosome or plasmids. A minimal replicon comprises just the essential elements for replication. i.e. the Rep system. Regarding plasmids, the first essential gene existing on the plasmids is a *rep* gene, which is responsible for the initiation of plasmid replication. The role of the *rep* gene is to ensure that the replication is balanced with the host cell growth cycle, as any failure would cause a burden for the host cell (Ebersbach and Gerdes 2005). Consequently, the *rep* gene is often coupled with the *cop* gene in order to control both replication and copy number of a plasmid.

The modules affecting plasmid propagation are *tra* genes, which encode DNA transfer and replication (Dtr) and Mating Pair Formation (MPF) proteins (Norman et al. 2009). Self-transmissible plasmids are able to transfer unaided when they contain both Dtr and MPF components, while mobilisable plasmids are only able to transfer in the presence of a self-transmissible plasmid when they possess Dtr components (Smillie et al. 2010).

1.2.1.2 *Stability systems*

Bacterial plasmids are usually stably maintained in their host cells (Hayes and Barillà 2006). High copy number plasmids can rely on random distribution among daughter cells. However, when the plasmid is present in only a few copies, a mechanism that ensures better than random distribution is required. Stability can be increased by the activity of a site-specific recombinase or resolvase that resolves the plasmid multimer form following plasmid replication and recombination (Hayes and Barillà 2006), which decreases their stability. There are several mechanisms that have been linked with plasmid stability and partitioning systems are the most significant. Partitioning systems ensure that the plasmids are actively moved into the proper position prior to cell division. This process is controlled by two coupled partitioning genes (*par*).

Plasmid partitioning systems are classified according to the cytoskeletal components that they encode (Ebersbach and Gerdes 2005). The first gene of a *par* operon encodes an ATPase, which is used to divide *par* loci into two types (Gerdes et al. 2000). The type I loci encode Walker-type ATPase, and type II loci encode actin-like ATPase (Bouet et al. 2005; Koonin 1993). The second gene of the *par* operon encodes a DNA-binding protein that recognizes varying numbers of direct and inverted repeats within a cognate centromere-like site (Davis et al. 1988; Møller-Jensen et al. 2003). Binding of the protein to centromere-like site, results the formation of partitioning (nucleo-protein) complex. The partitioning complex is the substrate for plasmid segregation, in which replicated plasmid molecules, often located at the mid-cell position, are moved in opposite directions. Figure 1.4 explains the mechanism and difference between these two types of partitioning systems. The best understood type I *par* system is *parAB*, where the *parA* encodes ATPase and *parB* is a DNA-binding protein (Gerdes et al. 2000). The centromere-like region of *parAB* system, *parS* sequence is localised downstream of the *par* operon. The well-studied type II *par* locus is encoded by plasmid R1. Here, the ATPase is designated *parM*, where the DNA-binding protein is call *parR*, and the cis-acting site is *parC* (Dam and Gerdes 1994).

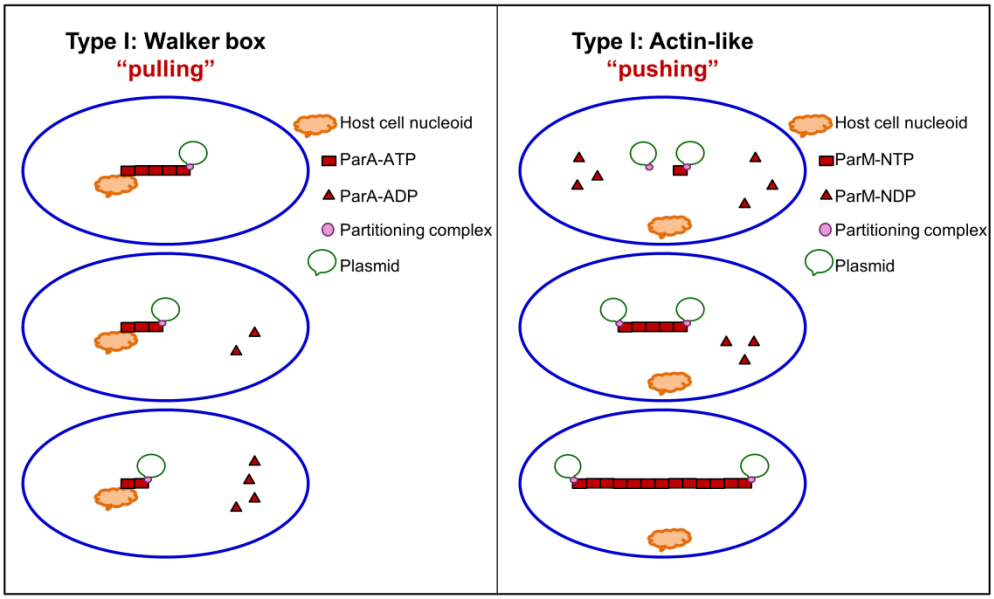


Figure 1.4: Schematic models for type I and II plasmid partitioning systems.
In type I, the ParA-ATP (red square) binds co-operatively to host cell nucleoid, forming a nucleating core from which filaments form (polymerisation). Subsequently, a growing filament contacts with a ParB-centromere partitioning complex (pink circle), which stimulates the ATPase activity resulting in dissociation of ParA-ADP subunits (red triangles) and polymer retraction. The partitioning complex is pulled along with the retreating ParA polymer. The ParB-plasmid is thus pulled in the opposite direction to the growth of the filament, and moves around its position towards host cell nucleoid. In type II, the dynamically unstable ParM filaments are stabilised and propagate only when each end is captured by a ParR-centromere partitioning complex. Continuous insertion of ParM-ATP or ParM-GTP subunits on the both ends of the filament pushes the plasmid molecules apart. Conversion of ParM-ATP to ParM-ADP leads to destabilisation of the filament, thus allowing the entry of another ParM-ATP. At cell division, plasmids localise near opposite cell poles, thus ending in daughter cells. Figure source http: //www.dukeschumacherlab.com/research.html.

1.2.1.3 *Maintenance systems*

Toxin-antitoxin (T/A) systems are commonly encoded by plasmids and function to maintain plasmid stability (Jensen and Gerdes 1995). T/A systems are also present in chromosomal DNA, where their functions can vary significantly (Magnuson 2007). The most common organization of a T/A system is where the unstable antitoxin gene precedes the stable toxin gene. The regulation of plasmid stability by a T/A module occurs when the antitoxin binds to the toxin to neutralize its toxic effects. The antitoxin is synthesized at greater rates to inhibit the function of the toxin, any disruption of this continuous synthesis it would allow the toxin to be free and to potentially kill the host (Gerdes et al. 2005). When the toxin is liberated, daughter cells that do not inherit the copy of the plasmid with an intact T/A system will eventually be eliminated by a post-segregational killing (PSK) mechanism by the deleterious activity of the toxin (Van

Melderen and Saavedra De Bast 2009). PSK or addition systems are effective methods to get rid of potential intercellular competition and, thus, contribute to plasmid maintenance.

There are three different types of T/A systems prevalent in bacterial organisms. Type I T/A systems are composed of proteins for both the toxin and antitoxin components. Genetically, the antitoxin gene is located upstream of the toxin gene forming an operon. An example for type I T/A system is the *ccd* locus, which is primarily discovered in plasmids and in the chromosomes of a few Gram-negative bacteria (Pandey et al. 2005). The *ccd* locus consists of the *ccdA* and *ccdB* gene pair, in which the CcdB prevents cell division while CcdA suppresses the activity of the CcdB toxin. The CcdB toxin inhibits DNA replication by the inactivation of the DNA gyrase (Critchlow et al. 1997). The binding of the *ccdA* and *ccdB* genes prevent the activation of the Lon-dependent proteolysis of CcdA (Van Melderen et al. 1994).

The type II T/A system consists of an antisense RNA antitoxin and a protein toxin complex, and was first represented by the *hok*-Sok system in the R1 plasmid (Gerdes et al. 1990). The translation of the killer mRNA is repressed by the action of small unstable antisense RNA (Sok) that is complementary to the leader translation of reading frame (*mok*) that overlaps with the toxin encoding gene (*hok*) (Hayes 2004). As translation of the killer gene *hok* is coupled to that of the overlapping reading frame *mok*, the antisense RNA inhibits killer gene expression indirectly.

The third type of T/A system consists of a direct interaction between an RNA antitoxin and toxin protein. The ToxIN system discovered in the phytopathogen *Erwinia carotovora* functions in the abortion of infections by different phages (Fineran et al. 2009). The *toxI* gene suppresses the bacteriostatic effects of the ToxN protein as it prevents phage replication or synthesis.

1.2.2 *Plasmid accessory region*

One of the reasons that plasmids were studied extensively in the late twentieth century was their ability to convey genes that can have a considerable impact on human health and bacterial evolution (Jackson et al. 2011). Accessory regions generally include the genes that confer specific phenotypic characteristics to the host cell. The earliest described accessory function of plasmids was antibiotic resistance. After a few decades, bacterial genome sequencing accelerated the discovery of other types of accessory genes carried by plasmids (Mira et al. 2010). These are diverse, and include symbiosis genes that contribute to nitrogen fixation, virulence factors that might facilitate colonization of eukaryotic cells, and even mechanisms for detoxification of heavy metal substances. In most cases, these modules have significant differences in their G+C content compared to the plasmid backbone (Norman et al. 2009).

1.2.3 *Plasmid classification*

Strategies for classifying plasmids have been phenotypic characterisation, physical characterisation of the plasmid DNA (including restriction pattern), incompatibility testing, replicon typing and whole plasmid sequencing (Smalla et al. 2000). Incompatibility testing is, at present, the most common way to classify plasmids, and approximately 40 different groups of plasmids have been defined on the basis of incompatibility (Inc) properties (Couturier et al. 1988; Viegas et al. 1997). The classification of plasmids by incompatibility is based on the idea that two different plasmids cannot co-exist within the same host cell. Whenever two plasmids are incompatible (thus belonging to the same incompatibility group), one of them is segregated (Couturier et al. 1988). A scheme to classify plasmids based on incompatibility has been created (Datta and Hedges 1972). These groupings have been used to classify plasmids originating from Gram-negative and Gram positive bacteria.

1.3 *Salmonella plasmids*

The genome of *S.* Typhimurium 4/74 (Richardson et al. 2011) used in this study includes three plasmids, namely pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$. pSLT$^{4/74}$ (also known as the *Salmonella* virulence plasmid) is a 93KB plasmid that encodes 109 protein coding genes and pColIb$^{4/74}$ that encodes 106 genes. Finally, pRSF1010$^{4/74}$ is a low molecular-weight plasmid that carries 11 protein coding genes.

1.3.1 *Virulence plasmid*

Several *Salmonella* species possess a high molecular-weight plasmid ranging from 50kb to 100kb in size, which is required to cause systemic disease (Gulig and Curtiss 1987; Gulig et al. 1990) in mouse. Strains cured of this large virulence plasmid remain invasive, but lose the ability to grow intracellularly in macrophages (Gulig and Doyle, 1993; Gulig et al. 1998). The genes encoded by the plasmid are also required for bacterial replication in the liver and spleen (Guiney et al. 1995). The plasmid pSLT possesses a well-conserved 7.8 kb region encoding five *spv* (*salmonella* plasmid virulence) genes (for a review see Gulig et al. 1993). The *spv* genes are arranged in two transcriptional units: the first consists of *spvR*, which encodes a positive activator and the latter includes *spvA, spvB, spvC* and *spvD*. SpvR regulates both transcriptional units, whereas the histone-like protein (H-NS), the stationary phase sigma factor (RpoS), the leucine-responsive regulatory protein (Lrp), the integration host factor (IHF), and the growth phase are involved in the control of *spvR* (Robbe-Saule et al. 1997; Marshall et al. 1999; O'Byrne and Dorman 1994). There are three genes from the *spv* locus required for the virulence phenotype in mice: *spvR* and two structural genes *spvB* and *spvC*. SpvB prevents actin polymerisation by ADP-ribosylation of actin monomers, while SpvC has phosphothreonine lyase activity and has been shown to inhibit MAP kinase signalling (Guiney and Fierer 2011). Both SpvB and SpvC are translocated into the host cell by the SPI-2 T3SS. The *spvC* gene encodes the phosphothreonine lyase enzyme, and a *Salmonella* strain lacking spvC, reduces inflammation in the infected mice (Takeshi et al. 2011). The functions of the other Spv proteins (SpvA and SpvD) are not yet known.

The complete set of *tra* genes in pSLT$^{4/74}$ are vital for plasmid conjugation and suggest that an ancestral strain of *Salmonella* acquired this plasmid by conjugation (Rotger and Casadesus 1999). The study by García-Quintanilla et al. (2008), found that the frequency of conjugation is high in the mouse gut, suggesting that the pSLT

conjugation is happening during the *Salmonella* infection. The pSLT-encoded Rck (Heffernan et al. 1992) is a 17 kDa outer membrane protein that interacts with the complement system to confer serum resistance (Ho et al. 2011). Interestingly, the protein encoded by the *tra* operon, TraT also confers weak serum resistance. The mechanism of TraT serum resistance is unclear, but *traT*-mediated serum resistance is detectable in S. Typhimurium carrying freshly transferred plasmids (Glöckner et al. 1997; Guiney et al. 1994). The resistance disappears after a few generations, and is only effective against the alternative pathway. The pSLT plasmid also encodes *tlpA*, which encodes an apparent thermometer that regulates its own temperature-dependent transcription (Hurme et al. 1997). The stability of pSLT$^{4/74}$ is maintained by the *parA/B* Walker-box type partitioning system and the *ccdA/B* type I PSK system. García-Quintanilla et al. (2008) demonstrated that pSLT can be cured by inducing the Bile-mediated disruption of the *ccdA/B* PSK system.

The *pef* (plasmid encoded fimbriae) locus of pSLT$^{4/74}$ contains five genes (*pefBACDI*) and additional ORFs including *orf6, orf7, orf8, orf9* and *orf11*. The known gene *rck*, which confers resistance to serum, was found between *orf9* and *orf11.*A study by Ahmer et al. (1998) revealed that *orf8, orf9, rck* and *orf11* are regulated by a chromosomal gene *sdiA*. They renamed these ORFs *srgA, srgB* and *srgC* (*sdiA* regulated genes). The *pef* locus produces a thin (2-5 nm in diameter) flexible fibrillae composed of the major fimbrial subunit PefA (Nicholson and Low 2000). PefA expression is positively regulated by the global post-transcriptional regulator CsrA (Sterzenbach et al. 2013). PEF (plasmid-encoded fimbrial protein) has been shown to be important for adhesion to murine small intestine and fluid accumulation (Bäumler et al. 1996). Transcription of the *pef* operon, induced under acidic conditions, is modulated by the PefI regulatory protein (Nicholson and Low 2000). PefI represses PEF production by activating the Lrp mediated inhibition of DNA methylation within the *pef* promoter region (Nicholson and Low 2000). PefI expression also inhibits transcription of flagellar proteins and consequently cell motility (Wozniak et al. 2009).

1.3.2 *Colicinogenic plasmid*

The colicinogenic plasmid pCol1B$^{4/74}$ (86 kb) belongs to the incompatibility I1 (IncI1, also called IncIα) group of enterobacterial conjugative plasmids. Colicin is released into the environment to reduce the competition from other bacterial strains. pCol1B$^{4/74}$ also produces the immune protein (gene: *imm*) to protect its own cell from the toxic effect of

colicin. IncI1 plasmids produce two kinds of sex pili, a thin flexible pilus (*pil* operon) and a thick rigid pilus (*tra* operon) (Bradley 1984). IncI1 plasmids carry a multiple inversion system called shufflon (Komano et al. 1987), and encode the *sog* gene for suppression of *Escherichia coli dnaG* mutations (Wilkins et al. 1981).

The stability of pCol1B$^{4/74}$ is maintained by the *pnd* system, which kills the newborn plasmid-free cells (type II PSK; Nielsen and Gerdes 1995). The gene *pndA* encodes a stable toxin and the PndB gene encoding an unstable RNA antitoxin. As translation of the killer gene *pndA* is coupled to that of the overlapping reading frame *pndC*, the antisense RNA PndB inhibits killer gene expression indirectly. In addition, the stability of pCol1B$^{4/74}$ is maintained by a *stbA/B* partitioning system, which is similar to the *parM/R* actin-like ATPase partitioning system. Plasmid pCol1B is responsible for horizontal gene transfer via conjugation to *E. coli* during infection of the murine gut (Stecher et al. 2012).

1.3.3 *Resistance plasmid*

RSF1010$^{4/74}$, a small (8.6 KB) multicopy plasmid confers resistance to streptomycin (gene: *strA, B*) and sulphonamides (gene: *sul2*) and is mobilisable but not conjugative (Sakai and Komano, 1996). It was first recovered from an *E. coli* strain isolated from pig faeces and is one of the prototypes for the IncQ incompatibility group. RSF1010 has a very broad host range, because it can replicate in a wide range of bacteria, both Gram-negative and Gram-positive, and has been used as the basis of broad host range cloning vectors for many different organisms. The analysis of streptomycin resistance in Danish isolates of *Salmonella* Typhimurium confirms that *strA* and *strB* genes are important for streptomycin resistance in this bacterium (Madsen et al. 2000).

1.4 *Bacterial non-coding regulatory RNAs*

Bacterial non-coding small RNAs (sRNAs) are an emerging class of post-transcriptional regulators of coding gene expression; they are highly structured and contain several stem-loops. The sRNAs are identified in diverse prokaryotes like Archaea (Klein et al. 2002; Dennis and Omer 2005), Cyanobacteria (Axmann et al. 2005), Actinobacteria (Panek et al. 2008) and in Firmicutes (Silvaggi et al. 2006). However, the majority of the sRNA have been identified and studied in *Enterobacteriaceae*. The sRNAs are comparatively small in size, ~50 to 300 nt (Storz et al. 2011) and are involved in different regulatory mechanisms, including RNA-RNA base pairing (Altuvia et al. 1998) and RNA-protein interactions (Liu et al. 1997). In RNA-RNA base-pairing, many sRNAs bind to the ribosome-binding sites (RBS) of their mRNA targets, thus inhibiting ribosome entry on mRNAs (Storz et al. 2004). The sRNA OxyS inhibits ribosomal binding and translation by pairing with the short sequence overlapping the RBS of the *fhlA* mRNA (Altuvia et al. 1998). In contrast, the sRNA MicF inhibits translation and destabilizes the *ompF* mRNA by pairing with the 5'- untranslated region of the *ompF* transcript (Andersen et al. 1989; Andersen and Delihas 1990). The non-coding RNAs can further be categorised as *cis*-encoded antisense RNAs and *trans*-encoded sRNAs, based on their genomic location and functional role.

1.4.1 *Trans-encoded small RNAs*

Generally, the *trans*-encoded sRNAs are encoded in the intergenic region (IGR – the DNA sequence between two coding regions) and regulates the expression of mRNA from a distant location. They exhibit one or more short imperfect complementarity stretches to their target mRNAs. For example, RyhB sRNA interacts with *sodB* mRNA by a nine base-pair complex (Geissmann and Touati 2004). OxyS sRNA targets *fhlA* mRNA through the formation of two short duplexes of 9 and 7 base pairs (bp), respectively (Argaman and Altuvia 2000). Systematic mutational analysis of SgrS–*sopD2* RNA duplexes revealed that a single G-U base pair is the key to mediating SgrS repression of *sopD* mRNA (Papenfort et al. 2012). However, the sRNAs can also regulate multiple mRNA targets, e.g., DsrA is shown to directly act on the *hns* and *rpoS* mRNAs (Lease et al. 1998; Majdalani et al. 1998). OxyS sRNA is expressed in response to oxidative stress condition in the cell and regulates the expression of ~40 genes (Altuvia et al. 1997). GcvB was shown to regulate a large number of mRNAs in *E. coli* and *Salmonella*; e.g., it regulates the expression of OppA and DppA mRNAs in *E. coli,* which transport oligopeptides and dipeptides respectively (Urbanowski et al.

2000) and regulates multiple ABC transporters in *Salmonella* by targeting C/A-rich regions inside and upstream of RBS (Sharma et al. 2007). Most of the sRNAs studied in the enterobacterial family act on *trans*-encoded target mRNAs through short base-pairing interactions are commonly aided by the RNA chaperone, Hfq (Valentin-Hansen et al. 2004; Sittka et al. 2008). Hfq was first discovered in the mid-1960s in a screen for host proteins required for replication of Q-beta bacteriophage in *E. coli* and is capable of simultaneously interacting with two distinct RNA molecules. Hfq significantly enhances the hybridisation and duplex formation between sRNA and target mRNA (Vogel and Luisi 2011). Figure 1.5 explains the Hfq mediated sRNA regulation on expression of target mRNA.

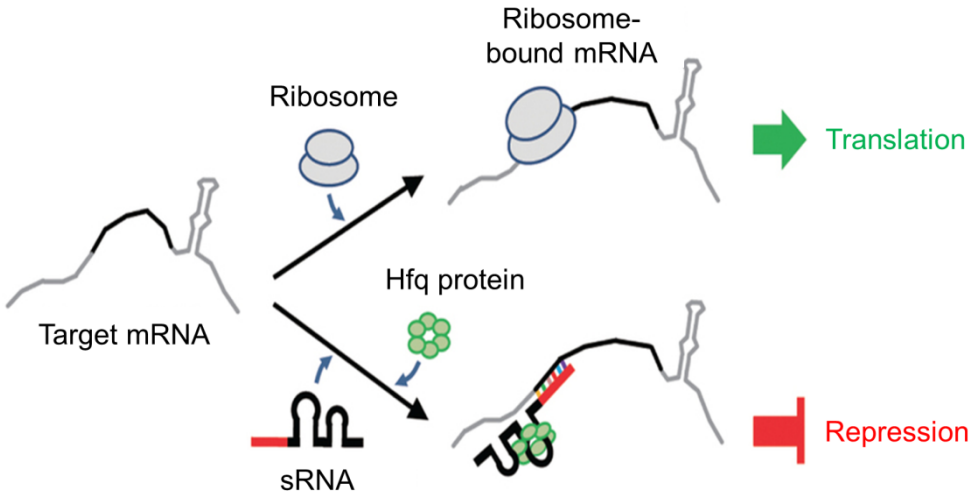


Figure 1.5: Example mechanism of *trans*-encoded sRNA alters the expression of target mRNA in Hfq-dependent manner.
The sRNA has the sequence (red coloured region in sRNA) complementary to the target mRNA, which compete with ribosome and binds to the target mRNA. The step-loop structure of the sRNA (black coloured region in sRNA) recruits the Hfq protein that facilitates the binding of the sRNA to the target mRNA. The duplex formation between sRNA and target mRNA enables sRNA to repress the translation of mRNA. Figure adapted from Yoo et al. (2013) with minor modifications.

1.4.2 *Cis-encoded antisense RNAs*

The *cis*-encoded antisense transcripts (asRNAs) are defined as non-protein-coding RNA molecules which are transcribed in the opposite strand to protein coding genes. They range from ~100 to 300 nt (Brantl 2002; 2007); but can be longer (700 to 3500 nt, André et al. 2008). One extraordinarily long asRNA in *Protocholorococcus* strain MED4 is 7000 nt and overlaps 14 genes of a ribosomal protein operon (Stazic et al. 2011). Individual studies in a variety of bacterial species determined the presence of asRNAs

in bacterial systems. However, the development of a new approach based on high-throughput sequencing techniques has revolutionized the identification of asRNAs (Georg et al. 2009). Previously, it was believed that the bacterial genomes mainly encode proteins, but accumulating evidence suggests that extensive antisense transcription occurs in bacteria (Selinger et al. 2000; Cho et al. 2009; Dornenburg et al. 2010; Lasa et al. 2011). Although some antisense transcripts have specific regulatory functions, it has been proposed that others result from 'transcriptional noise' generated by nonspecific transcription initiation or weak promoters that become fixed within genes by evolutionary constraints on the coding sequence (Struhl 2007).

Generally, asRNAs regulate the expression of the genes located in the sense strand on the DNA. The first asRNA was discovered in 1981, and is involved in the replication control of *E. coli* plasmid pColE1 (Tomizawa et al. 1981). The asRNAs in bacteria are functionally classified as 5' overlapping (head to head), internal and 3' overlapping (tail to tail) asRNAs based on where the asRNA binds on the sense mRNA to form the duplex (Lee and Groisman 2010; Stazic et al. 2011). As asRNA interacts with target RNA, the secondary structures of both interacting molecules are altered resulting in duplex (dsRNA) formation. In the majority of cases, the duplex formation will result in complete degradation of both RNAs, for example the *isiA/IsrR* sense/antisense pair (Dühring et al. 2006). In a few cases, the asRNA can stabilise the sense mRNA, for example *gadX*/GadY sense/antisense pair (Opdyke et al. 2010). Recent evidence suggests that *cis*-encoded regulation of gene expression by asRNAs constitutes a distinct layer of control in bacteria (Opdyke et al. 2010; Sesto et al. 2013). Unlike chromosomal asRNAs, the functional role of plasmid-encoded asRNAs has been extensively studied, including regulatory roles in plasmid stability, replication, conjugation and control of post-segregational killing (Brantl 2002; 2007).

1.5 *Aims of this study*

The aims of this doctoral thesis can be divided into two broad areas. The first aim was to develop a pipeline to analyse the high-throughput data generated by the Hinton lab using a bioinformatics approach and the second aim was to study the RNA transcripts of *Salmonella* plasmids using different experimental methods.

1.5.1 *Development of pipeline*

RNA-Seq is a recently developed transcriptome profiling approach that uses next-generation sequencing technologies. The size of the data produced by current deep-sequencing methods varies from few hundred thousand to several million short reads, depending on the sequencing platform used. The major challenges with RNA-Seq data are to align millions of sequenced reads onto the reference genome and to perform the downstream analyses, including visualisation of RNA-Seq data and measuring transcript abundance for each gene. The effective methods to store, retrieve and process large amounts of data also play a major role in RNA-Seq data handling and can impact on the quality of the results. The first aim of this study was to develop an RNA-Seq analysis pipeline to efficiently map sequenced cDNA reads to the genome and to perform downstream processes with the help of publicly available software tools and custom Perl scripts. The development of the RNA-Seq pipeline and its applications are discussed in Chapter 3.

1.5.2 *RNA biology of Salmonella plasmids*

Plasmids are believed to play a major role in bacterial adaptation to environmental changes and contribute to overall plasticity of the bacterial genome plasticity. Plasmids are also important when carried by medically relevant bacteria due to their ability to mobilise and disseminate antibiotic resistance genes and virulence genes, which may impact on the pathogenicity of a bacterium. The second major research focus of this study was to investigate the transcriptional landscape of the three plasmids carried by *S.* Typhimurium 4/74. To study the plasmid gene expression and regulation, the RNA-Seq libraries generated from the bacterium grown under multiple environmental (also infection-relevant) conditions were utilised extensively.

Aims of the plasmid project are as follows:

- To identify plasmid-associated transcriptional start sites (TSS) under different growth conditions (Chapter 4 and 5).
- To identify antisense RNAs (asRNAs) and small RNAs (sRNAs) expressed by plasmids of *S*. Typhimurium under infection-relevant growth conditions (Chapter 4).
- To identify the expression profiles of plasmid-encoded protein coding gene as well as sRNA/asRNA across 23 different environmental, stress and infection-relevant growth conditions (Chapter 5).
- To identify a novel regulatory inter cross-talk between chromosomal regulatory genes and plasmid-encoded genes (Chapter 5).
- To study the regulatory functional role of some newly identified plasmid-encoded sRNA/asRNAs of *S*. Typhimurium (Chapter 6).

CHAPTER 2

MATERIALS AND METHODS

Standard microbiological and molecular biology techniques used routinely in this study are described in this chapter. All solutions and growth media were made with water purified through de-ionising and filtration cartridges, MilliQ® (Millipore), certified to a standard of at least 18.2 MΩ/cm, and referred to here as dH_2O. All steps involving DNA or RNA elution, or enzymatic reactions were done with molecular biology grade water (Sigma, Cat no: 95284), which is certified DNase and RNase free (DEPC treated H_2O). Solutions were stored at room temperature unless otherwise indicated. All media and solutions used throughout this study were autoclaved at 121°C, 15 lbs pressure for 15 minutes, or by filter-sterilisation. The glassware was autoclaved as above, and dried within the autoclave machine by exploring the dry-cycle option. The sterility of the autoclaved products was ensured using indicator tapes (VWR, Cat. No. 489-1311). Some solutions were filter-sterilised with 0.22 µm membrane filters (Millipore Express®).

2.1 Standard growth conditions, media and concentration of antibiotics

Bacteria were grown in Lennox broth (LB, Lennox 1955) medium in 250 ml flasks at 37°C, 220 rpm with normal aeration or on LB agar plates, both supplemented with the appropriate antibiotics throughout this study unless stated otherwise. Cultures were inoculated from a single colony grown overnight on plate at 37°C or overnight cultures (inoculated from single colony) were diluted 1:1,000 into fresh medium. Antibiotics (where appropriate) were used at the following final concentrations: 100 µg/ml ampicillin, 50 µg/ml kanamycin, 25 µg/ml chloramphenicol.

2.1.1 *Liquid media*

LB medium

Bacto-Tryptone	(Becton Dickinson, Cat. No. 211705)	10 g/L
Bacto-Yeast Extract	(Becton Dickinson, Cat. No. 212750)	5 g/L
Sodium chloride (NaCl)	(Sigma, Cat. No. S3014)	5 g/L

Dry reagents were initially dissolved in 800 ml dH_2O, pH adjusted to 7.0 with a few drops of 1M NaOH or HCl, before being made up to 1.0 litre with water, then autoclaved.

PCN medium

PCN (Phosphate Carbon Nitrogen medium pH 5.8, Löber et al. 2006) minimal medium was used to study the functional role of plasmid-encoded sRNA STncP1-1 using RNA-Seq approach (*Section 6.2.1.4*).

MES (pH 5.8)	(Sigma, Cat. No. M8250)	80 mM
Tricine	(Sigma, Cat. No. T5816)	4 mM
Ferric chloride ($FeCl_3$)	(Sigma, Cat. No. 236489)	100 µM
Potassium sulfate (K_2SO_4)	(Sigma, Cat. No. P0772)	376 µM
Sodium chloride (NaCl)	(Sigma, Cat. No. S3014)	50 mM
Glucose ($C_6H_{12}O_6$)	(Sigma, Cal. No. G8270)	22.2 mM
Ammonium chloride (NH_4Cl)	(Sigma, Cat. No. A9434)	15 mM
Magnesium sulfate ($MgSO_4$)	(Sigma, Cat. No. 63139)	1 mM
Calcium chloride ($CaCl_2$)	(Sigma, Cat. No. C3306)	0.01 mM
Sodium molybdate (Na_2MoO_4)	(Sigma, Cat. No. 24365)	10 nM
Sodium selenite (Na_2SeO_3)	(Sigma, Cat. No. S5261)	10 nM
Boric acid (H_3BO_3)	(Sigma, Cat. No. B6768)	4 nM
Cobalt(II) chloride ($CoCl_2$)	(Sigma, Cat. No. C8661)	300 nM
Cupric sulfate ($CuSO_4$)	(Sigma, Cat. No. C1297)	100 nM
Manganese(II) chloride ($MnCl_2$)	(Sigma, Cat. No. 529680)	800 nM
Zinc sulfate ($ZnSO_4$)	(Sigma, Cat. No. Z4750)	1 nM
K_2HPO_4/ KH_2PO_4	(Sigma, Cat. No. 60353/ P8416)	0.4 mM

Dipotassium phosphate/Monopotassium phosphate (K_2HPO_4/ KH_2PO_4)

K_2HPO_4/ KH_2PO_4 buffer was prepared separately by adding 850 µl of 1 M KH_2PO_4 and 9.15 ml of 1 M K_2HPO_4 into 80 ml of dH_2O. pH was adjusted to 5.8 and made up to 100 ml with dH_2O. Similarly, appropriate molar concentration solutions of each ingredient (chemicals) were prepared separately. Appropriate amounts of each solution to make 1 L medium were mixed in a sterile bottle and sterilised using 0.22 µm membrane filter (Millipore Express®) in a fresh sterile bottle and stored at room temperature.

LowMg^{2+} medium

The ingredients present in the LowMg^{2+} medium were same as in PCN medium except the concentration of magnesium sulphate. Here, a low concentration (10 µM) of magnesium sulphate was used. This medium was used to experimentally validate the expression of STncP1-1 by Northern blot (*Section 6.2.1.1*). LowMg^{2+} medium was also used to study the functional role of STncP1-1 using pulse-expression of sRNA followed by microarray analysis.

SOC medium

SOC (Super Optimal Broth) medium was used in the final step of bacterial transformation to obtain maximal transformation efficiency of bacterial cell (e.g. *E. coli* TOP10).

Bacto-Tryptone	(Becton Dickinson, Cat. No. 211705)	20 g/L
Bacto-Yeast Extract	(Becton Dickinson, Cat. No. 212750)	5 g/L
Sodium chloride (NaCl)	(Sigma, Cat. No. S3014)	0.5 g/L
Potassium chloride (KCl)	(Sigma, Cat. No. P9541)	0.186 g/L
Magnesium sulfate ($MgSO_4$)	(Sigma, Cat. No. 63139)	4.8 g/L
Glucose ($C_6H_{12}O_6$)	(Sigma, Cal. No. G8270)	3.6 g/L

2.1.2 *Solid media*

LB agar

LB agar was prepared by the addition of Bacto-Agar (Becton Dickinson, Cat No. 214010) to LB medium pH 7.0 at 15 g/L prior to autoclaving. Antibiotics were added to the solid media after autoclaving once the molten agar had cooled to between 45 to 50°C.

Green indicator plates

Green indicator plates used after P22 transduction to differentiate unstable P22 pseudo-lysogens from true lysogens were prepared as previously described (Maloy *et al.*, 1996).

Bacto-Tryptone	(Becton Dickinson, Cat. No. 211705)	8.0 g/L
Bacto-Yeast Extract	(Becton Dickinson, Cat. No. 212750)	1.0 g/L
Sodium chloride (NaCl)	(Sigma, Cat. No. S3014)	5.0 g/L
Bacto-Agar	(Becton Dickinson, Cat No. 214010)	15 g/L

After autoclaving, the base medium was allowed to cool to 45 to 55°C and 21 ml of 40% (v/v) glucose (VWR, Cat. No. 101176K), 25 ml of 2.5% (v/v) alizarin yellow G (Sigma, Cat. No. 20,670-9) and 3.3 ml of 2% (v/v) aniline blue (Merck, Cat. No. 1.16316) added. All additives were autoclaved prior to addition to Green Agar, except the aniline blue stock that was filter sterilised, and the alizarin yellow G that was dissolved by heating gently in a microwave.

2.1.3 *Storage of bacterial stocks*

Bacterial strain, plasmids and growth conditions used in this study are shown in Table 2.1, 2.2 and 2.3, respectively. Long-term storage of bacterial strains was at -80°C, as glycerol stocks. Glycerol stocks were made by adding glycerol to final concentration of 30% with the overnight culture in LB medium. Strains were streaked onto LB agar plates as required and incubated overnight at 37°C. Short-term storage of strains on agar plates was performed at 4°C for a maximum of one week.

Table 2.1: List of bacterial strains used in this study.

Strain	Genotype	Reference/source
E. coli		
TOP10	*mcrA* Δ(*mrr hsdRMS mcrBC*) Φ80*lacZ*ΔM15 Δ*lacX74 deoR recA1 araD139* Δ(*ara leu*)7697 *galU galK rpsL endA1 nupG* λ-	Invitrogen
***S.* Typhimurium**		
4/74	Wild type	Wray and Sojka, 1978, Richardson et al. 2011
SL1344	Wild type	Sanger Institute
D23580	Wild type	Kingsley et al. 2009
JH3762	4/74 ΔSTncP1-1	this study

Table 2.2: List of plasmids used in this study.

Name	Relevant fragment	Comment	Origin/ marker	Reference
pXG10-sf	P_{LtetO}-*lacZ::sf-gfp*	standard plasmid for directional cloning of a target mRNA as N-translational fusion to sf-GFP	pSC101* / Cm^R	Corcoran et al. 2012
pXG10-sf-*gtrA*	P_{LtetO}-*gtrA::gfp*	*Salmonella gtrA* translational fusion plasmid	pSC101* / Cm^R	this study
pXG10-sf-*galP*	P_{LtetO}-*galP::gfp*	*Salmonella galP* translational fusion plasmid	pSC101* / Cm^R	this study
pP_L		standard plasmid for directional cloning of sRNA	P15A/ Amp^R	Corcoran et al. 2012
pP_L-P1-1	P_{LlacO}-STncP1-1	*Salmonella* STncP1-1 inserted in pP_L plasmid	P15A/ Amp^R	this study
pBAD/HisA		standard plasmid for directional cloning of sRNA in L-Arabinose-inducible pBAD vector	pBR322/ Amp^R	Invitrogen
pBAD474P1AS_50	P_{BAD}-474P1AS_50	*Salmonella* 474P1AS_50 transcriptional fusion plasmid	pBR322/ Amp^R	this study
pBAD474P1AS_90	P_{BAD}-474P1AS_90	*Salmonella* 474P1AS _90 transcriptional fusion plasmid	pBR322/ Amp^R	this study
pBADSTncP1-1	P_{BAD}-STncP1-1	*Salmonella* STncP1-1 transcriptional fusion plasmid	pBR322/ Amp^R	this study
pBAD474P1AS_110	P_{BAD}-474P1AS_110	*Salmonella* 474P1AS_110 transcriptional fusion plasmid	pBR322/ Amp^R	this study
pBAD474P1AS_180	P_{BAD}-474P1AS_180	*Salmonella* 474P1AS _180 transcriptional fusion plasmid	pBR322/ Amp^R	this study
pJV300		2-plasmid control plasmid	P15A/ Amp^R	Sharma et al. 2007
pKP8-35		pBAD control plasmid	pBR322/ Amp^R	Invitrogen
pKD4		Template for mutant construction; carries a kanamycin resistance cassette	oriRγ/Kan^R	Datsenko and Wanner 2000
pKD46	ParaB-*γ-β-exo*	Temperature sensitive red recombinase expression plasmid	oriR101 /Amp^R	Datsenko and Wanner 2000
pCP20		Temperature sensitive FLP recombinase expression plasmid	oriR101 / Amp^R, Cm^R	Datsenko and Wanner 2000

Table 2.3: List of 23 infection relevant growth conditions used in this study.

Name	Growth conditions
EEP	Early exponential phase; Growth in Lennox broth to OD_{600} 0.1.
MEP	Mid exponential phase; Growth in Lennox broth to OD_{600} 0.3.
LEP	Late exponential phase; Growth in Lennox broth to OD_{600} 1.0.
ESP	Early stationary phase; Growth in Lennox broth to OD_{600} 2.0.
LSP	Late stationary phase; Growth in Lennox broth to OD_{600} 2.0, followed by a further 6 h.
25°C	Low temperature; Growth in Lennox broth to OD_{600} 0.3 at 25°C.
Cold shock	Growth in Lennox broth to OD_{600} 0.3, then transfer for 10 min from 37°C to 15°C.
pH3 shock	Growth in Lennox broth to OD_{600} 0.3; then cells were harvested by centrifugation, resuspended in fresh Lennox pH3.0 (adjusted with HCl) and grown for additional 10 min.
pH5.8 shock	Growth in Lennox broth to OD_{600} 0.3; then cells were harvested by centrifugation, resuspended in fresh Lennox pH5.8 (adjusted with HCl) and grown for additional 10 min.
NaCl shock	Growth in Lennox broth to OD_{600} 0.3; then addition of NaCl (Sigma, Cat. S3014) to a final concentration of 0.3 M for 10 min.
Bile shock	Growth in Lennox broth to OD_{600} 0.3; then addition of bile (Sigma, Cat. S9875) to a final concentration of 3% for 10 min (Prouty and Gunn, 2000).
LowFe^{2+} shock	Growth in Lennox broth to OD_{600} 0.3; then addition of 2,2'-dipyridyl (Sigma, Cat. D21, 630) to a final concentration of 0.2 mM for 10 min (McHugh et al., 2003).
Anaerobic shock	Growth in Lennox broth to OD_{600} 0.3 (50 ml); then filled into 50 ml closed Falcon tube and incubated without agitation at 37°C for 30 min.
Anaerobic growth	Static growth in Lennox broth to OD_{600} 0.3 in a completely filled and closed 50 ml Falcon tube.
Oxygen shock	Static growth in Lennox broth to OD_{600} 0.3 in a completely filled and closed 50 ml Falcon tube; then 15 min aerobic growth (baffled flask, 250 rpm).
NonSPI2	Growth in PCN medium (pH 7.4, 25 mM Pi) to OD_{600} 0.3.
InSPI2	Growth in PCN medium (pH 5.8, 0.4 mM Pi) to OD_{600} 0.3 (Löber et al., 2006).
LowMg^{2+} (InSPI2)	Growth in PCN medium (pH 5.8, 0.4 mM Pi) with 10 µM $MgSO_4$ (Sigma, Cat. 63139) to OD_{600} 0.3.
Peroxide shock (InSPI2)	Growth in PCN (pH 5.8, 0.4 mM Pi) to OD_{600} 0.3; then addition of H_2O_2 (Merck, Cat. 822287) to final concentration of 1 mM H_2O_2 for 12 min (Wright et al., 2009).
Nitric oxide shock (InSPI2)	Growth in PCN medium (pH 5.8, 0.4 mM Pi) to OD_{600} 0.3; then addition of 250 µM Spermine NONOate (Sigma, Cat. S150) for 20 min.
Temp10	Growth in Lennox broth to OD_{600} 0.3 at 25°C; then transfer to 37°C for 10 min.
Temp20	Growth in Lennox broth to OD_{600} 0.3 at 25°C; then transfer to 37°C for 20 min.
MAC	Growth within the macrophage cell (RAW264.7) for 8 h

The table adapted from Kröger et al. (2013).

2.2 *Nucleic acid techniques*

2.2.1 *Genomic DNA extraction*

A 5 ml of overnight bacterial (enteric bacteria, e.g., *S*. Typhimurium 4/74) culture was centrifuged for 1 min at 10,000 rpm to settle down all bacterial cells. The supernatant was discarded and the cells were resuspended in 200 µl of dH_2O. The cells were boiled for 5 min at 100^oC and vertex briefly to ensure the disruption of the cell wall. The mixture was centrifuged for 5 min at 14,000 rpm, room temperature and the supernatant was transferred into a new 1.5 ml Eppendorf cap tube. One ml of chloroform (Sigma, Cat. No. 25666) was added into supernatant and vortex for 30 sec to ensure mixing of the aqueous and chloroform phase. Then the mixture was centrifuged at 4°C for 10 min at 14,000 rpm and the aqueous (upper) phase was carefully transferred into a new 1.5 ml Eppendorf cap tube. The total genomic DNA concentration was measured (*Section 2.2.3*) and stored at -20°C until used.

2.2.2 *Total RNA extraction*

Total RNA was extracted using TRIzol® (Invitrogen). Bacterial culture was grown in an appropriate growth condition and necessary amount of culture was taken for RNA extraction (usually 4 to 5 OD_{600} for 60 to 80 µg of RNA yield). Appropriate amount of ice-cold phenol/ethanol solution was added to culture (e.g., add 1.6 ml of phenol/enthanol solution to 4 ml of culture) and mixed gently. The mixture was incubated on ice for 30 min but no longer than 2 h to stabilise the RNA and prevent the degradation. The mixture was centrifuged for 10 min at 4,000 rpm at 4°C. After discarding the supernatant, the bacterial pellet was dissolved in 1 ml TRIzol® reagent (Invitrogen). The mixture was transferred to 2 ml Phase lock tubes (Phase Lock Gel Heavy 2ml, 5PRIME, Cat. No. 2302830), and mixed gently. After addition of 400 µl chloroform, the samples were mixed by shaking and centrifuged for 15 min at 14,000 rpm at 15-20°C. The supernatant was transferred to a fresh reaction tube and the RNA was precipitated by addition of 450 µl of isopropanol. The RNAs was pelleted by centrifugation for 30 min at 14,000 rpm at 15-20°C. After a wash step with 350 µl of 75% ethanol and additional centrifugation for 10 min at 14,000 rpm at 4°C, the supernatant was discarded, the pellets air-dried, and subsequently dissolved in DEPC treated H_2O (Sigma, Cat no: 95284). The quality and quality of the RNA was measured (*Section 2.2.3* and *2.2.4*) and the RNA stored at -80°C until used.

Phenol/ethanol solution

Phenol (pH 4.3)	(Sigma, Cat. No. P4682)	5% (v/v)
Ethanol	(Sigma, Cat. No. E7023)	95% (v/v)

2.2.3 *Quantification of nucleic acid concentration*

Measurement of nucleic acid concentrations was performed using the NanoDrop ND-1000 Spectrophotometer (NanoDrop, Wilmington, Delaware, USA) with accompanying ND-1000 software. The quantification was performed by measuring the absorbance of 1 µl of eluted nucleic acid sample at A_{230}, A_{260} and A_{280} readings, which gives an indication of the purity of the nucleic acid in relation to possible contaminants.

2.2.4 *Assessment of RNA quality*

In order to ensure the RNA was of a suitable quality to be used for transcriptomic studies, the Agilent 2100 Bioanalyser was used (Agilent, Stockport, UK), as per the described protocol instructed by manufacturer. This uses nucleic acid dyes and capillary electrophoresis to measure the size of RNA species present in a sample. A decimal dilution of each RNA sample is carried out and 1 µl run in each well on the RNA Nanochip along with 150 ng RNA 6000 ladder (Ambion, 7152, Austin, Texas, USA). The quality of the RNA was visualised using an electropherogram generated by the accompanying Agilent Bioanalyser software. Degraded RNA showed only small peaks if any, and had an uneven background.

2.2.5 *DNase I digestion of RNA samples*

Appropriate amount (40 µg) of RNA in 79 µl of dH_2O was denatured for 5 min at 65°C and cooled on ice for 5 min. The reaction mixture was prepared by adding 10 µl of 10X DNase I buffer, including MgCl2 (Fermentas, Cat. No. EN0521), 1 µl of Superase-In RNase inhibitor (20 U/µl) (Ambion, Cat. No. AM2694) and 10 µl of DNase I (1 U/µl) (Fermentas, Cat. No. EN0521) and incubated for 30 to 45 min at 37°C. After the incubation, the DNase I digested sample was transferred to 2 ml Phase lock tubes (Phase Lock Gel Heavy 2ml, 5PRIME, Cat. No. 2302830) along with equal volume of (100 µl) P/C/I mix (phenol/chloroform/isoamyl alcohol Sigma, Cat. No. P3803) and mixed gently for 15 sec (do not vertex). The mixture was centrifuged at 15°C for 12 min at 13,000 rpm. The aqueous (upper) phase was transferred to fresh 1.5 ml Eppendorf tube (VWR, Cat. No. 700-5239) and 2.5 volume (~300 µl) of 30:1 mixture of ethanol/3

M sodium acetate (pH 6.5) was added. The mixture was incubated for at least one hour at room temperature or overnight at -20°C. The RNAs was pelleted by centrifugation for 30 min at 14,000 rpm at 4°C. After a wash step with 350 µl of 75% ethanol and additional centrifugation for 10 min at 14,000 rpm at 4°C, the supernatant was discarded, the pellets air-dried, and subsequently dissolved in DEPC treated H_2O (Sigma, Cat No: 95284). The RNA sample was incubated at 65°C, 900 rpm on thermo-shaker and vertex two to three times in between incubation. The quality and quality of the RNA was measured (*Section 2.2.3* and *2.2.4*) and stored at -80°C until used.

2.2.6 *Northern blot using DIG reagent kit*

To validate the expression of plasmid-encoded sRNA/asRNAs, Northern blot experiments were performed using the DIG Northern Starter Kit (Roche, cat no. 12039672910). The DIG labelled riboprobes was prepared and checked according to the manufacturer's instructions. The primer pairs used for making riboprobes are listed in Table 2.4. To detect plasmid encoding sRNAs, 5 µg DNase I digested (*Section 2.2.5*) total RNA was separated on 7% denaturing (8.3 M urea) polyacrylamide gel. After a 30 min transfer to Nylon membranes (Roche Diagnnostics GmbH, Cat. No. 11209272001) in a blotter at 350 mA and 4°C in the presence of 1× TBE buffer, the RNA was cross-linked to the membrane 1 min with UV-crosslinker (Bio-Link BLX 254, PeQLab Biotechnologies GmbH, UK). After pre-hybridization for 1 h in 15 ml DIG easy starter kit at 62°C, the DIG labelled riboprobe was added. The membrane was hybridized for a period of 12 - 14 h at 62°C and washed in 2X SSC buffer for 5 min at 15-20°C and washed twice in 0.1X SSC at 68°C for 15 min. After incubation in blocking solution for 30 min at 15-25°C, the membrane was incubated in antibody solution (DIG Northern starter kit, Roche Diagnostics Ltd, USA) for 30 min at 15-25°C. Then membrane was washed twice in washing buffer for 15 min at 15-25°C. Finally the membrane was incubated in detection buffer for 2-5 min at 15-25°C and the chemiluminescence (band) was observed in chemiluminator (FluorChem M FM0254, Labtech, France).

Polyacrylamide (7%) gel

PAA solution 40% (19:1 acrylamide/bisacrylamide)	175 ml
Urea (8.3 M)	500 g
10X TBE buffer	100 ml

Add dH_2O to a final volume of 1 L.

All gel equipment was cleaned with 70% ethanol and dH_2O before use (glass plates, spaces and combs etc.). Polymerisation was initiated by addition of 1/100 volume of ammonium persulfate (APS) and 1/1000 volume of N,N,N,N,-Tetramethylethylenediamine (TEMED). Prior loading, RNA samples were denatured for 5 min at 65°C in equal volume of 2X RNA loading dye.

RNA loading dye

95% formamide, 0.025% Sodium dodecyl sulphate (SDS), 0.025% bromophenol blue, 0.025% xylene cyanol FF, 0.025% ethidium bromide and 0.5 mM EDTA.

10X TBE buffer

0.89 M Tris, 0.89 M Boric acid and 20 mM EDTA (pH 8.0)

2X SSC buffer

300 mM Sodium chloride, 30 mM Sodium citrate and 0.1 % SDS (pH 7.0).

Washing buffer

0.1 M Maleic acid, 0.15 M Sodium chloride and 0.3% Tween 20 (pH 7.5).

Detection buffer

0.1 M Tris-HCl and 0.1 M Sodium chloride (pH 9.5)

Table 2.4: Primers used to validate the expression of plasmid-encoded sRNA/asRNAs are listed here.

Gene	Oligo name	Sequences	Purpose
474P1AS_30	P1_30_for	ggcggcacgctgtccgg	Riboprobe template for 474P1AS_30
474P1AS_30	P1_30_rev	gaattaatacgactcactataaggaggcggtcacgccac	Riboprobe template for 474P1AS_30
FinP	FinP_for	acacataggaacctcctcaaa	Riboprobe template for FinP
FinP	FinP_rev	gaattaatacgactcactatattaaatcgccgatacagggag	Riboprobe template for FinP
474P1AS_50	P1_50_for	actgacattattgatactccttc	Riboprobe template for 474P1AS_50
474P1AS_50	P1_50_rev	gaattaatacgactcactataaggaggtgccggacggg	Riboprobe template for 474P1AS_50
474P1AS_90	P1_90_for	gatcatcctccaggcccag	Riboprobe template for 474P1AS_90
474P1AS_90	P1_90_rev	gaattaatacgactcactataccacatatacagtcactcgat	Riboprobe template for 474P1AS_90
474P1AS_180	P1_180_for	atagcctgattggtggctatac	Riboprobe template for 474P1AS_180
474P1AS_180	P1_180_rev	gaattaatacgactcactatagatgaggtgtctaccc	Riboprobe template for 474P1AS_180
474P2AS_10	P2_10_for	gtattctgtgaggccccca	Riboprobe template for 474P2AS_10
474P2AS_10	P2_10_rev	gaattaatacgactcactataaaggaaaaccccccactatctt	Riboprobe template for 474P2AS_10
474P2AS_50	P2_50_for	gtctcatggcataatctccat	Riboprobe template for 474P2AS_50
474P2AS_50	P2_50_rev	gaattaatacgactcactataccgaaatgaaaaaaaaggcagg	Riboprobe template for 474P2AS_50
474P2AS_80	P2_80_for	gtctaggcattacagtagagc	Riboprobe template for 474P2AS_80
474P2AS_80	P2_80_rev	gaattaatacgactcactataagcagaaagccctggaga	Riboprobe template for 474P2AS_80
474P2AS_90	P2_90_for	gtcagttatagacagttctttttc	Riboprobe template for 474P2AS_90
474P2AS_90	P2_90_rev	gaattaatacgactcactatacgggcggatttgagaaaca	Riboprobe template for 474P2AS_90
STncP1-1	P1-1_for	accttatccagtgagcgct	Riboprobe template for STncP1-1
STncP1-1	P1-1_rev	gaattaatacgactcactatacacgccaaaacagagaagtaat	Riboprobe template for STncP1-1

2.2.7 *Rapid amplification of cDNA ends (5' RACE)*

Mapping of transcriptional start sites using 5' RACE followed the protocol described in Argaman et al. (2001). In detail, 12 µg of DNase I digested (*Section 2.2.5*) total RNA was adjusted to a volume of 87.5 µl by adding the required volume of dH_2O. 10 µl of 10X TAP buffer and 0.5 µl SUPERase-In RNase inhibitor were added and samples split into two reactions of 49 µl each. Following treatment of one reaction with 10 units tobacco acid pyrophosphatase (TAP) for 30 min at 37°C, 300 pmol of RNA-linker A4 (encoding a BseRI restriction sit) was added to both reactions.

RNA linker A4: GACGAGCACGAGGACACUGACGUGGAGGAGGGAGUAGAAA

The RNA sample was purified using P/C/I methods (for details see *Section 2.2.5*) and ethanol precipitated. Then the RNA was dissolved in 13.5 µl dH_2O, denatured for 5 min at 90°C and chilled on ice for 5 min. The RNA-linker ligation was performed overnight (12 h) at 17°C in presence of 40 units T4 RNA ligase, 1X RNA ligase buffer, 10% v/v DMSO (final concentration) and 20 units SUPERase-In RNase Inhibitor. Following P/C/I extraction and ethanol precipitation, 2 µg linker-ligated RNA was converted to cDNA using 100 pmol random hexamer primers and the Superscript III (200 units) reverse transcription kit in a 20 µl reaction. 10 min incubation at 25°C was carried out before addition of the reverse transcriptase, followed by four subsequent 15 min incubation steps at 42°C, 50°C, 55°C and 60°C. After heat inactivation of the reverse transcriptase for 5 min at 85°C, samples were treated with 1 unit RNase H (New England Biolabs) at 37°C for 20 min.

One µl of the cDNA samples was subsequently used as template in standard Taq polymerase PCR reactions using the sense primer JVO-0367, which anneals to the RNA-linker sequence, and an antisense primer that anneals within the gene of interest (Table 2.5). The PCR products were visualised on 3% agarose gels to ensure the enrichment in TAP treated samples (indicating primary transcripts). The enriched bands from gel were excised and purified using the QIAquick® (QIAGEN. Cat. No. 28704) Gel Extraction Kit following the manufacturer's instructions. The purified inserts were cloned to TOPO® vector (Invitrogen, Cat. No K4500-01) following the manufacturer's instructions, and chemically transformed to *E. coli* TOP10 cell. A negative control of TOT10® cloning was performed water instead of inserts. The TOP10 cells were incubated in SOC medium for 1 h at 37°C, 220 rpm. After the 1 h incubation, the cells were plated and incubated for overnight. The next day, 10 colonies

from each sample were transferred to 5 ml fresh LB medium for overnight incubation at 37°C, 220 rpm. The TOPO® vector was extracted using the QIAprep® (QIAGEN. Cat. No. 27104) Spin Miniprep Kit as per the manufacturer's instructions. The extracted samples were sequenced to confirm the initiating nucleotide of the transcripts.

Table 2.5: List of primers used for 5' RACE experiments.

Gene	Oligo name	Sequences	Purpose
	JVO-0367	ACTGACATGGAGGAGGGA	General forward primer
474P1AS_170	5R_P1_170	TGCTGGAATGCAGGATGAA	5'-RACE; 474P1AS_170
sulII	5R_sul2	ATGCCGGACCGAGGTCGA	5'-RACE; *sulII*

PCR reaction mix

Components	Volume	Final concentration
5X Hi-Fi Reaction Buffer	10 µl	1X
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	1 µl	0.2 µM
10 µM Reverse Primer	1 µl	0.2 µM
Template	1 µl	100 ng
Velocity DNA Polymerase	1 µl	2 units
DEPC treated dH_2O	up to 50 µl (35 µl)	

PCR cycler settings

Step	Temperature	Time	
Initial denaturation	95°C	5 min	
Denaturation	95°C	15 sec	30 cycles
Annealing	53°C	30 sec	
Elongation	72°C	90 sec	
Final extension	72°C	10 min	
Hold	4°C	∞	

2.3 *Preparation of competent cells and transformation*

Electroporation allows the introduction of DNA, most commonly plasmid (but sometimes linear PCR product) into bacterial cells. The technique consists of two parts: the production of electrocompetent or chemically competent cells and the introduction of DNA into these cells. Transformation by electroporation has advantages over

traditional chemical methods (i.e. $CaCl_2$ treatment); electrocompetent cells are quicker to produce and have greater transformation efficiency. The procedure followed is based on that described previously (Sambrook and Russell 2000). The competent cells (especially *S*. Typhimurium 4/74) were produced freshly on the day of use.

2.3.1 Electroporation

Transformations of pBAD -based plasmid DNA into the *S*. Typhimurium 4/74 strains for the sRNA pulse-overexpression experiments (see chapter 6) were performed using electroporation.

1. The recipient strain was incubated for overnight at a suitable temperature in 5 ml LB (pH 7.0) with or without appropriate concentration of antibiotics.
2. The next day, the culture was diluted 1:1000 in the 25 ml of LB and grown at the appropriate temperature at 220 rpm. The volume of culture is dependent on the volume of electrocompetent cells needed as they are concentrated 100-fold (i.e. 25 ml culture produces 250 µl competent cells).
3. When the culture reached OD_{600} 0.3 to 0.4, it was transferred to 50 ml pre-cooled centrifuge tubes and left on ice for few min. Then spun down at 4000 rpm, 4°C for 10 minutes. From this point on the cells must be maintained on ice or at 4°C as increased temperatures decrease transformation efficiency.
4. Decanted supernatant and washed pellet three times in 25 ml ice-cold 5% (v/v) glycerol in dH_2O. After the last wash, spun-down and re-suspend the cells in 250 µl volume of ice-cold 10% glycerol.
5. Mixed 50 µl competent cells and 100 ng DNA (plasmid DNA or PCR product) together on ice and placed in a pre-cooled Electroporation cuvette (Bio-Rad, Cat. No. 165-2086). The mixture was ensured that mixed well.
6. The cuvette was placed in Electroporator (Bio-Rad) GenePulser II and a single pulse was given.

Electroporator settings

Voltage	2500 V
Capacity	25 µF
Resistance	∞ Ω
Second	≥ 4.8 milliseconds
Cuvette	2 mm

7. Immediately added 1 ml pre-warmed SOC medium to the cuvette, and mixed by pipetting. Incubated the culture statically at 37°C for one hour.
8. After 1h incubation, 100 µl of this mix was plated onto agar plates (with or without appropriate antibiotic) to select for transformed cells. The remaining 900 µl was left at room temperature overnight. If colonies absent on the plated, then the bench culture could be concentrated and plated.
9. A negative control of transformation was performed using water instead of DNA

2.3.2 *Chemical transformation*

All transformations into *E. coli* TOP10 cells during this study were done by heat shock. The examples are transformation of TOPO-, pP_L- and pXG10-sf-based plasmid DNA into the TOP10 cells in the context of the 5' RACE (see chapter 4) and GFP-two-plasmid validation (see chapter 6) experiments.

1. To generate a batch of chemically competent TOP10 cells, 25 ml LB medium supplemented with 50 mM $MgCl_2$ were inoculated 1:1,000 with TOP10 overnight culture and the cells were grown to an OD_{600} of 0.3 to 0.4 (approximately 3 to 3.5 h).
2. The cells were transferred to 50 ml centrifuge tube and left on ice for 10 min. Then the cells were centrifuged for 10 min at 4000 rpm at 4°C and the supernatant was discarded.
3. The pellet was washed two times with ice-cold 15 ml TB buffer using pre-cooled pipettes. After the second wash, 20 ml TB buffer was added left to incubate for 20 min on ice then centrifuged for 10 min at 4000 rpm, 4°C.
4. The supernatant was discarded and the pellet was gently resuspended in 250 µl TB buffer.
5. During transformation, 100 ng plasmid DNA was added to 50-100 µl cells and the suspension was incubated on ice for 10 min.
6. The cells were transferred to a water bath heated to 42°C for 1 min then incubated on ice for 5 min. 1 ml SOC or LB medium was added and the cells were left to recover at 37°C, shaking for 1 h.
7. A mock transformation using water instead of vector DNA was used as a negative control.

8. After 1h incubation, 100 µl of this mix was plated onto agar plates (with or without appropriate antibiotic) to select for transformed cells. The remaining 900 µl was left at room temperature overnight. If colonies absent on the plated, then the bench culture could be concentrated and plated.

TB buffer

HEPES	(Sigma, Cat. No. H3375)	10 mM
$CaCl_2$	(Sigma, Cat. No. C3306)	15 mM
$MnCl_2$	(Sigma, Cat. No. 529680)	55 mM
KCl	(Sigma, Cat. No. P9541)	250 mM

All components except $MnCl_2$ were mixed and adjusted the pH to 6.7 with 1 M KOH. Then the $MnCl_2$ was added and filter sterilized the mixture using a 0.22 µm filter.

2.4 *Construction of plasmids*

2.4.1 *pBAD constructs (inducible sRNA/asRNA over-expression)*

Inducible sRNA/asRNA over-expression plasmids for pulse-expression followed by microarray experiments are prepared according to Papenfort et al. (2006). Construction of pBAD along with the sRNA is a four step process. Initially a vector backbone and inserts were amplified then ligated and then transformed into recipient cell. The insertion of interested DNA fragment was checked by PCR amplification followed by sequencing.

Table 2.6: Primer used for pBAD-sRNA/asRNA construction is listed here.

Gene	Oligo name	Sequences
	pBAD_for	CAACTCTCTACTGTTTCTCC
	pBAD_rev	TCTGCGTTCTGATTTAATCT
474P1AS_50	CL_P1_50_for	ACTGACATTATTGATACTCCTT
474P1AS_50	CL_P1_50_rev	GTTTTTTCTAGAGTACTGGCTCAACCGCTGTA
474P1AS_90	CL_P1_90_for	GTGCGGAATCGCGGTTCC
474P1AS_90	CL_P1_90_rev	GTTTTTTCTAGAGCGAATGAATTCCTGAAATGG
474P1AS_110	CL_P1_110_for	ACTCATAAACCGGAATTGCT
474P1AS_110	CL_P1_110_rev	GTTTTTTCTAGATGCAACGATAAATTAAGCAACAA
474P1AS_180	CL_P1_180_for	ATAGCCTGATTGGTGGCTATA
474P1AS_180	CL_P1_180_rev	GTTTTTTCTAGAGGGAACTGGTTCTGATGAGG
STncP1-1	CL_P1-1_for	ACCTTTATCCAGTGAGCGCT
STncP1-1	CL_P1-1_rev	GTTTTTTCTAGAGGCAGGACACTCCGTAAGAA
	pBAD_con_for	ATGCCATAGCATTTTTATCC
	pBAD_con_rev	TTATCAGACCGCTTCTGC

pBAD backbone amplification: The pBAD/*Myc*-HisA backbone was PCR amplified using the primer pairs pBAD_for/pBAD_rev (Table 2.6). VELOCITY DNA polymerase (Bioline, Cat. No. BIO-21099) was used for PCR amplification to minimise the sequential errors.

PCR reaction mix

Components	Volume	Final concentration
5X Hi-Fi Reaction Buffer	10 µl	1X
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	1 µl	0.2 µM
10 µM Reverse Primer	1 µl	0.2 µM
pBAD (Template)	1 µl	100 ng
Velocity DNA Polymerase	1 µl	2 units
DEPC treated dH_2O	up to 50 µl (35 µl)	

PCR cycler settings

Step	Temperature	Time	
Initial denaturation	98°C	5 min	
Denaturation	98°C	15 sec	30 cycles
Annealing	57°C	30 sec	
Elongation	72°C	90 sec	
Final extension	72°C	10 min	
Hold	4°C	∞	

The amplified PCR fragments were treated with XbaI restriction enzyme using FastDigest® XbaI (Termo Scientific, Cat. No. FD0684) kit as specified by the manufacturer.

***XbaI* reaction mix**

Components	Volume	Final concentration
Uncleaned PCR reaction	44 µl	
10X FastDigest® buffer	5 µl	1X
FastDigest® *XbaI*	1 µl	

The reaction was incubated for 5 min at 37°C followed by thermal inactivation of XbaI for 5 min at 80°C. Followed by XbaI treatment, the PCR fragments were digested with

DpnI to remove template plasmid DNA using FastDigest® DpnI (Termo Scientific, Cat. No. FD1704). 1 µl DpnI was directly added to heat-inactivated reaction mixture and incubated for 5 min at 37°C. The reaction was stopped by thermal inactivation for 5 min at 80°C. Finally, the PCR products were cleaned using QIAquick® (QIAGEN, Cat. No. 28106) PCR Purification Kit following the manufacturer's instructions.

Insert amplification: Inserts of interested sRNA/asRNAs were generated by PCR amplification using genomic DNA as a template and fusion primers (reverse primers, Table 2.5). The fusion primer added the 5' overhangs to the PCR products that were complementary to the pBAD plasmid.

PCR reaction mix

Components	Volume	Final concentration
5X Hi-Fi Reaction Buffer	10 µl	1X
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	1 µl	0.2 µM
10 µM Reverse Primer	1 µl	0.2 µM
Genomic DNA (Template)	1 µl	100 ng
Velocity DNA Polymerase	1 µl	2 units
DEPC treated dH_2O	up to 50 µl (35 µl)	

PCR cycler settings

Step	Temperature	Time	
Initial denaturation	98°C	5 min	
Denaturation	98°C	15 sec	30 cycles
Annealing	55°C	30 sec	
Elongation	72°C	90 sec	
Final extension	72°C	10 min	
Hold	4°C	∞	

The PCR fragments were purified using QIAquick® (QIAGEN, Cat. No. 28106) PCR purification kit. 4 µl of purified insert was mixed with 1 µl of 5X DNA loading dye and loaded into 3% agarose gel (3 gm of agarose in 100 ml of TAE buffer). The gel was run at 100 V for appropriate time and stained with RedSafe™ (JH Science, Cat. No. 21141) for 45 min to 1 h. The fragment size of the inserts was confirmed with the help of

HyperLadder™ (under UV light), which was run in the gel along with the inserts. Finally, the fragments concentration was determined by NanoDrop (*Section 2.2.3*).

DNA loading dye (5X)

Bromophenol blue (0.25%)	(Sigma, Cat. No. B0126)	125 mg
Xylene cyanol (0.25%)	(Sigma, Cat. No. X4126)	125 mg
EDTA (0.1 M)	(Sigma, Cat. No. E7889)	25 ml
Glycerol (30%)	(Sigma, Cat. No. G5516)	15 ml
dH_2O		10 ml

TAE buffer: 40 mM Tris, 20 mM Acetic acid and 1 mM EDTA (pH 8.0)

T4 DNA ligation: The next step in the construction of a recombinant plasmid was connecting the insert DNA into a compatibly digested vector backbone. This reaction, called ligation was performed by the T4 DNA ligase (New England Biolabs, Cat. No. M0202) enzyme in a microcentrifuge tube on ice. Vector and insert were mixed in a molar ratio of 1:5 (vector: insert) for the indicated DNA size. The molar ratio was calculated using the online tool (link: http://www.insilico.uni-duesseldorf.de/ Lig_Input.html).

Ligation reaction mix

Components	Volume
10X T4 DNA Ligase buffer	2 µl
Vector (3.8 kb)	50 ng (1 molar ratio)
Insert (0.2 kb)	13.1 ng (5 molar ratio)
DECP treated H2O	up to 19 µl
T4 DNA ligase	1 µl (20,000 units)

The negative control reaction was performed with water instead of insert. Gently mixed the reaction by pipetting up and down and centrifuged briefly. The reaction mixture was incubated overnight at room temperature. Next day, the reaction was stopped by placing the microcentrifuge tube on ice.

***Transformation and validation of recombinant pBAD*:** Electrocompetent cells (*S*. Typhimurium 4/74) were made freshly and newly made recombinant plasmids (pBAD-insert) were electroporated. The cells were immediately transferred to SOC medium

and incubated at 37°C for one hour. After the incubation, 100 µl of culture was plated on the LB agar containing ampicillin. The remaining 900 µl was left overnight at room temperature, so that if colonies do not grow on the plated, this could be concentrated and plated if required.

The following day, 10 colonies from each insert were selected for PCR experiments to check the presence of insert in the pBAD. The primer pairs pBAD_con_for/pBAD_con_rev annealing next to the insertion site and genomic DNA of a transformant as a template used for PCR experiments.

PCR reaction mix

Components	**Volume**	**Final concentration**
5X Mango*Taq*™ Buffer (coloured)	2 µl	1X
10 mM dNTPs	0.4 µl	200 µM
10 µM Forward Primer	0.4 µl	0.2 µM
10 µM Reverse Primer	0.4 µl	0.2 µM
Genomic DNA (Template)	0.4 µl	100 ng
Mango*Taq*™ DNA Polymerase	1 µl	5 units
DEPC treated dH_2O	up to 20 µl (15.4 µl)	

PCR cycler settings

Step	**Temperature**	**Time**	
Initial denaturation	98°C	10 min	
Denaturation	98°C	15 sec	30 cycles
Annealing	52°C	30 sec	
Elongation	72°C	90 sec	
Final extension	72°C	10 min	
Hold	4°C	∞	

PCR products were purified using QIAquick® PCR purification column and sizes were identified by agarose gel electrophoresis using a 2% agarose gel (fragment size = insert size + 107 nt). The insert was PCR amplified and size was confirmed by sequencing. Finally, a glycerol stock of each culture was made and stored at -80°C until needed. The recombinant pBAD-sRNA/asRNA constructs made using this method is listed in Table 2.2.

2.4.2 *pP_L construction (constitutive sRNA expression)*

Constitutive sRNA expression plasmids were prepared by Sequence and Ligase Independent Cloning (SLIC) methods according to Li and Elledge (2007). The principle of SLIC cloning is explained in Figure 2.1. Construction of pP_L- sRNA is a five step process, which is similar to the pBAD construction (see *section 2.4.1*) with the variation in ligation step. Initially vector backbone and inserts were amplified, and treated with T4 DNA polymerase to make single strand ends. DNA ligation step was performed in the PCR machine. Finally, the recombinant plasmid was transformed to recipient cells and confirmed the insertion of interested DNA fragment by PCR amplification followed by sequencing.

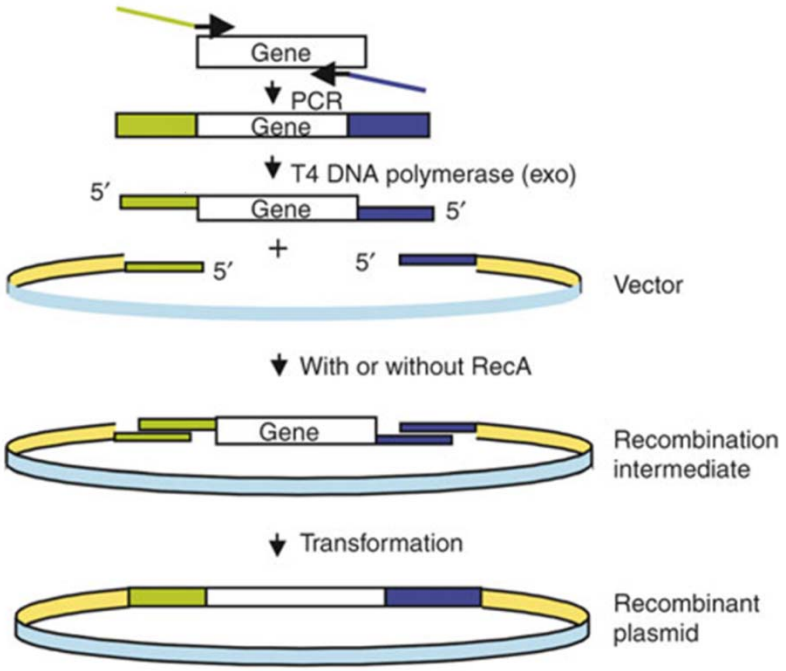


Figure 2.1: Principle of sequence and ligation independent cloning (SLIC) method.

The vector backbone is PCR amplified and *DpnI* digested. A sequence of interest is PCR-amplified with fusion primers generating a sequence overhang with the vector of interest (genomic DNA used as a template). T4 DNA polymerase is used to catalyse the reversal polymerization reaction in absence of dNTPs and single stranded DNA creating single-stranded, sticky ends in both, the insert and vector backbone. Annealing of T4 DNA polymerase treated insert and backbone generates a recombination intermediated which can be transformed into competent bacteria. Subsequently, selection pressure by antibiotics and the bacterial DNA repair mechanisms work together to generate transformants with correctly assembled plasmids. Figure from Li and Elledge (2007).

pP_L backbone amplification: The pP_L backbone was PCR-amplified with a proof-reading DNA polymerase using the primer pairs pP_L_for/pP_L_rev (Table 2.7). Purified pP_L plasmid used as a template for this amplification step. The PCR reaction mix and cycle settings were similar to "pBAD backbone construction" from *section 2.4.1*.

Table 2.7: Primers used for pP_L constitutive vector construction.

Gene	Oligo name	Sequences
	pP_L_for	TCTAGAGGCATCAAATAAAACGAAAGGC
	pP_L_rev	GTGCTCAGTATCTTGTTATCCGCTCAC
	pP_L_con_for	CGTATCACGAGGCCCTTTCGTC
	pP_L_con_rev	GCGGCGGATTTGTCCTACTCAG
STncP1-1	SL_P1-1_for	GTGAGCGGATAACAAGATACTGAGCACACCTTTATCCAGTGAGCGC
STncP1-1	SL_P1-1_rev	GCCTTTCGTTTTATTTGATGCCTCTAGAATCCGAGTGTATCTTTGC

The amplified PCR fragments were treated with DpnI using FastDigest® DpnI (Termo Scientific, Cat. No. FD1704) as specified by the manufacturer.

DpnI reaction mix

Components	Volume	Final concentration
Uncleaned PCR reaction	44 µl	
10X FastDigest® buffer	5 µl	1X
FastDigest® DpnI	1 µl	

The reaction was incubated for 5 min at 37°C followed by thermal inactivation for 5 min at 80°C. Finally, the PCR products were cleaned using QIAquick® (QIAGEN, Cat. No. 28106) PCR purification kit following the manufacturer's instructions.

Insert amplification: Inserts of sRNA (STncP1-1) was generated by PCR amplification using genomic DNA as a template and fusion primers (Table 2.6). The fusion primer added the 5' overhangs to the PCR products that were complementary to the pP_L plasmid. The PCR reaction mix and cycle settings were similar to "Insert amplification" from *Section 2.4.1*. The PCR products were purified using QIAquick® PCR purification kit.

T4 DNA polymerase treatment: Subsequently, inserts and vector backbone (pP_L) of the correct sizes were treated with T4 DNA polymerase to generate single-stranded DNA ends. The negative control reaction was performed with water instead of insert (STncP1-1).

T4 DNA reaction mix

Components	Volume	Final concentration
10X NEB Buffer 2	4 µl	1X
100 mM DTT	2 µl	5 mM
2 M Urea	4 µl	200 mM
BSA 100X (supplied by NEB)	0.4 µl	1X
PCR product (backbone or insert)	1 µg	25 ng/µl
T4 DNA Polymerase	1 µl	
dH_2O	up to 40 µl	

The reaction was incubated at 23°C for 20 min. 2 µl 500 mM EDTA was added to reaction mixture and incubated at 70°C for 20 min to stop the T4 DNA polymerase enzymatic reaction. Both vector and insert were purified using QIAquick® PCR purification kit and concentrations were determined by NanoDrop.

Annealing of backbone and insert: pP_L vector backbone and STncP1-1 insert with single-stranded 5' and 3' ends were mixed at a molar ration of 1:5 (vector: insert) and annealed using the touchdown protocol on a PCR machine (thermocycler). The vector and insert were annealed at 65°C for 10 min then decreased the temperature by 1°C with a 1 min hold for each degree (40 cycles).

PCR cycler settings:

Step	Temperature	Time
Initial denaturation	65°C	10 min
	65°C	1 min
Touchdown protocol	64°C	1 min
(decrease temperature by	63°C	1 min
1°C and hold for 1 min,	.	.
repeat for 40 cycles	.	.
between 65 to 25°C)	26°C	1 min
	25°C	∞

Transformation and validation of recombinant pP_L: Recipient chemically competent cells (*E. coli* TOP10 *recA*$^-$) were made freshly and newly made recombinant plasmid (pP_L-STncP1-1) was transformed (see *section 2.3.2*). The cells were immediately transferred to SOC medium and incubated at 37°C for I h. After the incubation, 100 µl of culture was plated on the LB agar containing ampicillin. The remaining 900 µl was left overnight at room temperature, so that if colonies do not grow on the plated, this could be concentrated and plated if required.

The validation of pP_L-STncP1-1 construct was same to the "Validation of recombinant pBAD vector" from *section 2.4.1*. Here, the control primer sets pP_L_con_for/pP_L_con_rev annealing next to the STncP1-1 insertion site used for PCR experiments. The size was identified by agarose gel electrophoresis using a 2.5% agarose gel (fragment size = insert size + 191 nt). A glycerol stock of positive clone was made and stored at -80^{o}C until needed. The recombinant pP_L-STncP1-1 construct generated using this method is listed in Table 2.2.

2.4.3 *pXG10-sf construction (constitutive mRNA::gfp fusion expression).*

The candidate target mRNA-*gfp* fusion plasmids were constructed by SLIC methods by Li and Elledge, 2007. The pXG10-sf recombinant plasmid construction is same to the construction of the pP_L plasmids. Primers used for pXG10-sf construction are listed in Table 2.8. Vector backbone was PCR amplified and DpnI digested. Insert contains the entire 5' untranslated region (5'UTR) and 45 to 60 bases of the coding sequence (CDS) were PCR amplified. Both vector and inserts were treated with T4 DNA polymerase and the inserts were inserted in front of GFP start codon using the touchdown protocol. Finally, the vector was transformed into chemical competent TOP10 cells. After PCR amplification the size of the insert was identified using a 1% agarose gel (fragment size = insert size + 1031 nt). Finally, the expression of GFP protein was determined using colony fluorescence experiment (*Section 2.7.1*) and stored at -80°C. The recombinant pXG10-sf-gene constructs (pXG10-sf-*gtrA* and pXG10-sf-*galP*) generated using this method are listed in Table 2.2.

Table 2.8: Primers used for pXG10-sf constitutive vector construction.

Gene	Oligo name	Sequences
	pXG10_for	GCTAGCGGATCCGCTGGCTCCGCTGC
	pXG10_rev	ATGCATGTGCTCAGTATCTCTATCAC
	pXG10_con_for	GCCGTAATATCCAGCTGAAC
	pXG10_con_rev	CTCATGAATTCGCCAGAACC
gtrAa	SL_gtrA_for	GTGATAGAGATACTGAGCACATGCATCTTTCCGAATCCGCTGATTT
gtrAa	SL_gtrA_rev	GCAGCGGAGCCAGCGGATCCGCTAGCATGCGTATGCATCCCATACA
galP	SL_galP_for	GTGATAGAGATACTGAGCACATGCATATTCAGTTTCAGTCATCATGAAA
galP	SL_galP_rev	GCAGCGGAGCCAGCGGATCCGCTAGCAACGCCGATATCCAGGCCA

2.5 *STncP1-1 deletion in S. Typhimurium 4/74*

An in frame deletion mutant of STncP1-1 (4/74-ΔP1-1) in *S*. Typhimurium 4/74 was prepared as described by Datsenko and Wanner (2000). Figure 2.2 explains the principle of making deletion mutant experiment. Removal of STncP1-1 sRNA from $pSLT^{4/74}$ is divided into four steps (see below).

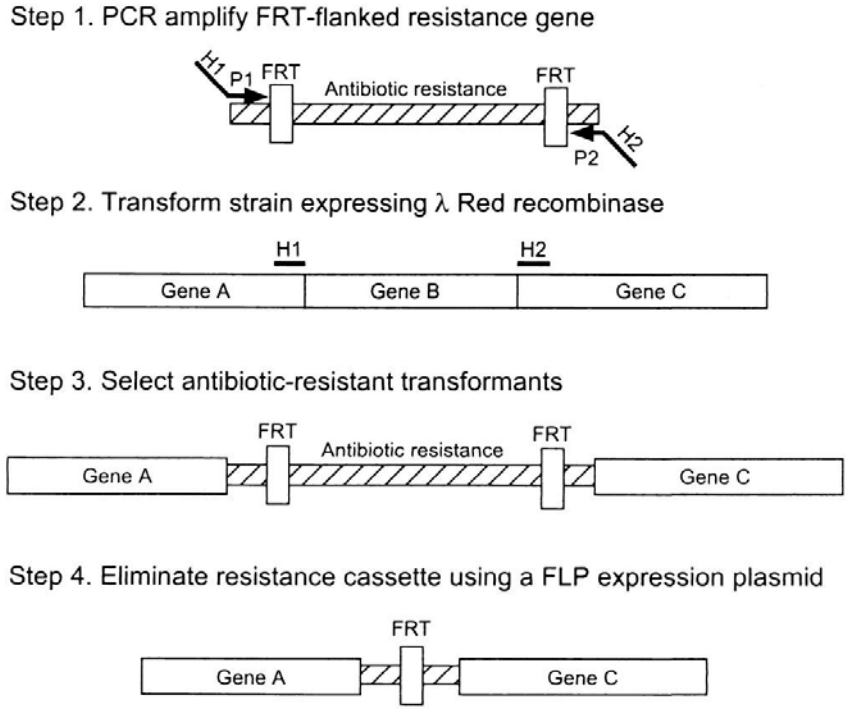


Figure 2.2: Principles of bacterial gene deletion method.

The strain from which the gene has to be knocked out is first transformed with pKD46 which encodes a specific recombinase. Meanwhile, the kanamycin cassette from pKD4 is amplified using the primers (P1/P2). Now, the insert contains the kanamycin cassette flanked by FRT sites (FLP recognition target). Then the insert is electroporated into the strain containing pK46 plasmid. Once the PCR fragment is transformed to the cell, the expression of recombinase (from pKD46 plasmid) is induced by adding arabinose. The recombinase helps to perform recombination between the targeted gene and the kanamycin (insert) cassette. This happens because of the homologus sequences present upstream and downstream of the kanamycin (insert) cassette. Finally the kanamycin cassette is cured using pCP20 plasmid. Figure from Datsenko and Wanner (2000).

2.5.1 *Amplification of the FRT-flanked Kan^R-cassette*

The forward deletion primer (P1-1_del_for, Table 2.9) containing of a stretch of 50 bases taken from downstream to the start site of STncP1-1 gene (5'-end of the deletion primer) H1 attached with 20 bases matching to pKD4 (3'-end of the primer) P1 was designed. Similarly, the reverse primer (P1-1_del_rev) containing of a stretch of 50 bases taken from upstream to the STncP1-1 gene end (5'-end of the deletion primer) H2 attached with 20 bases matching to pKD4 (3'-end of the primer) P2 was designed. This deletion primer pairs were used to PCR-amplify the FLP recognition target (FRT) site-flanked kanamycin resistance cassette of pKD4. Here, the plasmid pKD4 was used as a DNA template. The resulting PCR-fragment was 1576 bp in length.

Table 2.9: STncP1-1 deletion and test primers.

Gene	Oligo name	Sequences
STncP1-1	P1-1_del_for	AGTAAGCGACAGGATATTTGCTCCTTTTTCCTGATT GCATCCGGTTTTGCGTGTAGGCTGGAGCTGCTTC
STncP1-1	P1-1_del_rev	GTGTATCTTTGCCTTCAGAGGCTTTGGGCAGGACA CTCCGTAAGAAGATGCATATGAATATCCTCCTTA
STncP1-1	P1-1_tes_for	ACAAATCTTAACTAAATCTTAAGCGA
STncP1-1	P1-1_tes_rev	TTATGTAATGCTTCGGGCTGT

PCR reaction mix

Components	Volume	Final concentration
5X Hi-Fi Reaction Buffer	10 µl	1X
10 mM dNTPs	1 µl	200 µM
10 µM Forward Primer	1 µl	0.2 µM
10 µM Reverse Primer	1 µl	0.2 µM
pDK4 (Template)	1 µl	100 ng
Velocity DNA Polymerase	1 µl	2 units
DEPC treated dH_2O	up to 50 µl (35 µl)	

PCR cycler settings

Step	Temperature	Time	
Initial denaturation	98°C	5 min	
Denaturation	98°C	15 sec	30 cycles
Annealing	60°C	30 sec	
Elongation	72°C	90 sec	
Final extension	72°C	15 min	
Hold	4°C	∞	

The PCR products were purified using QIAquick® PCR purification kit.

2.5.2 *Transformation of pKD46 into wild-type 4/74*

Electrocompetent cells of *S*. Typhimurium were made freshly and appropriate amount of pKD46 plasmid was electroporated (*See section 2.3.1*). After the electric pulse the culture was immediately transferred to 1 ml SOC medium and incubated at 30°C for 1 h for recovery. Then 100 µl of this mix was plated onto LB agar plates containing ampicillin (100 µg/ml) and incubated overnight at 30°C. The remaining 900 µl was left overnight at room temperature, so that if colonies do not grow on the plated, this could be concentrated and plated if required. Next day, the presence of colonies on LB agar containing ampicillin was observed.

Plasmid pKD46 (see Table 2.2) contains an ampicillin cassette as selection marker, a L-arabinose inducible recombinase, and features temperature sensitive replication via *repA101ts* that allows for plasmid maintenance at 30°C and plasmid curing at 37 – 42°C. The *repA101ts* gene encodes the replication protein (RepA101T) is rapidly inactivated at temperature of 37 – 42°C and is therefore not able to mediate replication initiation at elevated temperature.

2.5.3 *Transformation of Kan^R-cassette into wild-type containing pKD46*

Overnight culture of *S*. Typhimurium 4/74 containing pKD46 was inoculated to 25 ml of LB medium supplemented with 0.2 % L-arabinose and 100 µg/ml ampicillin, and incubated at 30°C until the culture reached OD_{600} of 0.3 - 0.4. Electrocompetent cells were generated using the 10% glycerol. Different concentrations of inserts (50 ng, 100 ng and 250 ng) were transformed into competent cells using electric pulse as described in *section 2.3.1*. The culture was immediately transferred to 1 ml SOC medium and incubated at 37°C for 1 h for recovery. Then 100 µl of culture was plated onto both LB agar containing ampicillin (100 µg/ml) and LB agar containing kanamycin (50 µg/ml), and incubated overnight at 37°C. The remaining 800 µl was left on bench at room temperature. If the culture was failed to grow on kanamycin containing plated, the 800 µl bench culture was spun and streaked on the fresh kanamycin containing plates, and incubated overnight.

The next day colonies from both kanamycin plates were re-streaked and incubated overnight at 37°C. Then the colonies from kanamycin plate were lysed and a control PCR was performed to verify recombination of the kanamycin resistance cassette at

the anticipated genomic location. Test primers pairs (Table 2.8) and the kanamycin inserted mutant genomic DNA (as a template) used for this PCR-based verification. Test primers were designed to align more than 20 – 40 nt upstream and downstream of the deletion site. The size if the inserted fragment was determined using 1% agarose gel. Here the wild-type genomic DNA used as a template for negative control.

PCR reaction mix

Components	Volume	Final concentration
5X Mango*Taq*™ Buffer (coloured)	2 µl	1X
10 mM dNTPs	0.4 µl	200 µM
10 µM Forward Primer	0.4 µl	0.2 µM
10 µM Reverse Primer	0.4 µl	0.2 µM
Genomic DNA (Template)	0.4 µl	100 ng
Mango*Taq*™ DNA Polymerase	1 µl	5 units
DEPC treated dH_2O	up to 20 µl (15.4 µl)	

Mango*Taq*™ DNA polymerase kit (Bioline, Cat. No. 21078).

PCR cycler settings

Step	Temperature	Time	
Initial denaturation	95°C	10 min	
Denaturation	95°C	15 sec	30 cycles
Annealing	57°C	30 sec	
Elongation	72°C	2 min	
Final extension	72°C	10 min	
Hold	4°C	∞	

2.5.4 *Transduction of Kan^R-cassette into pKD46 free background*

To avoid the recombinase-mediated recipient genomic DNA damage (recombination in the location other than the target site), the P22 transduction experiment was performed. P22 is a generalised transducing phage, specific to *Salmonella* Typhimurium, useful for moving selectable genetic markers or plasmids from one strain to another. This protocol followed that of Gemski and Stocker (1967) and was carried out as detailed previously (Maloy et al. 1996). We use a high transducing derivative, P22 HT105/1 *int-4* was used here, which is deficient (*int*) form of P22.

Preparation of transducing-Lysate: 5 ml culture of donor cells (4/74-Kan^R mutant) was incubated for overnight in LB medium containing 50 µg/ml of kanamycin. Then 25 µl of overnight culture was inoculated into 25 ml LB medium (1:1,000-fold dilution) and grown at 37°C, 220 rpm, until an OD_{600} of 0.15-0.2 (if grown too far the phage yield will be low). 250 µl of stock P22 (stock should contain at least 10^{10} phage/ml raised on a strain of *Salmonella* lacking any selectable markers) added into 25 ml of culture and incubate for at least 6 hours (usually overnight) during which time lysis occurs. Next day morning, 500 µl of chloroform was added to lyse any remaining cells. The mixture was mixed gently by inverting the centrifuge tube and left the mixture at 4°C for at least 10 min (do not vortex as this reduces titre). Then the mixture was centrifuged for 10 minutes at 6000 rpm, 4°C in centrifuge to remove cell debris. The mixture was Filter sterilised using a syringe-filter (0.22 µm). Finally, 500 µl of chloroform was added to the lysates and mixed together before storage at 4°C. Lysate can successfully be used for transduction for over 1 year.

P22 Transduction: Similar to donor cells, the recipient cells (4/74 wild-type) was grown in 5 ml culture for overnight in LB medium without any antibiotics. Then 200 µl of recipient cells were mixed with 10 µl P22 lysate (prepared from strain containing the selectable marker we wish to transduce) in a sterile 1.5 ml Eppendorf tube and incubate at 37°C for 60 minutes. This allowed expression of the acquired antibiotic resistance genes in transductants. Plated out all 210 µl onto LB agar plates containing kanamycin (50 µg/ml) antibiotic to select for transductants and incubated overnight at 37°C. At the same time the negative control was carried out by spreading 100 µl recipient cells (with no phage) and 50 µl P22 lysate (with no cells) on separate plates to ensure that neither of these on their own will yield antibiotic resistant colonies. The transductants (colonies) from kanamycin plated were picked and purified immediately after overnight growth to minimise the formation of lysogens. The purification step was involved in re-streaking of kanamycin positive transductants onto fresh kanamycin containing LB agar supplemented with 10 mM EGTA.

To verify that P22 phage was eliminated from the mutant strain, the mutant strain was plated on green agar plated (see solid medium in *section 2.1.2*). Green plates are poorly buffered rich medium with an excess amount of glucose and a mixture of pH indicator dyes. The mutant transductants were re-streaked twice on green indicator plates to visualise unstable P22 pseudo-lysogens. Cells that are pseudo-lysogens

appear dark green on these plates, cells that are P22-free are light green or white, and were used for subsequent experiments. As P22 H105/1 *int-4* cannot form true lysogenic strains that for large, light green or white colonies were considered to be phage free. Control PCR using the same test primers as before was performed to verify that the transduction procedure had integrated the kanamycin resistance cassette at the correct genomic (plasmid) location. The resulting PCR product was purified and size was determined using the 1% agarose gel.

2.5.5 *Elimination of Kan^R-cassette using pCP20*

To avoid the polar effects in microarray and RNA-Seq approach-based STncP1-1 target identification experiments, the kanamycin cassette was excised from the mutant 4/74 (plasmid) strain. To excise the antibiotic resistance cassette from the target location in the genome, plasmid pCP20 was transformed into freshly made electrocompetent cells (*section 2.3.1*) of the deletion strain with kanamycin resistance cassette. Plasmid pCP20 (see Table 2.2) contains a chloramphenicol resistance cassette as selection marker and encodes yeast FLP-recombinase. The feature of temperature sensitive replication via *repA101ts* is similar to pKD46 replication at 30°C Here also the plasmid is maintained at 30°C and cured at 37 – 42^{o}C. The FLP-recombinase recognises the FRT sites adjacent to the kanamycin cassette copied from plasmid pKD4 and catalyses excision of the antibiotic, leaving the single FRT site 34 bp length as a scar.

1 ml of SOC medium was added to the transformation mixture and the culture was incubated shaking at 200 rpm for one hour at 30°C for recovery. Then, 100 µl of reaction mixture was plated onto LB agar containing chloramphenicol to select for transformed cells. The remaining 900 µl was left at room temperature overnight, so that if colonies do not grow on the plated, this could be concentrated and plated if required. The plated were incubated overnight at 30°C, without shaking. Colonies appeared on the LB agar were re-streaked twice and incubated at 30°C for overnight. Then the temperature was increased from 30 to 42°C to remove the pCP20 from cells. The re-streaking procedure was repeated twice and incubated at 42°C. Here, the third control PCR was performed using the same test primers and genomic DNA of kanamycin excised strain as a template. The wild-type genomic DNA used as a control the check the fragment size with kanamycin excised strain. Finally, a glycerol stock of resulting strains from each steps were made and stored at -80°C until needed.

2.6 *sRNA target identification analysis*

To identify the mRNA targets of plasmid-encoded sRNA/asRNAs, the pulse-expression for sRNA followed by Microarray experiments were performed as descried in Papenfort et al. 2006 with minor modifications. Figure 2.3 shows the principles of pulse-overexpression experiments.

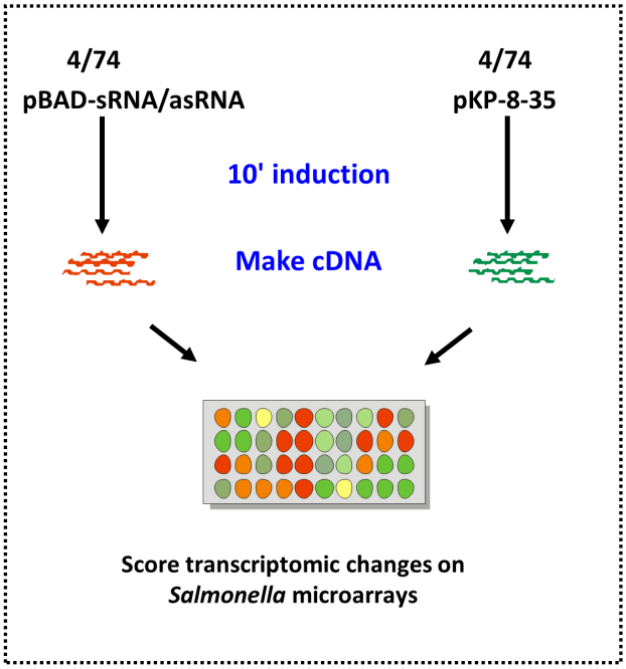


Figure 2.3: Principle of sRNA pulse-expression followed by microarray experiments. See text for details. Figure was adapted from Papenfort et al. 2006 with some modifications.

2.6.1 *Pulse-expression experiments*

In case where the sRNA located antisense to coding gene (asRNA), it could not be deleted. Because, the deletion of asRNA could affect the expression of sense coding gene. So, the pBAD constructs were made (pBAD-asRNA, e.g., pBAD-474P1AS_50) and electroporated into *S.* Typhimurium 4/74 wild-type background. Simultaneously, the pKP-8-35 also transformed into 4/74 wild-type background, which is used as a control for pBAD-asRNA experiments. Here, by comparing the expression level of asRNA across the 22 environmental conditions, the growth condition for asRNA pulse-expression was chosen where the asRNA was known to be low (expression levels), to avoid target coding gene silencing from plasmid asRNA expression (native) to interfere with the pulse-expression experiments.

In contrast, where the sRNA was encoded in an intergenic region (IGR) and could be deleted from the plasmid without disrupting other transcripts. So, the sRNA deletion mutant strain (4/74-ΔSTncP1-1) was generated and used as genetic backbone for transient pulse-expression experiments. The pBAD-STncP1-1 construct was generated and transformed into 4/74-ΔSTncP1-1 mutant background. Simultaneously, the pKP-8-35 also transformed into same mutant background. As transcriptomic data was available for 4/74 grown under 22 environmental conditions, the growth condition under which the pulse-expression experiment was performed could be chosen after evaluation of the known sRNA expression profiles.

Two growth conditions were chosen for each sRNA target identification and two biological replicated were generated for each growth conditions. Both strains, 4/74-pBAD-sRNA and 4/74-pKP-8-35, were grown overnight in 5 ml of LB medium at 37°C. Next day, the cells were washed twice with the chosen grown medium and 25 µl of inoculum was transferred to 250 ml flask containing 25 ml of growth medium (1:1,000 dilution ratios). Flasks were placed on the pre-warmed water both (to appropriate temperature) and incubated. Once the OD_{600} was reached at the desired time-point, 20% of L-arabinose was added to yield a final concentration of 0.2% L-arabinose in the culture. The cultures were incubated for another 10 min before adding 0.25 volume of stop solution (phenol/ ethanol in 5:95 % ratio). Total RNA was then extracted using TRIzol® (See *Section 2.2.2*). Finally, the differences in the genes at their expression levels were determined using the microarray experiments. The targets of sRNA/asRNAs investigated in this study are listed in Table 2.10.

Table 2.10: List of sRNA/asRNAs investigated by pulse-expression and microarray.

Clones	Control	Growth conditions
4/74 wild-type pBAD-474P1AS_50	4/74 wild-type pKP-8-35	LowMg^{2+}, OD_{600} 0.3
4/74 wild-type pBAD-474P1AS_90	4/74 wild-type pKP-8-35	LowMg^{2+}, OD_{600} 0.3
4/74 wild-type pBAD-474P1AS_110	4/74 wild-type pKP-8-35	LowMg^{2+}, OD_{600} 0.3
4/74 wild-type pBAD-474P1AS _180	4/74 wild-type pKP-8-35	LowMg^{2+}, OD_{600} 0.3
4/74-ΔSTncP1-1 pBAD-STncP1-1	4/74-ΔSTncP1-1 pKP-8-35	LowMg^{2+}, OD_{600} 0.3
4/74 wild-type pBAD-474P1AS_50	4/74 wild-type pKP-8-35	LB-Lennox, 25^{o}C, OD_{600} 0.3
4/74 wild-type pBAD-474P1AS_90	4/74 wild-type pKP-8-35	LB-Lennox, ESP, OD_{600} 2.0
4/74 wild-type pBAD-474P1AS_110	4/74 wild-type pKP-8-35	LB-Lennox, ESP, OD_{600} 2.0
4/74 wild-type pBAD-474P1AS _180	4/74 wild-type pKP-8-35	LB-Lennox, 25^{o}C, OD_{600} 0.3
4/74-ΔSTncP1-1 pBAD-STncP1-1	4/74-ΔSTncP1-1 pKP-8-35	LB-Lennox, TEMP10, OD_{600} 0.3

2.6.2 *Microarray*

DNA microarrays allow the simultaneous measurement of the global gene expression under defined experimental conditions. Total RNA samples were reversed transcribed to cDNA and labelled with the Cy3 fluorophore whereas the reference genomic DNA was labelled with the Cy5 fluorophore as described on the IFR website (link: http://www.ifr.ac.uk/safety/microarrays/#protocols). The SALSIFY microarrays used in this thesis for sRNA target hunting experiments (Chapter 6) were designed by Dr. Karsten Hokamp, Dept. Gentics, Trinity College Dublin, Ireland. The arrays include 60mer oligonucleotide probes of 4715 coding genes from SL1344 (including 217 plasmid-encoded coding genes) with, on average two distict probes per gene. It also contains probes for 141 published sRNAs (Kröger et al. 2012).

Labelling of RNA: Total RNA extracted from different transcriptomic samples were digested with DNase I enzyme. The labelled cDNA was synthesised from total RNA

directly through a reverse transcription reaction, which reduces the bias in the data, e.g., bias introduced by PCR amplification. The first step involved the reverse transcription of 10 µg extracted RNA into corresponding cDNA. To the RNA, 5 µg (1.7 µl, Invitrogen, Cat. No. 48190-011) random hexamer primers were added in a total volume of 9.4 µl (made up with DEPC treated H_2O). The RNA was denatured by incubation at 70°C for 5 min and then chilled on ice for 10 min to allow primers annealing. The primed template was then added to the reverse transcription reaction (AffinityScript multi-temperature Reverse Transcriptase kit, Stratagen, Cat. No. 6000109).

Reverse transcription and labelling reaction

Components	Volume	Final concentration
10X Reverse Transcription buffer	2 µl	1X
DTT (0.1 M)	2 µl	10 mM
Primed total RNA sample	9.4 µl	
dNTPs (25 mM dATP,dGTP,dTTP; 10 mM dCTP)	0.6 µl	0.75 mM dATP, dGTP, dTTP; 0.3 mM dCTP
Cy3 dCTP (1 mM)	2 µl	0.1 mM
AffinityScript RT	4 µl	

The reaction was incubated protected from light at 25°C for 10 min, shaking, to mix. Subsequently, incubated overnight at 42°C in a thermolcycler PCR block. The reverse transcription reaction was stopped by the addition of 1.5 µl 20 mM EDTA (pH 8.0) and any remaining RNA was hydrolysed by the addition of 15 µl 0.1 M sodium hydroxide and heated the sample at 70°C for 10 min. the alkali was neutralised by adding 15 µl 0.1 M HCl. The cDNA was purified using a MiniElute (QIAGEN, Cat. No. 28206) PCR purification kit with a final elution volume of 16 µl DEPC treated water. Concentration and dye incorporation of the sample cDNA were measured by NanoDrop. If the samples signal at 570 nm was more than 5 fold higher than the background signal at 570 nm (caused by the absorption of the Cy5-labbled control cDNA) and the dye incorporation was between 8 and 20 pmol Cy3/µg (manufacturer's guidelines), the samples were stored at -20°C protected from light until hybridisation to a single microarray.

Labelling of genomic DNA: A labelled-cDNA library of 4/74 genomic DNA was performed using the BioPrime® DNA Labelling System (Invitrogen, Cat. No. 18094-

011). The reaction involved the incorporation of a fluorophore into the genomic DNA using random primers (octamers), dNTPs spiked with Cy5-labelled cytidine triphosphate and Klenow polymerase. To 21 µl (2 µg) genomic DNA, 20 µl of 2.5X random primers (octamers)/reaction buffer mix from the BioPrime DNA labelling kit was added and the primers were denatured by boiling at 100°C for 5 min. The samples were placed on ice for at least 5 min for primer annealing to occur. The primer templates were added to the labelled-cDNA synthesis reaction.

Amplification and labelling reaction:

Components	Volume	Final concentration
Primed genomic DNA sample	41 µl	2 µg
dNTPs (1.2 mM dATP,dGTP, dTTP; 0.6 mM dCTP)	5 µl	0.12mMdATP,dGTP,dTTP; 0.06 mM dCTP
Cy5 dCTP (1 mM)	3 µl	0.06 mM
Klenow polymerase	1 µl	40 units

The samples were spun briefly and incubated overnight at 37°C, in the dark to prevent dye bleaching. The labelling reaction was stopped by the adding 5 µl 0.5 M EDTA (pH 8.0) and the DNA was cleaned using a QIAquick PCR purification kit, eluted with 41 µl DEPC treated water. Concentration and dye incorporation of the control cDNA were measured by NanoDrop. If the signal at 670 nm was more than 5 fold higher than the background signal at 670 nm (caused by the absorption of the Cy3-lebbled control cDNA) and the dye incorporation was between 8 and 20 pmol Cy5/µg (manufacturer's guidelines), the samples were stored at -20°C protected from light till hybridisation. The amount of control cDNA library generated in the above protocol was sufficient for 8 microarrays which equals one microarray slide.

Hybridisation, washing and measurement: Eight individual hybridisation reactions were set up for the eight arrays per slide. 5 µl Cy5-labelled control cDNA generated from genomic DNA were added to each tube containing 15 µl Cy3-labelled sample cDNA synthesised from RNA samples.

Hybridization reaction

Components	Volume
Cy3-labelled sample cDNA	15 µl
Cy5-labelled control cDNA	5 µl
Blocking solution (Agilent)	5 µl

2X Hybridization buffer (Agilent)	25 µl

The reaction mixture was heated at 95°C in the dark for 2 min to denature the cDNA. Then reaction mixture was centrifuged at 14,000 rpm for 1 min, room temperature. A gasket slide was placed in the hybridisation chamber. The eight wells on the gasket slide were filled with one hybridisation reaction mixture and respective loading positions were recorded. The array slide was placed on the top of the gasket slide with the arrays facing down. The hybridisation chamber was closed tightly and incubated at 65°C overnight in the hybridisation oven.

After the overnight incubation, the hybridisation chamber was opened and the gasket slide was separated from the array slide in a bath of washing buffer 1. Subsequently, the slide was transferred into a bath containing washing buffer I pre-warmed to 37°C and washed for 10 min in the dark. The washing buffer was agitated using a magnetic stirring bar. Then, the slide was transferred to the bath containing washing buffer II at room temperature and washed twice for 5 min in wash buffer II. The washing was performed in the dark. Finally the slide was air-dried and stored at room temperature protected from the light till the measurement was taken. Array slides were scanned on the High-Resolution C Scanner (Agilent), and the data was transferred and analysed using the software tool GeneSpring 7.3 (Agilent).

Wash buffer I

900 mM NaCl, 90 mM Na acetate and 0.05% Triton (pH 7.0)

Wash buffer II

15 mM NaCl, 1.5 mM Na acetate and 0.05% Triton (pH 7.0)

2.7 *Two-plasmid assay for sRNA-target validation*

Two-plasmid assay was performed to validate the putative mRNA targets of plasmid-encoded sRNA STncP1-1 (chapter 6). This was performed in TOP10 cells (*E. coli recA*$^-$) according to protocol from Urban and Vogel (2007) and Corcoran et al. (2012). The pP_L-based STncP1-1 transcriptional (constitutive) sRNA-expression vector and pXG10-sf-based putative target (e.g., *gtrA*) translational vector fusions were generated using the SLIC cloning protocols (*section 2.4.2*). In the *gtrA::gfp* fusion of the pXG10-sf vector, a stretch of DNA sequence taken from *gtrA*, was inserted in front of *gfp* gene.

The DNA sequence stretch contains the entire 5' UTR and few nucleotides of the coding sequence (33 aa) of the corresponding putative target gene. *E. coli* Top10 chemical competent cells were freshly made and the pLP-STncP1-1 and pXG10-sf-*gtrA::gfp* vector was co-transformed. As a control, same pXG10-sf-*gtrA::gfp* fusion vector and pP_L-based constitutive nonsense-RNA-expression (pJV300) vector were co-transformed to TOP cells. GFP measurement was taken in both test and control samples using the colony fluorescence imaging (CFI) and flow cytometric (FC) analysis. The CFI and FC experiments were done by the methods explained in Sharma et al. (2011). The intensity of GFP fluorescence between test and control was compared.

Principle: The entire sequence of sRNA is inserted in pP_L constitutive expression vectors. The 5' UTR along with few nucleotides of the coding sequence of the target mRNA is inserted in front of *gfp* gene (translational fusion) in pXG10-sf constitutive expression vector. These two vectors are co-transformed into TOP10 cells (test - TOP10_pP_L_**STncP1-1**_pXG10-sf-***gtrA****::gfp*). Control samples contain pJV300 vector instead of pP_L-STncP1-1 (control - TOP10_pJV300_pXG10-sf-***gtrA****::gfp*). Plasmid pJV300 is also a pP_L-based constitutive vector but encodes nonsense RNA, which is unrelated to STncP1-1 targets. The intensity of GFP fluorescence between test and control was compared. If the sRNA binds with the 5' region (5' UTR or 5' coding region) of the target mRNA, which results the degradation of target mRNA (*gfp* mRNA also). So, the reduced GFP fluorescence level is observed in test compared to control sample.

Limitation: Two-plasmid assay was designed to validate the negative regulation of the sRNAs (Sharma et al. 2011 and Corcoran et al. 2012).

2.7.1 *Colony fluorescence imaging*

The double transformants (TOP10_pP_L_sRNA_pXG10-sf-*gtrA::gfp* and TOP10_pJV300_pXG10-sf-*gtrA::gfp*) were streaked in a single plate contains LB agar supplemented with both ampicillin (pP_L) and chloramphenicol (pXG10-sf). The plates were incubated overnight at 37°C. Next day, the fluorescence levels were detected using ImageQuantTM LAS4000 (GE, USA) (filter: 510DF10). The GFP fluorescence intensity of the test sample was compared with control.

2.7.2 *Flow cytometric analysis*

TOP10 cells carrying *pXG10-sf* with *gene::gfp* fusion and either a control plasmid (pJV300*)* or constitutive sRNA expression plasmid *(*pP_L) were grown in parallel to stationary phase in Lennox broth containing both ampicillin (100 µg/ml) and chloramphenicol (25 µg/ml) at 37^oC. Cells from 100 µl of culture was harvested, washed twice in PBS and resuspended in 1 ml of PBS with 4% formaldehyde. Measurement took place in a BD FACSCanto II Flow Cytometer (excitation at 488 nm, emission filter 530/30 nm). Forward angle light scatter (FSC), side scatter (SSC), and the fluorescence of the cells was recorded. Data analysis was carried out using FlowJo 7.6.4 (Tree Star, Inc.)

PBS buffer

137 mM NaCl, 2.7 mM KCl and 10 mM Na_2HPO_4/KH_2PO_4 (pH 7.4)

2.8 *Sample preparation for RNA-Sequencing*

To identify putative mRNA target of $pSLT^{4/74}$-encoded regulatory small RNA STncP1-1 using RNA-Seq approach, the RNA samples from *S*. Typhimurium 4/74 wild-type and STncP1-1 deletion mutant strains were prepared and sent for sequencing (chapter 6). The growth condition was chosen by comparing the expression of STncP1-1 across 22 environmental conditions. STncP1-1 sRNA was highly expression in $LowMg^{2+}$, salt shock (NaCl shock) and InSPI2 condition, relatively (Figure 6.3). The STncP1-1 target hunting using pule-expression analysis was done using the $LowMg^{2+}$ stress condition. In $LowMg^{2+}$, the bacterium was grown in PCN medium with 10 µM Mg^{2+} (low concentration of Mg) until OD_{600} reaches 0.3. The same $LowMg^{2+}$ stress condition was chosen here in RNA-Seq based mRNA target hunting, so that the results from both experiments can be compared. But, very poor quality of total RNA was yielded due to high Magnesium stress in $LowMg^{2+}$. So, the InSPI2 growth condition was chosen, which is PCN medium with 1 mM Mg^{2+} (no Magnesium stress).

S. Typhimurium 4/74 wild-type and STncP1-1 deletion mutant were grown in parallel (no replicates) under SPI2-inducing conditions (PCN, OD_{600} 0.3; Löber et al. 2006). The total RNA was extracted using the TRIzol® (Invitrogen) method (*section 2.2.2*). Then the samples were treated with DNase I to remove unwanted DNA in the samples (*section 2.2.5*). The purity of the RNA was determined using RNA Bioanalyser (*section*

2.2.4) and sent for sequencing. Sequencing was carried out using Illumina HiSeq 2000 system by Vertis Biotechnologie AG (Ferising, Germany).

2.9 *Prediction of Sigma70 promoter consensus motifs*

To identify Sigma70 promoter consensus motifs, an unbiased motif-search was performed using algorithm MEME 4.4.0 (Bailey et al. 2006) as described by Kröger et al. (2012). The downstream DNA sequences (up to 50 nt) of the chosen TSS were extracted from the plasmids of S. Typhimurium 4/74 using the custom Perl scripts. σ^{70} (RpoD) binding sites are composed of two discrete blocks (-10 and -35) with variable spacing between blocks. So, the sequences were divided into two segments ranging from -2 to -35 (for the -10 motif) and from -25 to -50 (for the -35 motif) relative to TSS and these were searched using MEME with the following parameters: Motifs could range in size from 4 to 10 bp, each DNA sequence was expected to contain one sequence matching the motif, and only the forward DNA strand was analysed. The chromosomal σ^{70} binding motifs were adapted from Kröger et al. (2012). Sequence logos were generated using WebLogo (Crooks et al. 2004), and base positions within the motifs are number using the co-ordinate system of Shultzaberger et al. (2007).

2.10 *Identification of chromosomal-like transcriptional signatures in plasmids*

The use of similarity metrics to identify classes of co-expressed genes from a transcriptomic compendium of is well explained (Kröger et al. 2013). The Pearson correlation was used to identify genes that showed a similar expression profile over the 23 environmental conditions using GeneSpring GX7.3 (Agilent Technologies, Inc.). The representative genes from each SPI of *Salmonella* (e.g., *ssaG* for SPI-1 and *prgH* for SPI-2) were selected to identify plasmid genes that shared SPI-like patterns. The chromosomal genes including *rpsM* and *entA* were also selected to find similar patterns in plasmids. The plasmid-encoded genes that shared a correlation coefficient of > 0.7 with representative gene, was considered as the similar expression patterns.

CHAPTER 3

DEVELOPMENT OF A DATA ANALYSIS PIPELINE FOR RNA-Seq LIBRARIES

3.1 *Introduction*

The transcriptome is defined as the total set of transcripts, mRNAs and non-coding RNAs in a population of cells grown under specific conditions. Transcriptomic analysis provides the important information that is needed to understand gene structure and function (Sorek and Cossart 2010; McGettigan 2013). Various technologies have been developed to deduce and quantify the transcriptome of a cell, including hybridization and tag-based approaches. Hybridization-based approaches typically involve incubating fluorescently labelled cDNA with custom-made microarrays or commercial high-density oligo microarrays. In the past decade, microarrays were the most widely used methodology for transcriptomic analysis, but they have some limitations such as cross-hybridization between paralogous genes (Wang et al. 2009). In Serial Analysis of Gene Expression (SAGE) and Massively Parallel Signature Sequencing (MPSS) methods, the mRNA population in a sample is sequenced in the form of small tags that correspond to fragments of those transcripts, but these do not allow the discovery of new genes (MacLean et al. 2009). In contrast, tiling microarrays can be used to discover new genes, but they require large amounts of input RNA and have other limitations that affect sensitivity and specificity (Ozsolak and Milos 2009). The limitations of these methods, which include bias towards mRNAs rather than the non-coding RNAs, and the development of next-generation sequencing led to the development of a new approach for transcriptomic analysis called RNA-Seq (RNA sequencing) (Wang et al. 2009; McGettigan 2013). RNA-Seq is an accurate, strand-specific, quantitative method that can detect the complete transcriptome, including all antisense transcripts in a bacterial cell.

3.1.1 *RNA-Sequencing*

In the RNA-Seq approach, total RNA is converted into a library of cDNA. An optional, but important step before construction of a cDNA library is the depletion of ribosomal RNA (rRNA), which is an abundant component (>80% to 90%) of the total RNA sample

(O'Neil et al. 2013). Successful depletion of the rRNA fraction from the sample can enrich mRNA, thereby improving transcriptome coverage. Various methods have been developed for mRNA enrichment, including ribosomal RNA capture (Passalacqua et al. 2009), enzymatic degradation of processed RNA (Sharma *et al.,* 2010), selective polyadenylation of mRNA (Amara and Vijaya, 1997), and antibody capture of RNAs that interact with a specific protein (Sittka et al. 2008). The standard RNA-Seq procedure involves PCR amplification of cDNA and sequencing of the amplified libraries with high-throughput sequencing platforms (Wang *et al.,* 2009). The complete workflow of the RNA-Seq approach is explained in Figure 3.1. Different sequencing platforms, including Illumina (Illumina®), SOLiD (Life technologies™) and 454 pyrosequencing (Roche) are currently being used for RNA-Seq (Ozsolak and Milos 2009; Mutz et al. 2013). The Illumina sequencing platform involves the massively parallel sequencing of millions of fragments using reversible terminator-based sequencing chemistry (http://www.illumina.com/technology/next-generation-sequencing.ilmn).

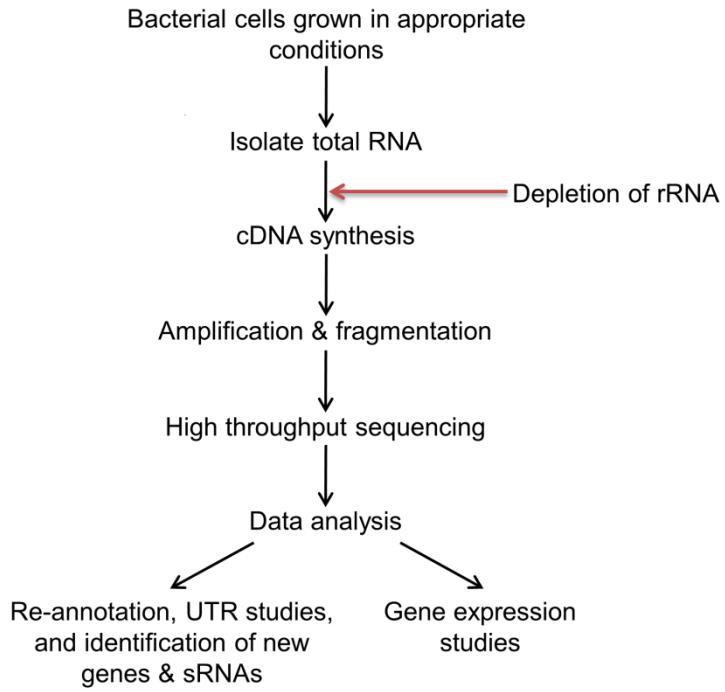


Figure 3.1: Overview of the RNA-Seq approach.
The UTR is the 5' or 3' untranslated region of a gene. Depletion of rRNA from total RNA sample is an optional but important step (see text for other details).

Most protocols for the construction of cDNA libraries involve a PCR amplification step (Kozarewa et al. 2009). However, PCR-amplified cDNA libraries have reduced complexity compared to a total RNA pool because different fragments amplify with unequal efficiency (Kozarewa et al. 2009). This reduces the detectability of some RNA species in the RNA pool by excessive amplification of other RNAs (PCR bias). To overcome this limitation, a new RNA-Seq protocol was recently developed, where the RNA fragment is reverse transcribed on the surface of the Illumina Genome Analyzer flowcell (FRT-Seq) (Mamanova et al. 2010). This method is strand-specific, amplification-free and was intended to avoid any ambiguities arising from the addition of unspecific nucleotides by reverse transcriptase (Mamanova et al. 2010).

A critical requirement of RNA-Seq is the alignment of millions of short reads to a reference genome. The utilization of different sequencing platforms for RNA-Seq can introduce technical problems that complicate alignment strategies due to short read lengths and low sequence quality. Once the reads have been obtained from sequencers, they are aligned to a reference genome to produce a genome-scale transcription map that includes both the transcriptional structure and level of expression for each gene. In 2012, there were over 60 several software tools available for mapping (Fonseca et al. 2012), including BWA (Li et al. 2008), SOAP (Li and Durbin 2009), Bowtie (Langmead et al. 2009; Langmead and Salzberg 2012) and Segemehl (Hoffmann et al. 2009). They have been adapted to different technology and protocol developments and implement various algorithms which result in a range of computational speed and memory requirements. Segemehl uses error-tolerant suffix arrays to compute exact and inexact seed methods to map the reads. The mapped reads are then visualised with the help of browsers such as the Integrated Genome Browser (IGB; Nicol et al. 2009) or JBrowse (Skinner et al. 2009) to allow the manual re-annotation of gene structure and identification of new genes.

RNA-Seq also promises to give highly accurate gene expression measurements (Wang et al. 2009; Wolf 2013). Compared with microarrays, RNA-Seq is able to measure the absolute concentration of transcripts. Biological replicates are required to give robust quantification of differential gene expression (Roh et al. 2010; Croucher and Thomson 2010) but the greater cost of library preparation and sequencing in RNA-Seq often prevents this (Mortazavi et al. 2008). To overcome this issue, statistical methods have been developed to model the expected distributions of sequence reads mapped to a

locus in different samples. Different normalization methods have been implemented to calculate the abundance of each gene expressed in RNA-Seq samples including RPKM (reads per kilobase per million reads; Mortazavi et al. 2008) and TPM (transcripts per million transcripts; Li et al. 2010; Wagner et al. 2012; 2013), and some tools are also designed to study differential gene expression between libraries derived from two different growth conditions, like EdgeR (Robinson et al. 2010) and DESeq (Anders and Huber 2010).

Here, I present a newly-developed RNA-Seq analysis pipeline to efficiently map and analyse the RNA-Seq data generated from *S*. Typhimurium 4/74. The pipeline first checks the quality of the sequenced cDNA reads. Then the reads are mapped to a reference bacterial genome, and the uniquely mapped cDNA reads are formatted for visualisation. Finally, the expression levels of genes within and between conditions are calculated from the uniquely mapped reads (see below) using the TPM measurement. The key findings include the successful mapping of ~1.6 billion short cDNA reads onto the *S*. Typhimurium 4/74 genome. Of the 106 RNA-Seq libraries processed and analysed using my pipeline, the data from 41 have already been published in Kröger et al. (2012; 2013). The pipeline was also used to map the cDNA reads generated from *Leptospira* Copenhageni (Caimano et al. 2014) and *Acinetobacter baumannii* (Dr. Carsten Kröger pers. comm.).

3.2 *Results and discussion*

3.2.1 *Rationale for the development of the RNA-Seq data analysis pipeline*

RNA-Seq is a complicated, multi-step approach involving sample preparation, amplification, fragmentation and sequencing. A single improper operation within any of these steps would result in biased or even unusable data. Therefore, it is always good practice to check the quality of the RNA-Seq data before starting the analysis. The RNA-Seq data analysis pipeline begins with checking the quality of libraries. The output reads from different samples were formatted from FASTQ to FASTA format (Figure 3.2). Then the cDNA reads were aligned to the reference genome. The cDNA reads that failed to align were discarded from the analysis. In most cases these unaligned cDNA reads are probably from the same source, but have poor sequencing quality.

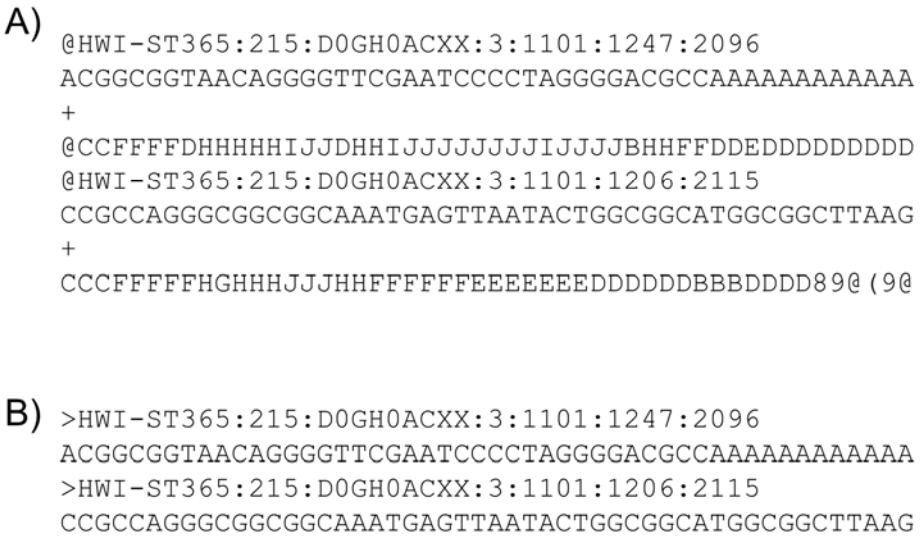


Figure 3.2: Examples for FASTQ and FASTA format sequences.
A) FASTQ format is a text-based format to store both nucleotide sequence and their corresponding quality scores. It uses four lines per read (Cock et al. 2010). Line 1 begins with a '@' character and is followed by a sequence identifier and an optional description. Line 2 is the nucleotide sequence, while the third line begins with '+' character and is optionally followed by the same sequence identifier. Line 4 encodes the quality values for the sequence in line 2 (see source for scores detail: http://support.illumina.com/help/SequencingAnalysisWorkflow/Content/Vault/Informatics/Sequencing_Analysis/CASAVA/swSEQ_mCA_FASTQFiles.htm). B) FASTA is also a text-based format with only two lines. Line 1 begins with a '>' character and is followed by a sequence identifier. The second line contains the nucleotide sequence.

The cDNA reads that mapped to the genome were separated into two categories; uniquely mapped reads – the reads that mapped to only one location in the genome, and multi-mapped reads – the reads that mapped to more than one location in the genome. Only the reads that uniquely mapped to the genome were used for downstream analysis, including visualisation and expression analysis. This way, the data processed using this pipeline will not represent paralogus or repeats.

The graphical files for visualisation were generated from unique reads to visualise the RNA-Seq data. The absolute expression of a gene in a condition was derived from RNA-Seq data by considering the number of cDNA reads mapped to the gene, gene length and the total number reads mapped uniquely to genome. Subsequently, the differential expression of a gene was calculated by comparing the absolute expression of a particular gene between different conditions. The graphical files from multi-mapped reads were also generated for some visualisation purposes. Figure 3.3 depicts the workflow which covers the entire RNA-Seq analysis pipeline. I wrote custom Perl scripts to process every step of the RNA-Seq pipeline.

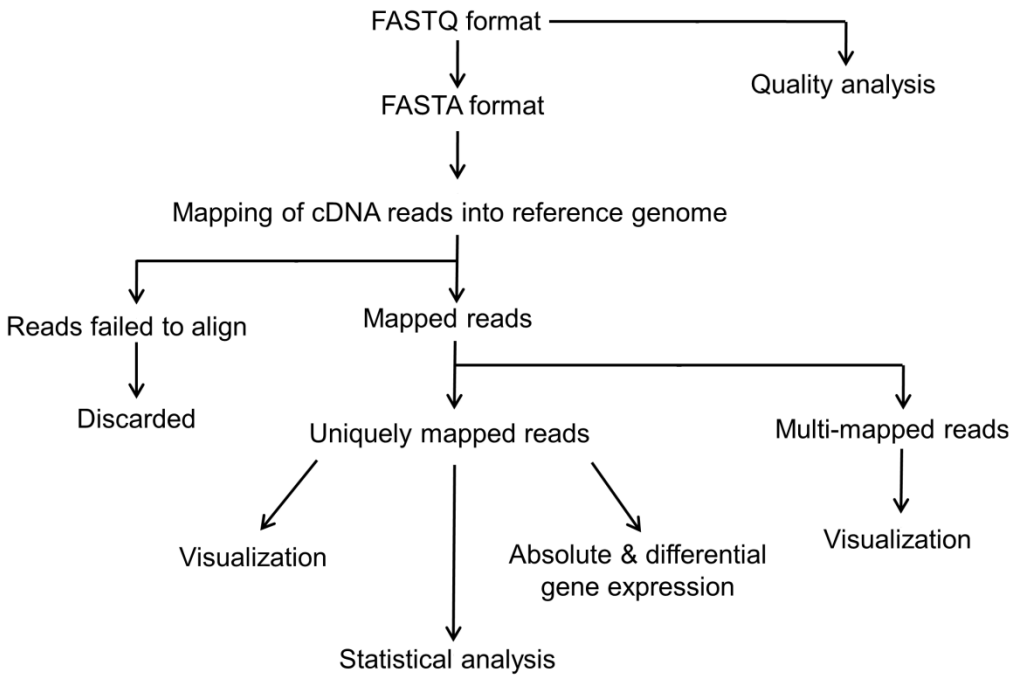


Figure 3.3: The workflow of RNA-Seq data analysis pipeline.

See text for other details.

The distribution of unique reads across the genome was calculated using the Perl script written by me. Figure 3.4 illustrates the statistical analysis of cDNA reads distributed across a bacterial genome. The cDNA reads were categorised according to genomic location, which includes reads mapped to annotated genes, antisense to annotated genes and reads mapped to intergenic regions (IGR). The reads mapped to known non-coding RNAs including rRNA, tRNA, tmRNA (Moore and Sauer 2007) and published regulatory small RNAs were also categorised. The reads mapping antisense to the annotated coding genes and IGR are likely to represent transcripts of new coding or non-coding genes.

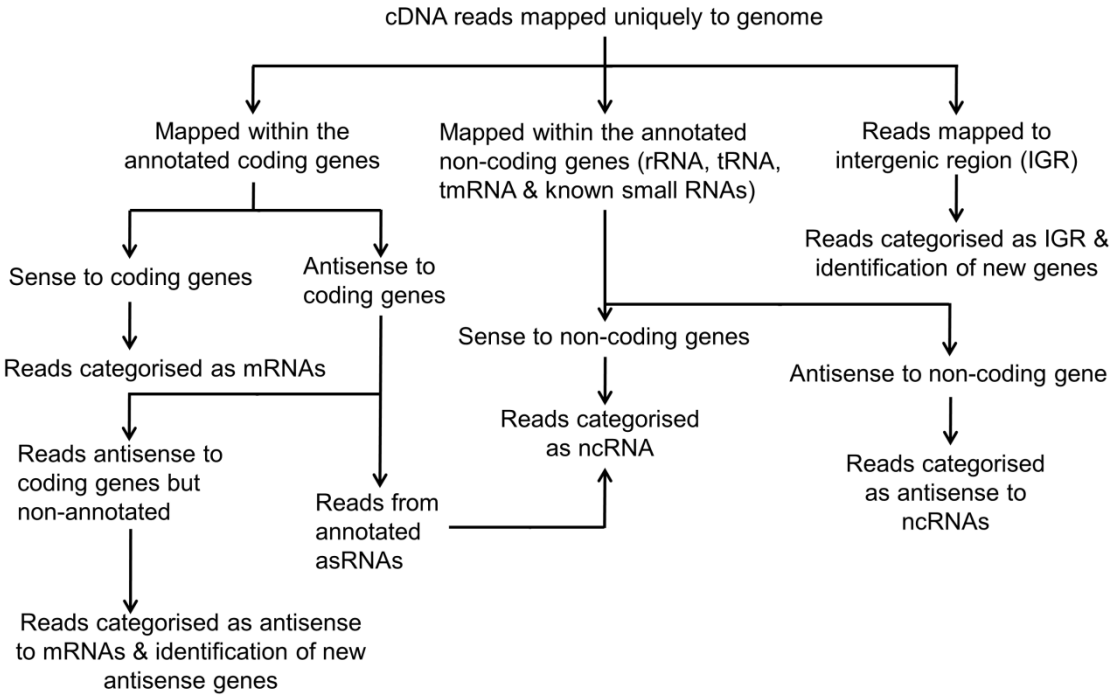


Figure 3.4: Statistical analysis of cDNA read distribution across a bacterial genome.

See text for other details.

3.2.2 *Assessment of bioinformatics pipeline*

The analysis of RNA-Seq libraries (sequenced cDNA libraries) was largely dependent on bioinformatic tools to support the different steps of the process. Errors within any of these steps could lead to wrong results. To assess the efficiency of the newly developed RNA-Seq analysis pipeline, six FRT-Seq datasets generated by the Hinton lab at TCD were used. The bacterial growth conditions and sequencing methods of these libraries are explained in Table 3.1. Custom Perl scripts were used to process the data in each step of the RNA-Seq analysis pipeline, except quality analysis and

mapping. The data processing steps explained here involved my RNA-Seq analysis pipeline and were performed in a Linux environment.

Table 3.1: Description of FRT-Seq libraries used in this study.

Bacterium	Growth condition	Library	Description	Sequencing method	Sequencing platform
Salmonella enterica serovar Typhimurium 4/74	Early stationary phase	ESP-TEX	Primary RNA enriched	Reverse transcription process performed in flowcell surface of Illumina GA (FRT-Seq)	Illumina Genome Analyzer II
		ESP-DEP	rRNA depleted		
		ESP			
	SPI-2 inducing conditions	InSPI2-TEX	Primary RNA enriched		
		InSPI2-DEP	rRNA depleted		
		InSPI2			

ESP – Early stationary phase (LB medium, OD_{600} 2.0)

InSPI2 – *Salmonella* pathogenicity island-2 inducing conditions (PCN, OD_{600} 0.3; Löber et al. 2006).

TEX – Enrichment of primary transcripts by degrading mature RNA population using enzyme terminator exonuclease (Sharma et al. 2010).

DEP – Enrichment of bacterial mRNA by removing rRNA from total RNA using MICROB*Express*™ kit from Ambion®.

ESP and InSPI2 libraries were generated by Dr. Carsten Kröger, Hinton lab, Trinity College Dublin (Kröger et al. 2012; 2013).

3.2.2.1 *Read quality analysis*

Combining all FRT-Seq libraries from both growth conditions, a total of 87,268,648 reads were sequenced with a read length of 42 nucleotides. The quality of the cDNA reads was assessed using the FastQC (http://www.Bioinformatics.bbsrc.ac.uk/projects/fastqc) software tool and the results were analysed manually. Figure 3.5 shows the output from a test data set of high quality (see the figure for source). In comparison, the FastQC results for the FRT-Seq dataset indicate poor quality due to sequencing errors (Figure. 3.6). With the exception of ESP-TEX, all datasets had good quality values for the first 20 base positions, followed by reduction in quality for the remaining bases.

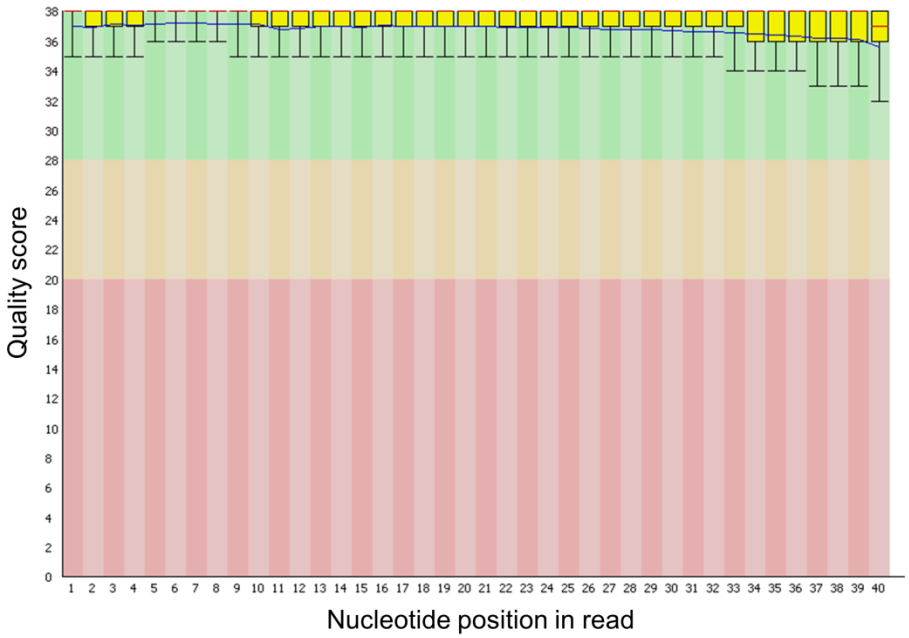


Figure 3.5: An example of good quality dataset from FastQC software tool for the analysis of next-generation sequencing reads.

The Y-axis on the graph shows the quality scores. The higher the score the better the base call. The green colour indicates good quality scores, the brown colour indicates moderate scores and the maroon colour indicates the low quality scores. The X-axis indicates nucleotide positions. The central red line is the median value; yellow boxes represent the inter-quartile range (25-75%); upper & lower whiskers represent the 10% & 90% points; and the blue line represents the mean value of the quality score. Figure source: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/good_sequence_short_fastqc.html.

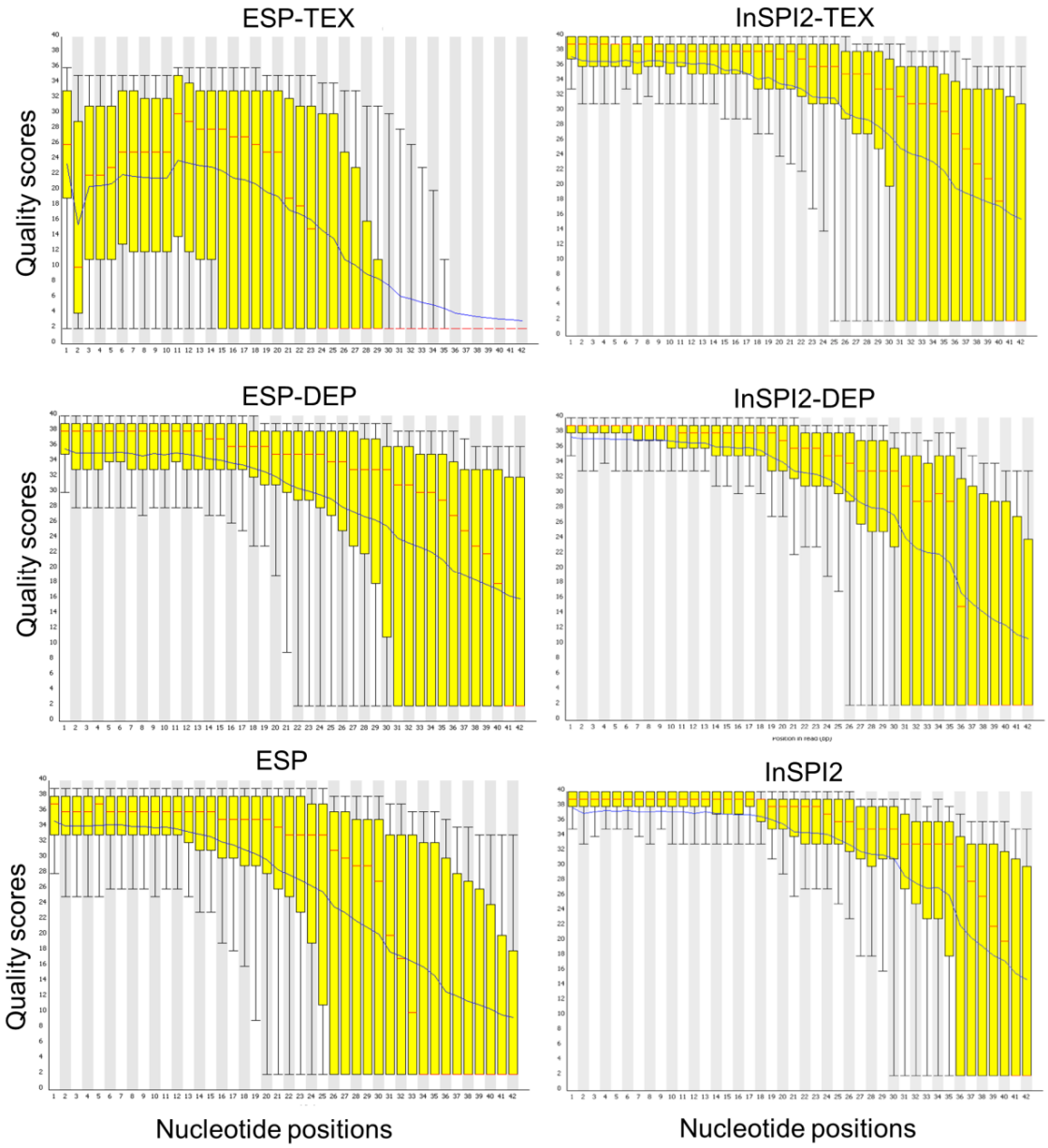


Figure 3.6: Per base sequence quality of *Salmonella* FRT-Seq libraries.
Other details as in the legend to Figure 3.5.

3.2.2.2 *Alignment of cDNA reads to the bacterial genome*

Initially, cDNA reads were formatted from FASTQ format to FASTA format using the Perl script `FQtoFA.pl`. The formatted reads were mapped onto the reference bacterial genome with a publicly available short reads mapping tool named Segemehl (Hoffmann et al. 2009; http://www.bioinf.uni-leipzig.de/Software/segemehl/). Segemehl was chosen for mapping because it showed high accuracy in mapping results compared to other mapping tools (Hoffmann et al. 2009) at the time when the RNA-Seq pipeline was developed. To reduce non-specific alignment, only mappings with 100%

accuracy (without introducing any mismatches or gaps in the alignment) over the full read length (42 nt) were used. The command used for running Segemehl is given below.

Segemehl usage:

```
segemehl.x –A 100 –x index.idx –d SL1344_chro.fa SLP1.fa SLP2.fa
SLP3.fa –q cDNA_reads.fa –u wt_out.fa -t 4 > output.sam
```

Which have the following meanings:

-A % of matches per reads in semi-global alignment (100%)

-x generate the database index and store to disk

-d list of files of database (genomic sequence files)

-q file of query sequence (cDNA reads file)

-u file for unmatched reads

-t threads to use more than one processor at a time

The output file was generated in SAM format (See source for detail: http://chagall.med.cornell.edu/NGScourse/SAM.pdf).

The Illumina reads were mapped onto the *S.* Typhimurium SL1344 genome (http://www.sanger.ac.uk/resources/downloads/bacteria/salmonella.html) due to lack of annotation of *S.* Typhimurium 4/74 (Richardson et al. 2011) at the time of the RNA-Seq analysis. There is only a difference of eight SNPs between the two genomes (seven substitutions and one frame-shift mutation in the chromosome). The three plasmids are identical. The *S.* Typhimurium SL1344 strain is a histidine auxotrophic strain derived from wild-type *S.* Typhimurium 4/74 (Hoiseth and Stocker 1981).

The script `Extract_unique_reads.pl` was used to separate the uniquely mapped reads from the Segemehl output file. The mapping statistics of cDNA reads of the six FRT-Seq libraries are presented in Table 3.2. The total number of reads mapped from all libraries was very low compared to the total number of reads sequenced due to poor sequence quality (*section 3.2.2.1*). A recent study showed that sequencing of 50 million non-rRNA cDNA reads provides complete coverage of bacterial transcripts at one conditions and at least 5-10 million non-rRNA reads would be needed to cover the majority of bacterial transcripts (Haas et al. 2012). Here, the numbers of uniquely mapped reads from each FRT-Seq libraries were too low (see below) to permit downstream analysis, including visualisation and gene expression studies. An alternative method was needed to increase the number of reads mapped to reference genome.

Table 3.2: Analysis of FRT-Seq data mapped to *S*. Typhimurium SL1344.

Libraries	Reads sequenced (42 nt)	Total no. of reads mapped (%)	Uniquely mapped reads (%)
ESP-TEX	67,908	1,119 (**1.6**)	287 (**0.4**)
ESP-DEP	24,585,564	4,810,278 (**19.5**)	1,137,702 (**4.6**)
ESP	18,563,218	3,828,280 (**20.3**)	553,831 (**2.9**)
InSPI2-TEX	22,728,348	1,477,089 (**6.5**)	144,994 (**0.6**)
InSPI2-DEP	14,048,735	55,504 (**0.4**)	14,126 (**0.1**)
InSPI2	7,274,875	83,295 (**1.1**)	17,090 (**0.2**)

The mapping of cDNA reads onto the bacterial genome was done without introducing any mismatches or gaps in the alignment (100% accuracy).

3.2.2.3 *Step-wise mapping methods*

The number of cDNA reads mapped to the reference genome could be increased by reducing the mapping accuracy. This would introduce mismatches, which reduces the mapping quality and potentially lead to non-specific alignments (Ruffalo et al. 2012). To avoid mismatches in the alignment and at the same time increase the quantity of uniquely mapped reads, a novel method was developed to map the datasets. This is an iterative base-clipping approach, which involves the stepwise removal of nucleotides from the 3' end of cDNA reads until the remaining sequence can be accurately mapped to the genome or drops below a minimum read length (Figure. 3.7A). Trimming involved the removal of up to 22 nucleotides from the 3' end of the sequenced reads in a stepwise fashion. These shortened reads were then mapped onto the bacterial genome. A dataset generated in this way, which contains reads of different lengths from 42 to 20 nucleotides is hereby referred as 'SW20' (step-wise shortening to 20 nucleotides). The SW20 method produces a number of mapped read output files (SAM format), and the `Stepwise_unique_extract.pl` script was used to extract the uniquely mapped reads from these files. Based on the SW20 mapping approach, the number of uniquely mapped reads increased in all libraries, e.g., from 2.9% to 13.2% for the ESP library (Figure. 1.6B). The mapping statistics of the FRT-Seq libraries after the step-wise mapping method are shown in Table 3.3.

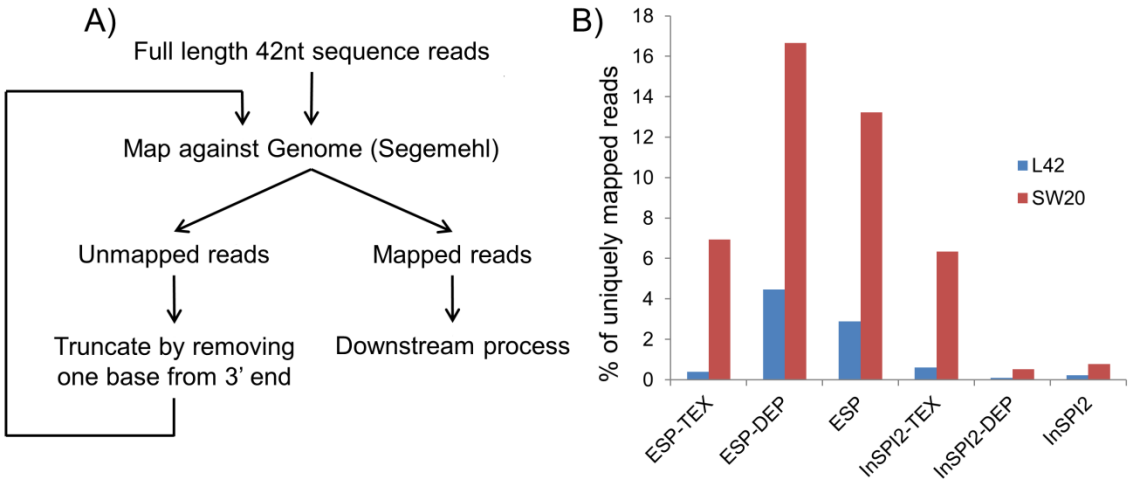


Figure 3.7: Re-mapping of cDNA reads that failed to align to the genome using the Step-wise (SW20) method.

A) Stepwise (SW20) mapping analysis. Downstream processes are visualisation of mapped reads and calculation of absolute and relative expression level of the genes. B) The total proportion of uniquely mapped reads was compared between L42 vs. SW20.

Table 3.3: Mapping statistics of FRT-Seq Salmonella libraries using SW20 method.

Library	Total no. of sequenced reads	Total no. of mapped reads	No. of unmatched reads	No. of uniquely mapped reads into genome
ESP-TEX	67,908	20,415 (**30.1**)	47,493 (**69.9**)	4,711 (**6.9**)
ESP-DEP	24,585,564	14,933,167 (**60.7**)	9,652,397 (**39.3**)	4,093,744 (**16.7**)
ESP	18,563,218	14,359,503 (**77.4**)	4,203,715 (**22.6**)	2,456,792 (**13.2**)
InSPI2-TEX	22,728,348	10,968,198 (**48.3**)	11,760,150 (**51.7**)	1,441,335 (**6.3**)
InSPI2-DEP	14,048,735	280,300 (**2.1**)	13,768,435 (**97.9**)	72,108 (**0.5**)
InSPI2	7,274,875	255,249 (**3.5**)	7,019,626 (**96.5**)	56,726 (**0.8**)

*Uniquely mapped reads are defined as "reads mapped to a single genomic location", and so do not include reads that map to paralogous genes, including most rRNA and tRNA genes.

To determine the efficiency of the SW20 method, 22 nucleotides were removed from the 3' end of all cDNA reads, and only the remaining 20 nucleotides (L20) of the reads were mapped against the genome. This was compared against mapping of the full length reads (L42).

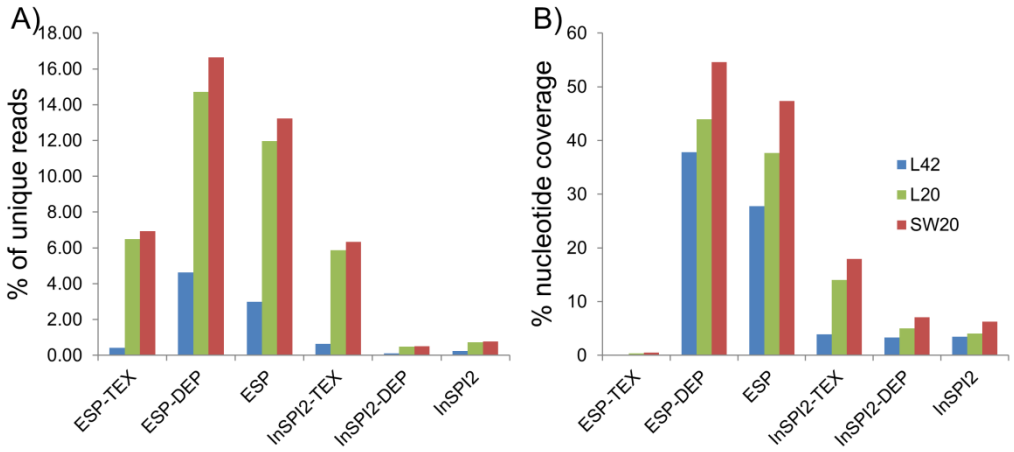


Figure 3.8: Comparison between L42 vs. L20 vs. SW20 approach for mapping sequenced reads.

A) Comparison between three different mapping methods according to the total proportion of uniquely mapped reads. B) comparison between three mapping methods according to their genomic nucleotide coverage. The L42, L20 and SW20 are defined in *Section 3.2.2.3*.

A small amount of variation was observed in the numbers of cDNA reads mapped uniquely to the genome between SW20 and L20 (Figure. 6.8A). However, the genomic read coverage (percentage of nucleotide covered by uniquely mapped cDNA reads) was increased considerably by the SW20 trimming approach compared to L20, e.g., from 43.9% to 54.6% for the ESP-DEP library (Figure. 6.8B).

3.2.2.4 *Read distribution of FRT-Seq libraries*

The RNA-Seq protocol was modified with different methods to enrich the mRNA transcripts (*Section 3.1.1*) including depletion of rRNA. The comparison of the number of cDNA reads mapped against rRNA genes between rRNA depleted and RNA-Seq (without any depletion) libraries would reveal the efficiency of the depletion method. For FRT-Seq, two of the libraries were enzymatically treated (TEX) to degrade the mature RNA transcripts (Sharma et al. 2010) and two were processed using the MICROB*Express* kit to deplete rRNA molecules (DEP) from the total RNA pool (see Table 3.1). To analyse the efficiency of these depletion methods, the read distribution of FRT-Seq libraries was calculated using the Perl script `Read_distribution_calculator.pl` and compared. The Perl script uses the mRNA, sRNA rRNA and tRNA annotations and uniquely mapped cDNA reads as input to calculate the read distribution. All reads mapped to the genome (both unique and multi mapped reads)

were categorized according to genomic locations with the help of a custom Perl script (Figure 3.9A).

Reads mapped to rRNAs and tRNAs comprised ~90% to 95% of the total reads. The rRNA depletion method reduced the rRNA population from 64% to 36% in ESP-DEP compared to ESP, suggesting that the MICROB*Express* kit based methods had a considerable effect. In contrast, a slight reduction (6.2%) in rRNA fraction was observed in InSPI2-DEP compared to InSPI2. The mRNA, sRNA, antisense to mRNA, SRP (signal recognition particles) RNA transcripts were represented by only ~5 -10% of reads. To study these transcripts in detail, the distribution of reads mapped uniquely to the genome was categorized (Figure 3.9B). Approximately 20% of reads correspond to regions annotated as coding sequences in all of the libraries (except InSPI2-TEX) and 10% of the reads represent regions annotated as sRNAs. Many RNA-Seq protocols do not provide information regarding the strand specificity of the transcription (Croucher and Thomson, 2010). Our data show that the FRT-Seq methodology is highly strand-specific because very few reads were mapped antisense (AS) to rRNA and tRNA annotated regions.

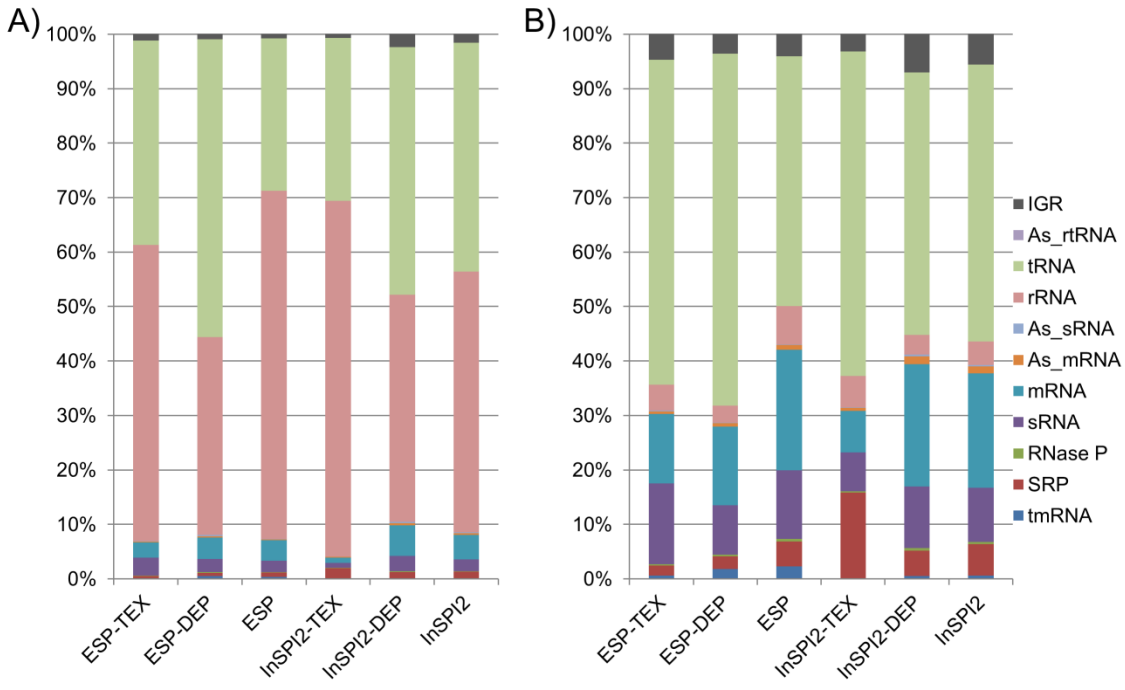


Figure 3.9: Read distribution of *Salmonella* FRT-Seq libraries.
A) Distribution of all mapped cDNA reads (both multi and uniquely mapped reads). B) Distribution of reads mapped uniquely to the genome. SRP - signal recognition particles.

3.2.2.5 *Visualisation of FRT-Seq libraries*

Bacterial genomes comprise both coding and non-coding genes, including mRNA, rRNA, tRNA, sRNA and antisense transcripts. The 5' or 3' untranslated regions (UTRs) of coding genes are known to be important regulatory elements, including riboswitches and binding sites for sRNAs (Serganov and Nudler 2013; Desnoyers et al. 2013; Chao et al. 2012; Kröger et al. 2013). To identify the UTRs, sRNAs and antisense RNAs which were not previously annotated, the uniquely mapped cDNA reads were formatted into graphical files for visualisation. The `Segemehl2IGB.pl` script was used to generate graphical files from mapped cDNA reads. The script generated two sets of graphical files for the plasmids and the chromosome of *Salmonella* (for positive and negative strands). Each file contains two columns of information; the first column indicates the chromosome position (co-ordinates), and the second column indicates the read coverage (number of times the position covered by cDNA reads).

The read coverage used for the graphical display of RNA-Seq data was normalised to compensate for the difference in library sizes within the dataset. The normalization method works as follows: the number of reads covering each nucleotide position was divided by the total number of mapped reads in the library and then multiplied with the number of mapped reads from the smallest library in the dataset, using the formula:

$$N = \frac{10^6 . A}{C} X B$$

Here, A is the read coverage of a nucleotide position, C is the total number of reads mapped (both unique and multi-mapped) in the library to the bacterial genome (library size), B is the smallest library size in the dataset, and N is the normalized read coverage of a nucleotide position.

The FRT-Seq data graphical files were normalised using the script `Normalise_IGB.pl` and visualised using IGB (Nicol et al. 2009) along with annotations. Figure 3.10 shows examples of the visualisation analysis including identification of transcriptional start sites (TSS), new genes, 3' or 5' UTRs, and operons using FRT-Seq libraries. The ESP-TEX library was not included for the IGB visualisation due to its poor quality and the small number of reads that mapped to the genome. The newly identified TSS were grouped into five categories according to their physical location as described in Sharma et al. (2010): primary TSS have a greater

number of cDNA reads within ≤500bp upstream of the annotated gene start; secondary TSS associate with the same gene but have fewer cDNA reads compared to the primary TSS; internal TSS lie within an annotated gene on the same strand; antisense TSS are situated within or near (≤100bp) a gene encoded on the opposite strand; orphan TSS are located in true intergenic regions (≥500bp).

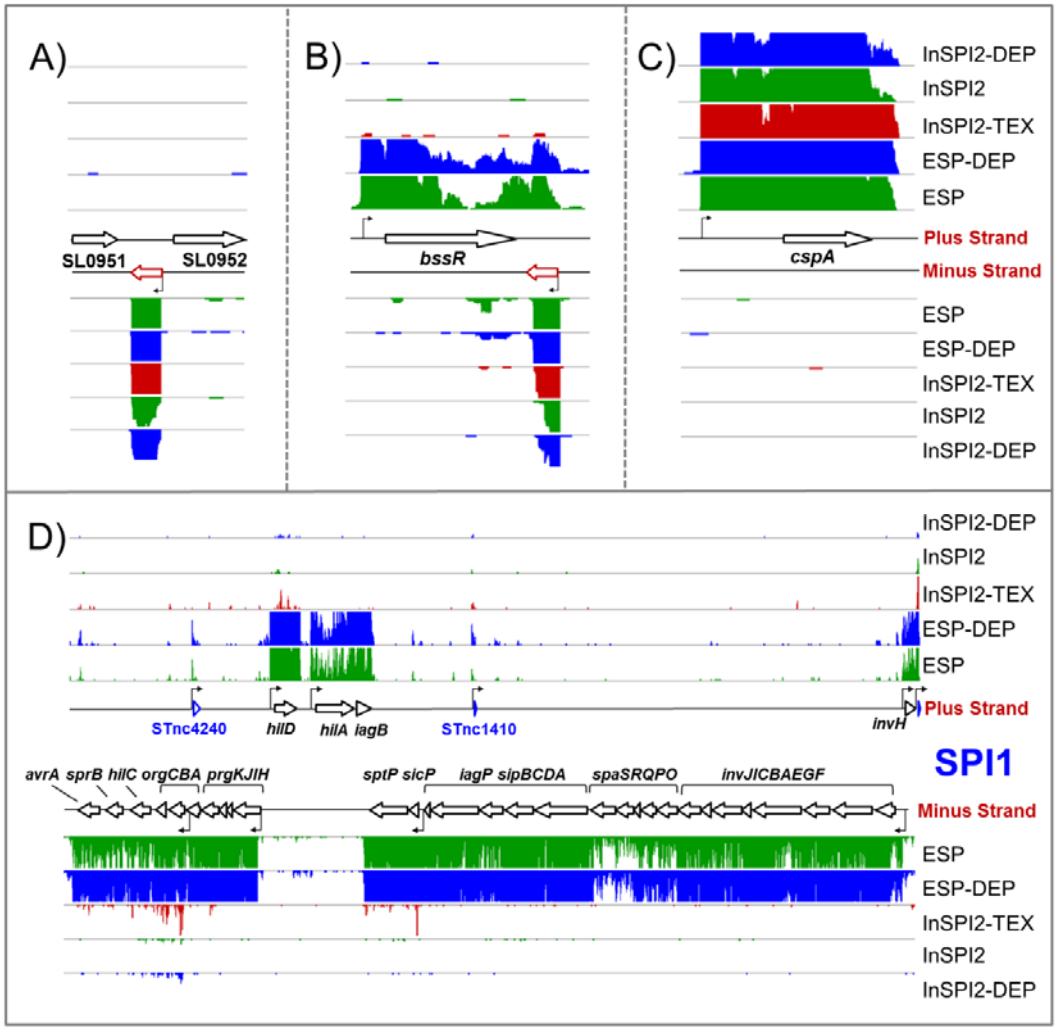


Figure 3.10: Visualisation of FRT-Seq libraries in IGB.
RNA-Seq can be used to improve genome annotation by graphical representation of the data. A) Discovery of a new sRNA gene within the Gifsy-2 phage, subsequently named STnc1080 (Kröger et al. 2013); B) Identification of antisense transcripts (asRNAs), subsequently named STnc1200 (Kröger et al. 2013); C) Detection of untranslated regions (UTRs) and transcriptional start sites (TSS); D) Determination of operon structures. In all panels the vertical scale is 0-20 normalized cDNA reads for every library. The annotated protein coding genes are labelled in black arrows and small RNAs are in blue arrows. The red arrows depict the newly identified genes. The bent arrows indicate the transcriptional start sites.

3.2.2.6 *Gene expression analysis of FRT-Seq data*

High-throughput sequencing technology is rapidly becoming the standard method for measuring gene expression levels (Tarazona et al. 2011). I refer to the absolute expression level of a gene as the transcript abundance within a single library, as defined in Kröger et al. (2013). Transcript abundance is calculated from the number of cDNA reads mapped to the gene and depends on the length of the transcript, the library size and also on the expression levels of the other transcripts (Tarazona et al. 2011; Young et al. 2012). Different approaches have been developed to normalise the data and to allow gene expression levels to be compared between samples, including RPKM (Mortazavi et al. 2008) and TPM (Li et al. 2010; Wagner et al. 2012; 2013). The relative expression of a gene is derived by comparing the absolute expression levels between libraries.

In my RNA-Seq analysis pipeline, the absolute gene expression levels are represented as TPM values. These values provide a more stable measurement (Wagner et al. 2012; 2013) than the still widely used RPKM and FPKM (fragments per kilobase per million), which use the mean length of a transcript in the transcriptome as a normalisation factor. This means that RPKM/FPKM values change between libraries if the mean expressed transcript length differs, even though the transcript abundance might stay the same. TPM values were also used for calculating differential expression. This prevents bias resulting from extremely different conditions, e.g., between EEP and LSP samples (data not shown), which violate assumptions made by commonly used methods, such as DESeq and EdgeR. They work on assumptions that only a small percentage of genes are differentially expressed between two conditions and the numbers of up- and down-regulated genes between two conditions are nearly identical (Dillies et al. 2013). TPM values are not affected by this (Li et al. 2010).

The TPM values were calculated using the Perl script `Gene_reads_cover2TPM.pl`, which was written by Dr. Karsten Hokamp, Department Genetics, Trinity College Dublin, Ireland. The expression values of the *Salmonella* genes in ESP and InSPI2 conditions are listed in Supplementary Dataset 3.1. The genes with TPM value ≥10 were considered as expressed (Kröger et al. 2013). The mean was taken from the TPM value if more than one biological or technical replicates were available for one condition. The relative expression was derived by comparing the absolute expression of a gene between two different conditions (Supplementary Dataset 3.1). If the TPM

value of a gene was lower than the cut-off score (10 TPM), first the value was adjusted to the cut-off, then the relative expression was derived.

3.2.3 *Application of RNA-Seq analysis pipeline*

The newly developed RNA-Seq analysis pipeline was applied to the high-throughput sequencing data generated from the Hinton lab, Trinity College Dublin, Ireland. A total of 106 libraries were generated during this study (Supplementary Dataset 3.2). No rRNA depletion was done prior to the RNA-Seq.

The libraries are:

- 23 sRNA-Seq libraries extracted from 23 infection-relevant conditions. Here, the RNA samples were processed using MirVana^{TM} kit (Ambion) to enrich the RNA transcripts which have a sequence length of <500 nucleotides.
- 31 RNA-Seq libraries that were generated from 22 infection-relevant conditions, which includes biological replicates (6 X ESP and 2 X LSP, and 2 X InSPI2) and technical replicates (3 X MEP).
- Eight dRNA-Seq libraries from MEP, ESP, LSP and InSPI2 to identify the TSS of genes expressed under these conditions.
- Eight MAC libraries constructed from the *Salmonella* grown within the macrophage cells.
- Nine target hunting libraries to identify the putative targets of five *Salmonella*-encoded sRNAs (four from chromosome and one pSLT$^{4/74}$-encoded).
- Six pool libraries were generated by pooling the RNA samples extracted from multiple growth conditions to identify TSS of *Salmonella* genes in a single step.
- 19 mutant libraries constructed from the RNA samples that were extracted from 17 regulatory mutants to study the role of these regulatory proteins at the transcriptomic level.

Combining all libraries, a total of 1.75 billion short reads were sequenced with the read length varying from 50 to 100 nucleotides. The reads were sequenced using different sequencing platforms including FLX Genome Sequencer (Roche), Genome Analyser II and HiSeq 2000 (Illumina). The library size varies from a hundred thousand to forty million sequence reads. All libraries were analysed using my RNA-Seq pipeline. The genomic sequence *S*. Typhimurium 4/74 was used as a reference for mapping after it was published by Richardson et al. (2011) (NCBI Acc: NC_016857). Using the SW20

method, ~1.6 billion cDNA reads (~90%) were mapped successfully to the reference genome. The number of reads mapped from each library varied from 43% to 97%. Of the 1.6 billion reads mapped, 37% (~648 million) of cDNA reads were unique to the genome. Approximately 2% of cDNA reads mapped to the three plasmids of *Salmonella*. The uniquely mapped reads alone were included for visualisation and gene expression studies. The reads were formatted into graphical files for visualisation using the IGB browser. The absolute expression levels were expressed as TPM values (Li et al. 2010; Wagner et al. 2012; 2013). The analysis results from 7 libraries were published in Kröger et al. (2012) and 34 libraries in Kröger et al. (2013).

3.2.4 *Limitations of the RNA-Seq pipeline*

The limitations of the RNA-Seq pipeline are as follows:

- To map the cDNA reads onto the genome the mapping tool Segemehl (Hoffmann et al. 2009) was chosen. At the time the RNA-Seq pipeline was developed, Segemehl showed high accuracy and fewer mapping errors than other tools. But mapping cDNA reads using Segemehl is computationally more expensive than other short read mapping tools, such as Bowtie (Langmead et al. 2009) and BWA (Li et al. 2008). The advent of Bowtie2 (Langmead and Salzberg 2012) brought a tool that is highly accurate and makes as few errors as Segemehl but runs much faster. The pipeline was built in a way that the mapping tool can easily be switched.

- The RNA-Seq pipeline aligns the cDNA reads using the software tool Segemehl (Hoffmann et al. 2009) with the accuracy set to 100%. The cDNA reads that fail to map due to low quality nucleotides at the 3' end were mapped into genome using the SW20 method. However, reads with wrongly called bases at the start of the reads cannot be mapped by the SW20 method. This issue can be solved by modifying the SW20 approach to trim low quality nucleotides at the 5' ends or by allowing a small number of mismatches.

- Mapping of cDNA reads using the SW20 approach is a very time-consuming computer-intensive process due to iterative mapping cycles that are required after each trimming step. This issue can be solved by choosing a mapping tool that a built-in option to trim low quality bases from the 3' ends of the reads, such as Bowtie2 or Novoalign (http://www.novocraft.com/main/index.php).

3.3 *Conclusion*

RNA-Seq has proven particularly powerful for the identification of novel genes and the detection of low abundance transcripts (Mutz et al. 2013). It replaces the microarray as the technology of choice for transcriptomic analysis, providing access to a greater dynamic range of RNA expression levels (Young et al. 2012). Technological advances in high-throughput sequencing necessitate improved computational tools for processing and analysing large-scale datasets in a systemic manner. For that purpose, I have developed an RNA-Seq pipeline that provides many different types of information by a multifaceted analysis that begins with the determination of the quality of raw cDNA reads and produces gene expression levels. The open source software tools FastQC and Segemehl (Hoffmann et al. 2009) were included in the pipeline to determine the sequenced read quality and for mapping, respectively. The TPM (Wagner et al. 2012; 2013) measurement was used for *Salmonella* gene expression studies. Custom-made Perl scripts were used for each steps of data processing. The cDNA reads that mapped uniquely onto the genome were output as graphical format data to visualise the RNA-Seq data using IGB (Nicol et al. 2009). The graphical data were normalised to allow comparison of the expression levels of RNA-Seq data during visualisation. The efficiency of the pipeline was determined using six FRT-Seq datasets. The poor quality of FRT-Seq data spurred the development of a novel read-mapping approach called the Step-wise method. The data extracted from two FRT-Seq libraries (ESP and ESP-DEP) were published in Kröger et al. (2012).

The RNA-Seq pipeline has been used extensively and successfully to analyse 106 RNA-Seq libraries generated by the Hinton lab during this study. Combining all libraries, 1.75 billion cDNA reads were sequenced using different sequencing platforms, of which 1.6 billion reads were mapped to the reference genome. The number of reads mapped from each library varied from 43% to 97% with an average of 89% (combining all libraries). Approximately 37% of cDNA reads were mapped uniquely to the genome, which is ~650 million reads with a genomic coverage of more than 6000 fold. Two percent of cDNA reads were mapped uniquely to the three plasmids of *Salmonella*. The chromosomal data from 7 libraries were published in Kröger et al. (2012) and 34 libraries in Kröger et al. (2013). Similarly, 12 RNA-Seq libraries produced from the African strain *S*. Typhimurium D23580 and sequenced using the SOLiD™ sequencing platform have also been analysed using my RNA-Seq pipeline (Hinton lab, University of Liverpool, UK).

Apart from *Salmonella*, the pipeline has also been extensively used to process RNA-Seq data obtained from the bacterial species *Leptospira interrogans* Copenhageni and *Acinetobacter baumannii.* A total of six libraries were generated from *Leptospira* grown under two different growth conditions (three biological replicates from each condition). Combining six libraries, a total of 62 million cDNA reads were sequenced using the Illumina sequencing platform, and analysed using my RNA-Seq pipeline. The results extracted from *Leptospira* libraries were published in Caimano et al. (2014). Similarly, six high-throughput sequencing libraries comprising 63 million cDNA reads were mapped and analysed using my RNA-Seq pipeline (Dr. Carsten Kröger pers. comm.). The data from these libraries are yet to be published.

CHAPTER 4

SALMONELLA PLASMID GENE EXPRESSION IS CHARACTERISED BY PERVASIVE ANTISENSE TRANSCRIPTION

4.1 *Introduction*

Plasmids are key players in the team of mobile genetic elements that increases *Salmonella* adaptability and diversity. One example is, the multi-drug resistance locus encoded by the virulence plasmid of strain *S*. Typhimurium D23580, which is absent from the virulence plasmid of strains *S*. Typhimurium 4/74 and LT2. In contrast, 4/74 contains a colicin toxin encoded by pCol1B$^{4/74}$, whereas the entire plasmid is absent in strains D23580 and LT2. It is therefore important to study the expression of plasmid genes, including the location of promoters and non-coding RNAs (ncRNAs). Advancements of RNA-Seq technologies, including dRNA-Seq, provide an opportunity to pursue these studies. Using RNA-Seq, the transcriptional architecture, promoters and ncRNAs of bacterial chromosomes have been studied in several bacterial species, such as *Escherichia*, *Chlamydia, Campylobacter, Helicobacter, Salmonella* and *Mycobacterium* (Mendoza-Vargas et al. 2009; Perkins 2009; Albrecht 2011; Sharma et al. 2010; Kröger et al. 2012, 2013; Dugar 2013; Cortes 2013). A recent study by Seo et al. (2012) identified 263 TSS in five plasmids of *Klebsiella* using ChIP-chip experiments. Here, I present the transcriptional organisation of three plasmids of *Salmonella* using a strategy which combines the RNA-Seq, dRNA-Seq and chromatin immunoprecipitation coupled with microarray (ChIP-chip) approaches.

Key findings include the identification of 120 TSS and 113 RpoD-binding regions, which encompass 86 promoter regions and 33 antisense transcripts in the plasmids of strain 4/74. In addition, we identified 83 TSS and 23 antisense transcripts from pSLT-BT^{D23580} of the African ST313 *S*. Typhimurium strain. We determined that all the antisense transcripts of pSLT are conserved between S. Typhimurium strains 4/74 and D23580. The data are consistent with the replication of all *Salmonella* plasmids being tightly regulated by antisense transcription. Overall, the definition of the *Salmonella* plasmid transcriptome structure and promoter maps provides a new insight into the gene regulation and transcriptomic evolution of plasmids.

4.2 *Results and discussion*

4.2.1 *Identification of Transcription Start Sites of Salmonella plasmids*

Promoters are the molecular biological signals that are crucial for the regulation of the expression of genes (Maetschke et al. 2006). Promoter regions are difficult to locate by computational methods because they lie at a variable distance upstream of their associated genes, and can be poorly conserved. It is essential to experimentally determine the site of transcription initiation of a gene in order to locate its promoter region. A combination of RNA-Seq and dRNA-Seq approach-based analyses was performed to identify the TSS of plasmid-encoded genes of *Salmonella* grown to early stationary phase (ESP). ESP is an infection-relevant growth condition associated with high levels of expression of the *Salmonella* pathogenicity island 1 (SPI-1) virulence genes that are responsible for invasion of epithelial cells (Pfeiffer et al. 2007). The reproducibility and robustness of data was ensured by using three biologically independent replicates. Two biological replicates were sequenced with 454-pyrosequencing (ESP Biol. Rep 1 & 2 and ESP [dRNA-Seq] Biol. Rep. 1 & 2), and the third replicate was sequenced on the Illumina High-Seq platform (ESP Biol. Rep. 3 and ESP [dRNA-Seq] Biol. Rep. 3). Because of the different sequencing methods, the depth of the library (number of cDNA reads sequenced and mapped) varies between different biological replicates. The sequencing reads were mapped to the genome of *S*. Typhimurium 4/74 using the pipeline developed in Chapter 3, described in *Section 3.2.2.3*. The mapping statistics and other details of RNA-Seq and dRNA-Seq libraries used here are presented in Supplementary Dataset 3.2.

Overall, the majority (~96%) of uniquely mapped reads mapped to the chromosome, while ~4% mapped to plasmids. The distribution of cDNA reads mapping uniquely to *Salmonella* plasmids was calculated as described in *Section 3.2.2.4* and shown in Figure 4.1. In plasmids, only ~23% of sequenced reads mapped to protein-coding regions, and ~60% of reads mapped antisense to protein-coding genes. Approximately 17% of unique reads mapped to intergenic regions (IGR), which includes the reads mapped to the 5' and 3' untranslated (UTR) region of protein-coding genes.

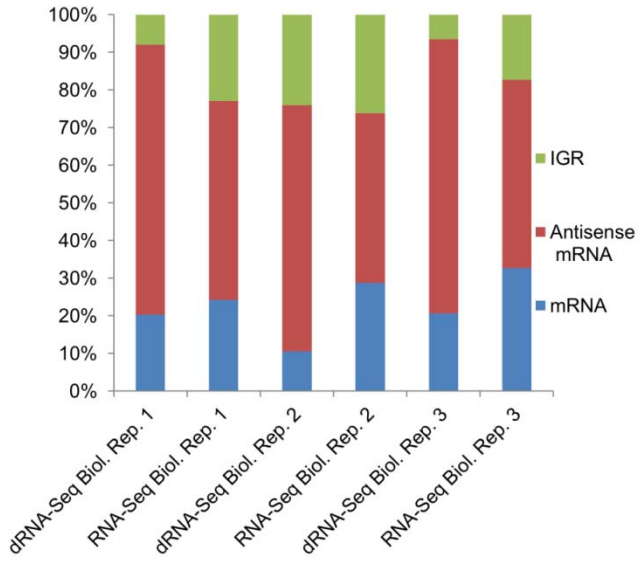


Figure 4.1: Distribution of uniquely mapped reads to different locations on the three plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.
The relative levels of cDNA reads mapped uniquely to coding regions (mRNA), antisense to coding regions (Antisense mRNA) and intergenic regions (region between two adjacent coding gene - IGR) are shown. Read distribution was calculated as described in *Section 3.2.2.4*. Total RNA extraction at ESP was done by Dr. Carsten Kröger, Hinton lab, TCD. I mapped sequencing data to the genome and analysed the plasmid cDNA reads in detail.

Plasmid-associated TSS were identified by the method of Kröger et al. (2012), as described in *Section 3.2.2.5*. The dRNA-seq data often confirmed the TSS that were clearly apparent from RNA-Seq data, as seen for *rlgA* from pSLT$^{4/74}$ (Figure 4.2A). This rigorous and conservative approach for identifying TSS required the same nucleotide location (+1) to be observed in at least two biological replicates. A total of 120 TSS from the three plasmids (Table S4.1) were identified at single nucleotide resolution. The plasmid-coded TSS were classified into five promoter categories (Figure 4.2B) according to Kröger et al. (2012), and the categorization is explained in Figure 4.2C. We assigned 50 primary TSS (pTSS) to protein-coding genes that reveal the main promoters of genes involved in plasmid replication, transfer (conjugation and mobilization), maintenance and partitioning systems (Table S4.1). Our dRNA-Seq based approach also revealed the pTSS of other biologically important genes including the gene responsible for production of colicin (*cib*), the gene conferring sulphonamide resistance (*sul2*) and the gene encoding the thermo-regulator (*tlpA*). In the pTSS of FinP and *traJ*, we observed that there is a 2 nt difference between the dRNA-Seq data and the published start sites (Camacho et al. 2005; based on the primer extension

method). We have already shown that 31 of the 37 published chromosome-associated TSS of *Salmonella* lie within 2 nt of the TSS identified by us (Kröger et al. 2012).

A total of nine secondary TSS were assigned to five functionally known genes (*traS, ccdA* and *repA2* in pSLT$^{4/74}$; *excA* and *sul2* genes in pCol1B$^{4/74}$ and pRSF1010$^{4/74}$ respectively) and four hypothetical proteins. Twenty-four internal TSS were identified in plasmids, which is slightly (~10%) higher than the proportion of internal TSS observed on the chromosome (9%). Overall, 36 antisense TSS were identified in the three plasmids of *Salmonella* at ESP condition (see below).

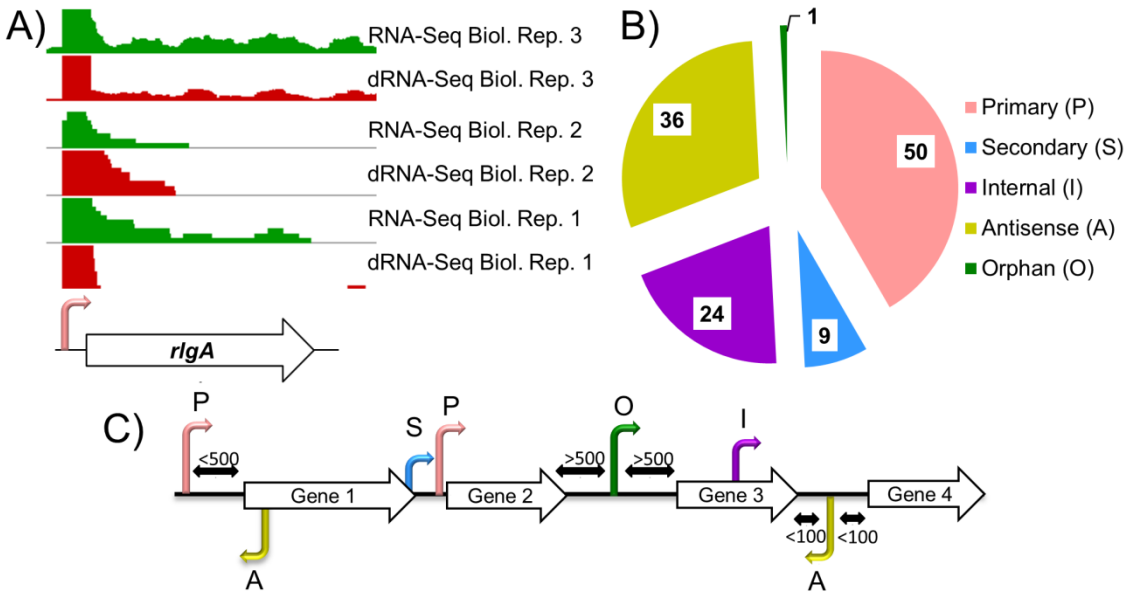


Figure 4.2: Identification and categorization of TSS from 4/74 plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.

(A) Visualization of mapped sequenced reads of three biological replicates of dRNA-Seq (red) and RNA-Seq (green) libraries in the Integrated Genome Browser (IGB) (Nicol et al. 2009). The vertical scale for the replicates 1 & 2 (both dRNA-Seq & RNA-Seq) is 0-10 and for dRNA-Seq_3, RNA-Seq_3 is 0-100 sequenced reads. The identified primary TSS of gene *rlgA* from pSLT$^{4/74}$ is indicated with a coloured arrow. B) Categorization of 120 TSS of the three plasmids identified in this study, of which 50% are antisense or internal relative to coding genes. (C) Explanation of the genomic context of the TSS localization (Kröger et al. 2012).

The dRNA-seq approach has already been used successfully for the analysis of chromosomal TSS in several bacteria (Sharma et al. 2010, Kröger et al. 2012, 2013). The use of the same technology for the identification of plasmid-coded transcripts

makes it vital to put the newly identified TSS from plasmids into a biological context. To validate the TSS, we conducted two 5'-RACE experiments, which exactly confirmed the primary TSS of *sulII* (gene responsible for sulphonamide resistance) from pRSF1010[4/74] (Figure 4.3A) and TSS of an antisense transcript 474P1AS_170 (antisense to *srgC*) from pSLT[4/74] (Figure 4.3B).

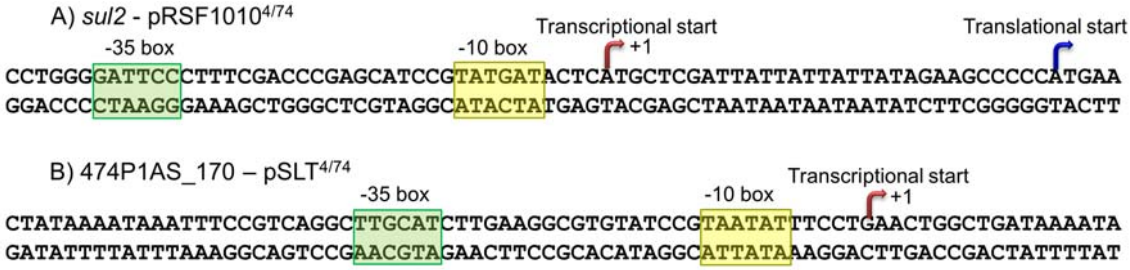


Figure 4.3: Experimental validation of newly identified TSS by 5' RACE.
The TSS of a gene conferring sulfonamide resistance (*sul2*) from pRSF1010[4/74] and an asRNA (474P1AS_170) from pSLT[4/74] were confirmed by 5' RACE. The predicted -35 & -10 boxes are marked with green and yellow boxes. The experimentally-determined transcriptional start sites are marked with red arrows.

4.2.2 *Antisense transcription is pervasive and conserved*

Deep-sequencing based studies have revealed that the level of chromosomal antisense TSS vary between Gram-negative bacteria. For example, 26% of TSS have been reported to be antisense to annotated coding genes of *E. coli* (Mendoza-Vargas et al. 2009; Dornenburg et al. 2010) and 12% in *Salmonella* (Kröger et al. 2012, 2013). A recent RNA-seq based study by Dugar et al. (2013) showed that 47-54% of TSS were antisense in *Campylobacter* isolates. In the plasmids of *Salmonella*, we identified 36 TSS antisense to annotated coding genes (Figure 4.4A), which is ~18% higher than the proportion of antisense TSS of the *Salmonella* chromosome at ESP (12%, Kröger et al. 2012). This finding led us to compare the proportion of cDNA reads mapped antisense to ORFs between plasmids and chromosome, which showed that a higher proportion of cDNA reads mapped antisense to coding genes of plasmids compared to the chromosome (Figure 4.4B). Approximately 48% of cDNA reads mapped antisense to plasmid ORFs. In contrast, only ~2% of cDNA reads mapped antisense to ORFs on the chromosome (Kröger et al. 2012). Widespread antisense transcription has already been reported in the chromosomes of some bacterial species, including *E. coli*, *Synechocystis* species PCC6803, *S. aureus* (Dornenburg et al. 2010; Mitschke et al. 2011; Lasa et al. 2011). Here, our discovery of numerous antisense transcripts in

Salmonella plasmids suggests that antisense transcription may be a common feature of plasmids.

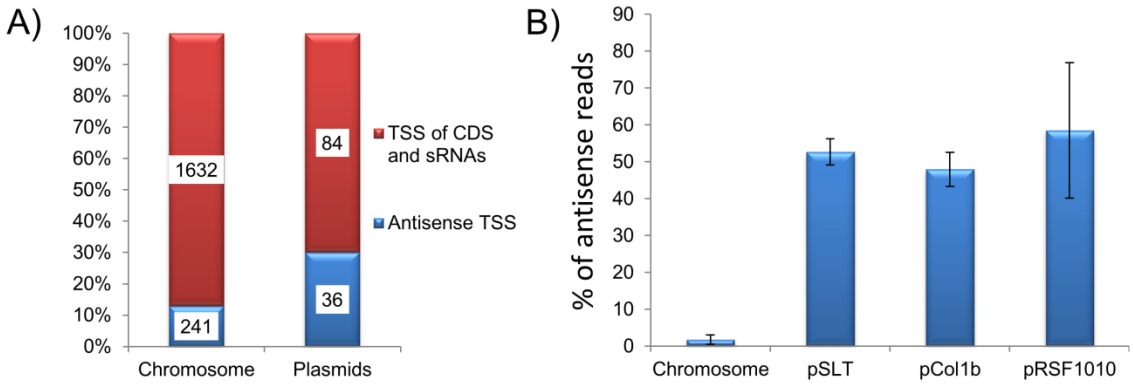


Figure 4.4: *Salmonella* plasmids show extensive antisense transcription.
A) The bar chart represents the total number of chromosomal (Kröger et al. 2012, 2013) and plasmid TSS, and the proportion of antisense TSS (blue). In total, 30% of the plasmid TSS are located antisense to an ORF, which is ~18% higher than for the *Salmonella* chromosome. B) The bar chart shows the percentage of mapped sequence reads located antisense to chromosome and plasmid ORFs (all biological replicates of RNA-Seq libraries were included in this study). In contrast to the chromosome, the ORFs of the plasmids show a higher level of antisense transcription.

Although it has been clear that antisense transcription is pervasive in both eukaryotes and prokaryotes (Lindberg and Lundeberg 2010; Georg and Hess 2011), the biological relevance of asRNA has been debated. Does antisense transcription have a functional role, or is it merely a manifestation of spurious RNA transcription machinery? Although some antisense transcripts have specific regulatory functions, others may result from 'transcriptional noise' generated by non-specific transcription initiation or weak promoters that become fixed within genes by evolutionary constraints on the coding sequence (Struhl 2007). Accumulating evidence suggests that many asRNAs are not conserved between closely-related species and are likely to be non-functional (Raghavan et al. 2012). To determine whether the asRNAs identified in this study are conserved between bacterial stains, we compared the occurrence of asRNAs between pSLT$^{4/74}$ and pSLT-BT^{D23580} from strain *S*. Typhimurium D23580 at the same ESP growth condition.

The nucleotide sequence similarities between pSLT$^{4/74}$ and pSLT-BT^{D23580} have been reported by Kingsley et al. (2009). These two plasmids differ by 29 SNPs and a single

large insertion in pSLT-BT^{D23580} which is a composite Tn*21*-like mobile genetic element. The genes encoded within this distinct transposable element are responsible for all the antibiotic resistance expressed by D23580, and termed the multi-drug resistant (MDR) locus. The organisation of conserved ORFs between two plasmids and the MDR locus from pSLT-BT^{D23580} is shown in Figure 4.5A. The MDR locus was inserted within the coding region of the *rlgA* gene, which encodes a resolvase-like DNA binding protein.

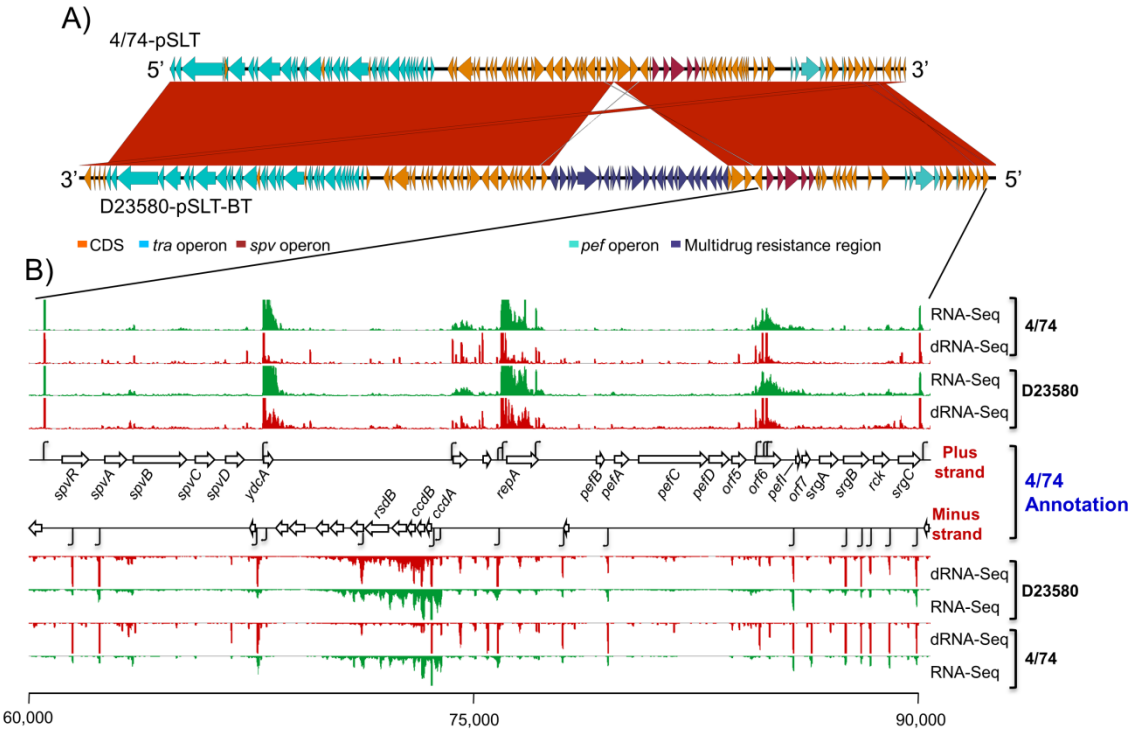


Figure 4.5: Conservation of TSS between the pSLT plasmids of *S*. Typhimurium 4/74 and D23580 at ESP.

A) Organisation of pSLT coding genes in strains compared between 4/74 and D23580. B) The IGB snapshot of a ~30kb region from pSLT displaying TSS conservation between 4/74 and D23580 based on uniquely mapped reads. The RNA-Seq libraries are green in colour and dRNA-Seq libraries are red in colour. The protein coding genes are marked with arrows and the direction of arrow shows the strand information. The functionally-characterised genes are named. The vertical scale for each library is 0-100 normalised cDNA reads. The RNA-Seq and dRNA-Seq of strain D23580 was performed by Dr. Disa L. Hammarlöf, Hinton lab, Liverpool University. Mapping of cDNA reads to D23580 genome was done by Dr. Roy R. Chaudhuri, Liverpool University.

RNA-Seq and dRNA-Seq libraries from *S*. Typhimurium D23580 at ESP condition were generated by Dr. Disa L. Hammarlöf Hinton lab, Liverpool University, UK. The library preparation, sequencing, and mapping protocols used for D23580 were the same as

for 4/74. The mapping of D23580 libraries were done by Dr. Roy R. Chaudhuri, Liverpool University, UK. The depth of the cDNA reads mapped to the D23580 strain was a little higher than the 4/74 data for both the RNA-Seq and dRNA-Seq libraries. Initially, TSS identification in pSLT-BT^{D23580} was done separately and compared with TSS of pSLT$^{4/74}$ (Table S4.1.1 and S4.2). A slight variation was observed in promoter categories, which may reflect differences in sequencing depth (number of reads sequenced) and the presence of the MDR locus in pSLT-BT^{D23580}. Four antisense TSS located in the regions conserved between both plasmids were identified in pSLT-BT^{D23580}, but were not observed in pSLT$^{4/74}$. This difference is likely to reflect the reduced sequence depth of the 4/74 RNA-Seq data.

4.2.3 *asRNAs arise from 5' or 3' UTRs in MDR locus of pSLT-BT^{D23580}*

Our RNA-Seq analysis identified nine pTSS of coding genes expressed within the MDR locus of pSLT-BT^{D23580} (Figure 4.6; Table S4.2), which includes the six drug resistance genes (*sul2, dhfr1, aadA1, qacE, cat* and *blaT*) and three genes involved in transposition (*tnpB, int* and *istA*). The genes encoding insertion elements (IS – *istA* & *istB*) were expressed as an operon. The genes that conferred sulphonamide and streptomycin resistances (*sul2, strA* & *B*) were expressed from a single promoter as a polycistronic mRNA, which is the same operonic structure observed in pRSF1010$^{4/74}$-encoded *sul2, strA* & *B* operon. Four coding genes with antisense promoters were identified within the MDR locus, including the *sul2 & dhfr1* genes that confer drug resistance, and two functionally unknown genes (pSLTBT_p047 & p055).

The pRSF1010$^{4/74}$-encoded asRNA 474P3AS_10 (which is 637 nt in length) is transcribed antisense to the 5' UTR of the *sul2* gene, and overlaps both the IGR between *sul2* and adjacent *repC*, and the 3' end of the *repC* gene. Interestingly, a similar asRNA/mRNA (D23P1AS_150 /*sul2*) transcription pattern was observed within the MDR locus of the pSLT-BT^{D23580}. The sequence similarity and protein coding ability of the D23P1AS_150 and 474P3AS_10 were analysed due to their transcript length (637 nt). The results showed that these asRNAs are 98% identical and do not encode protein. Further experimental and bioinformatics analysis will be required to analyse the functional role and evolution of these asRNAs.

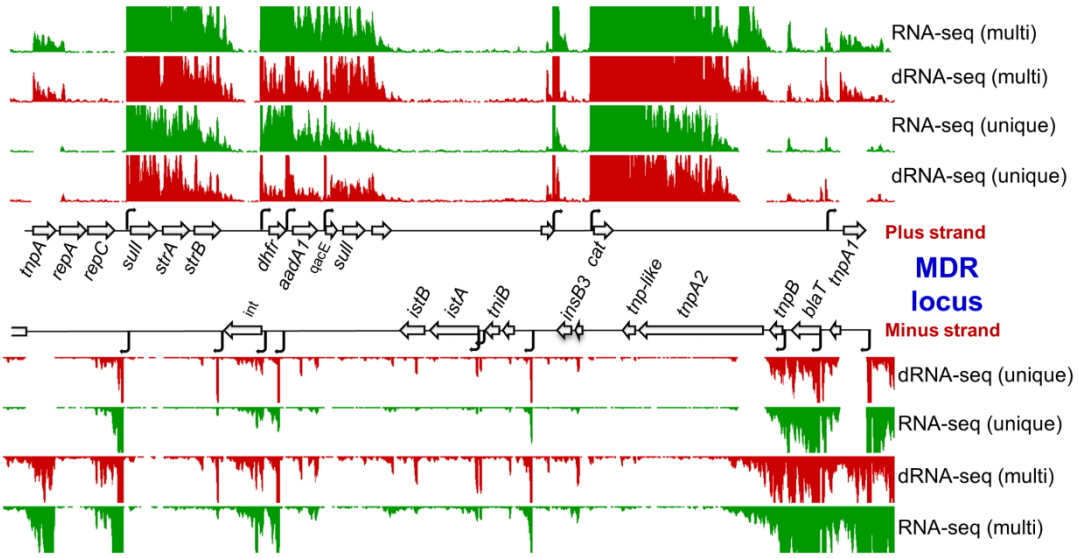


Figure 4.6: Visualisation of pSLT-BT^{D23580}-encoded multi-drug resistance (MDR) region in IGB, using ESP RNA-Seq data.

To identify asRNAs located antisense to the paralogous transposon (*tnp*) genes, we included both unique and multi mapped cDNA reads. The multi-mapped cDNA data (lanes labelled with multi) also include the uniquely mapped cDNA reads. The long 5' and 3' UTRs overlapping the *tnp* genes are clearly visible in multi-mapped RNA-Seq and dRNA-Seq data. The protein-coding genes are marked with arrows, and gene names are labelled in black. TSS are marked with curved arrows. The vertical scale is 0-100 normalized cDNA reads.

Antisense transcripts can originate from the long 5' or 3' UTRs of protein-coding genes (Hernández et al. 2006; Rasmussen et al. 2009; Wurtzel et al. 2012). The expression of asRNAs from 5' and 3' long UTRs overlapping one or more ORFs on the opposite strand were demonstrated in Gram-positive bacteria (Toledo-Arana et al. 2009, Lasa et al. 2011). Here, we discovered a long 3' UTR of the *cat* gene that overlaps the transposon-related genes on the opposite strand (*tnp-like, tnpA2* & *tnpB*) (Figure 4.6), located within the MDR locus. Similar patterns were not observed for the other two *tnpA* genes located at either end of the MDR locus (*tnpA* & *tnpA1*). Further analysis revealed that the cDNA reads mapped to the paralogous *tnpA* and *tnpA1* regions were removed from the data due to high sequence identity, as they were multi-mapped cDNA reads. To visualise the expression of these regions, we added multi-mapped cDNA reads to the data analysis. This revealed that all the transposon-related genes annotated within the MDR locus were overlapped by antisense transcripts (Figure 4.6; multi-mapped cDNA reads lane). For example, the 5' UTRs of pseudogenes pSLTBT_p030 (*rlgA*-like) and p055 overlap the nearby divergent *tnpA* (pSLTBT_p031 & p056) genes.

The excludon is a new regulatory concept that has recently been described by Sesto et al. (2013), who demonstrated that a genomic locus encoding an unusually long asRNA that spans divergent genes or operons with related or opposing functions. Our findings suggest the possibility of a regulatory (negative) linkage between these pSLT-BT^{D23580}-encoded asRNAs and their sense coding genes. Further experimental results are required to determine the functional role of these asRNAs.

4.2.4 *Many asRNAs located at 5' ends of the ORFs*

The detection of antisense transcripts has been difficult in bacteria, but with the advent of high throughput methods such as tiling arrays, direct labelling and, in particular RNA-Seq, the numbers of identified asRNAs have increased dramatically. Our RNA-seq based analysis identified 33 candidate asRNAs in 4/74 plasmids (pSLT$^{4/74}$ – 17; pCol1B$^{4/74}$ – 12; and pRSF1010$^{4/74}$ – 4) at ESP condition (Table S4.3). In addition, 23 asRNA candidates were identified in pSLT-BT^{D23580}, 17 of which were identified in the region shared between the virulence plasmids of both strains, and four were located within the MDR locus (Table S4.4). The remaining two candidate asRNAs (D23P1AS_140 and D23P1AS_140) were only identified in pSLT-BT^{D23580}, even though they were located in the conserved regions of both plasmids. However, a small number of cDNA reads were observed in the corresponding locations of these asRNAs in pSLT$^{4/74}$ (only in dRNA-Seq libraries; no reads were observed in RNA-Seq data), suggesting that these locations might also be transcribed in strain 4/74. Functionally characterized asRNAs involved in replication, conjugation and segregational stability of plasmid have been studied previously (Brantl 2002; 2007). Here, eight asRNAs were identified antisense to genes involved in the above mentioned functions, consistent with a regulatory role for these asRNAs.

In general, we observed two distinct patterns of expression when manually inspecting the mRNA/asRNA regions in IGB. The asRNA strand was more highly expressed than the mRNA region, which we termed "class I" (Figure 4.7A). In contrast, some mRNA/asRNA region showed similar expression levels on both strands, which we termed "class II" regions. Functionally-characterised asRNAs belonged to both classes of mRNA/asRNA expression. For example, FinP antisense to *traJ* was categorised as class I and PndB as class II, which is antisense to *pndC* gene. Binding of FinP to the *traJ* leader sequesters the *traJ* ribosome binding site, preventing its expression and repressing plasmid conjugation (Frost et al. 1989). Similarly, PndB inhibits the

translation of the *pndC* mRNA by blocking the Shine-Dalgarno (SD) sequence (Nielsen and Gerdes 1995). The *pnd* system stabilises the plasmid by killing of new-born plasmid-free cells (also termed post-segregation killing; Gerdes et al. 1990). The expression of FinP and PndB were also validated by northern blot (Figure 4.8B and 4.9C) at ESP.

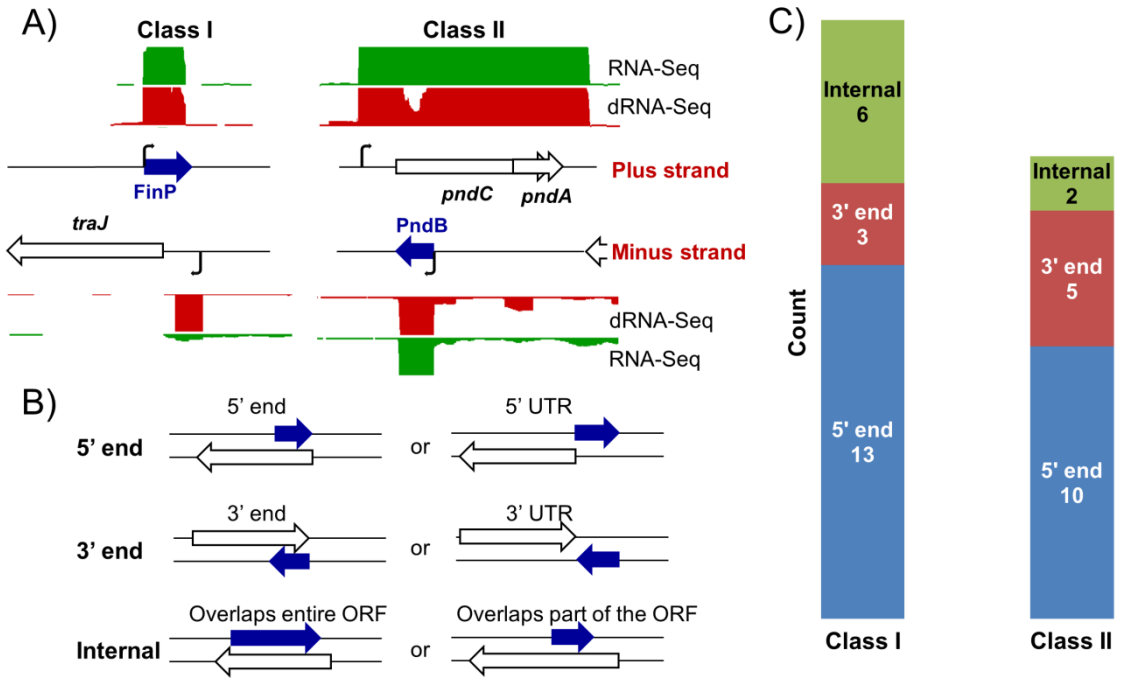


Figure 4.7: Categorization of asRNAs identified from *Salmonella* plasmids.
A) Two different expression patterns of mRNA/asRNA region were visualised in IGB. Class I: asRNA showed higher expression than the mRNA; Class II: asRNA and mRNA showed similar expression levels. Blue and black boxed arrows show the direction of asRNAs and mRNAs respectively. Black curved arrow shows the TSS. RNA-Seq and dRNA-Seq data from biological replicate 3 are presented; vertical scale 0-100 normalised cDNA reads. B) Schematic illustration of the asRNA categories based on their genomic location to the annotated ORFs; for category explanation see the text. Blue and black boxed arrows indicate the asRNA and coding gene respectively. C) Bar chart summarizing the distribution of the asRNA categories.

To classify the asRNA/mRNA regions of the *Salmonella* plasmids, the expression levels of asRNAs and sense mRNAs were calculated using the TPM measurement as described (*Section 3.2.2.6*). If the asRNA had a TPM value > 10 and the cognate mRNA < 10, it was classified as 'class I'. Alternatively, if the asRNA and sense mRNA had TPM values > 10, it was categorised as 'class II'. Out of 39 asRNAs identified in both strains (33 in 4/74 and six in D23580), 22 of the asRNAs were categorized as

class I. The asRNAs opposite to the *spv* and *pef* locus were included in this class. The remaining 17 asRNAs were class II including asRNAs antisense to genes involved in plasmid replication and maintenance, such as 474P1AS_90, 474P1AS_180 and 474P2AS_10 (Table S4.3 and s4.4). Antisense RNAs located complementary to genes conferring drug resistance (474P3AS_10, D23P1AS_150 & 160) were also classified as class II.

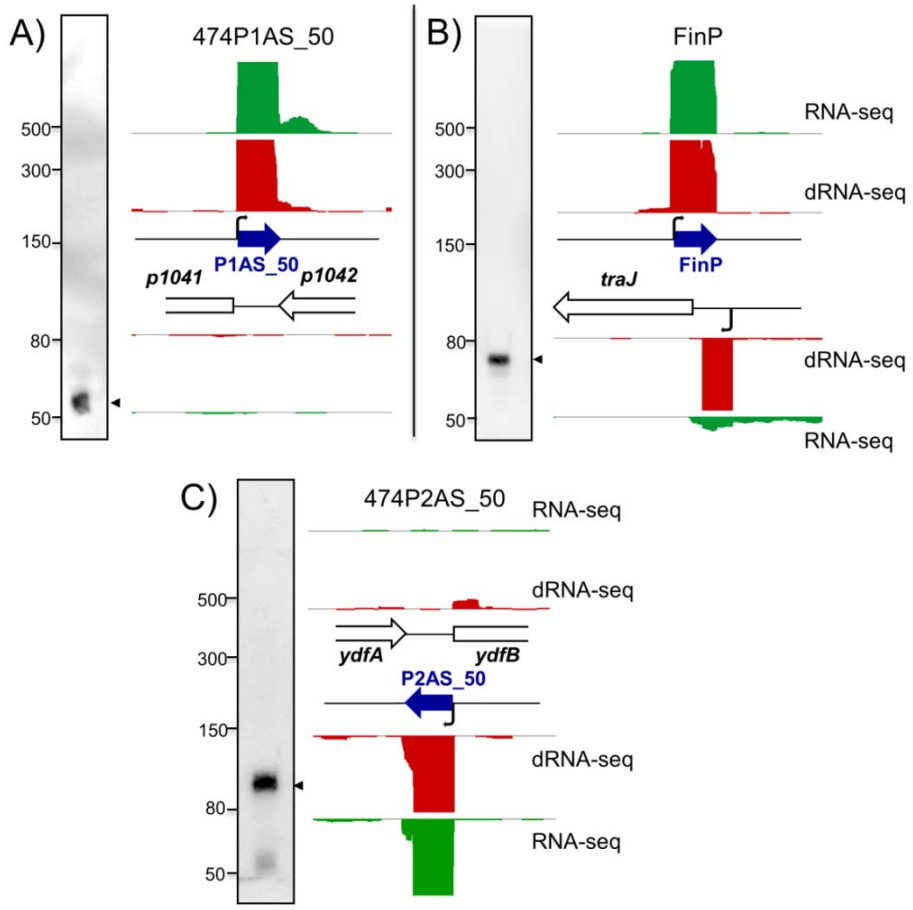


Figure 4.8: Validation of antisense transcripts overlapping the 5' end of the transcripts showing class I expression pattern (Class I/5' end).
Each figure contains a Northern blot (left panel) and RNA-Seq/dRNA-Seq data visualised in IGB (right panel). RNA-Seq (green) and dRNA-Seq (red) data from biological replicate 3 are presented; vertical scale 0-100 normalised cDNA reads. The estimated length of the asRNAs was deduced from RNA-seq data and is shown for each transcript. The orientation of the asRNAs (blue) and their adjacent flanking genes (black) indicated as arrows.

The asRNAs which have a regulatory role in bacteria are functionally classified as 5' overlapping (head to head), internal asRNAs or 3'overlapping (tail to tail) based on

where the antisense RNA binds on the mRNA to form the duplex (Lee and Groisman, 2010; Stazic et al. 2011). To determine the global picture of the location of plasmid-encoded asRNAs relative to genes, they were categorized according to the ORF annotations. If an asRNA overlapped the 5' coding region or 5' UTR of a gene, it was categorized as "5' end". If the asRNAs overlapped the 3' coding region or 3' UTR, it was described as "3' end". Alternatively, if the asRNA covered the entire coding region or if it was located in the middle of the gene, then it was categorized as "internal". The diagrammatic explanation of asRNA categories based on their genomic location is depicted in Figure 4.7B.

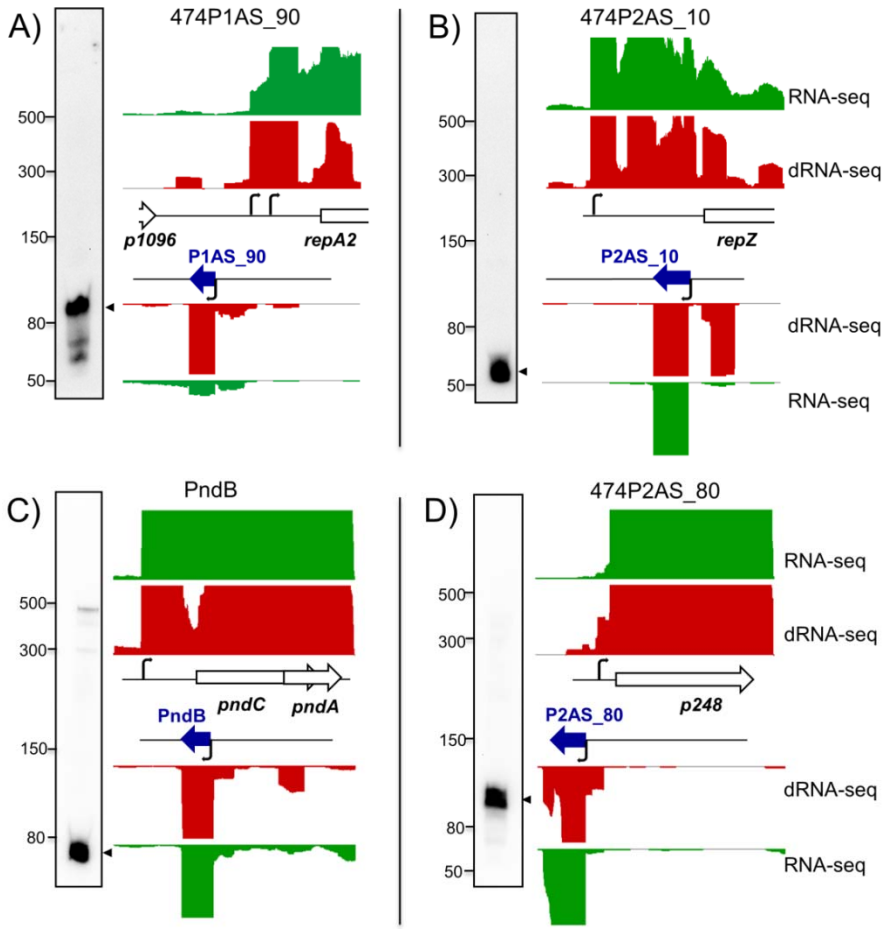


Figure 4.9: Validation of antisense transcripts with class II expression patterns and overlapping the 5' end of the transcripts (class II/5' end).
Details as in the legend to Figure 4.8

Our data showed that the majority of asRNAs (23) were found at the “5’ ends" of genes, which would be consistent with asRNA-dependent gene regulation occurring primarily at the 5’ end of the transcripts. Of the 23 “5’ end” asRNAs, 13 were further classified into “class I” (class I/5’ end), which includes asRNAs opposite to *traJ*, *spvA*, *pefB*, *pefI*, *rck* and *srgB* genes. Northern blot analysis confirmed the presence of asRNA opposite three genes (*traJ*, 474_p1041 and *ydfB)* from this category (Figure 4.8) of asRNAs. The remaining 10 asRNAs (from 5’ end) showed class II expression (class II/5’ end), including the asRNAs antisense to genes involved in plasmid replication (*repA*, *repC* & *repY*) and the segregational killing (*pndC*) process. The expression of four asRNAs from this category was validated by Northern blot with strand specific probes (Figure 4.9). Further analysis of “5’ end” asRNAs revealed that nine asRNA transcripts overlapped the 5’ UTR, including the SD region of the coding genes, with a head-to-head orientation (Table 4.1). Taken together, these data suggest that the “class I/5’ end” asRNAs could have a regulatory role. This possibility should be addressed experimentally.

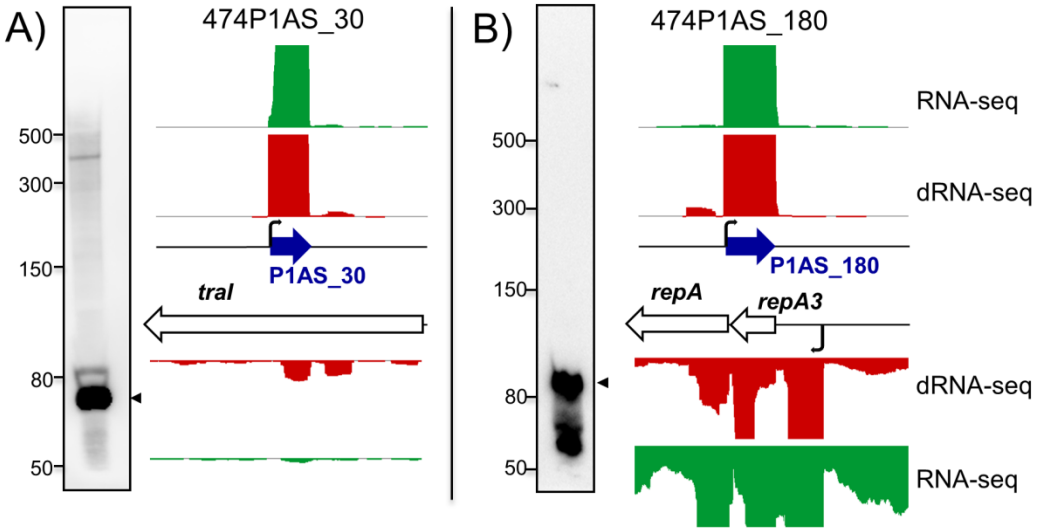


Figure 4.10: Experimental validation of antisense transcripts overlapping within the coding region of genes (internal).

A) 474P1AS_30 categorized as “class I/internal”. B) 474P1AS_180 classified as “class II/internal”. Other details as in the legend to Figure 4.8.

A total of eight asRNAs were identified in the “internal” category, which includes asRNAs antisense to the *traI*, *spvR*, *srgC* and *repA3* gene from pSLT$^{4/74}$; *trbB* and *traJ* genes from pCol1B$^{4/74}$; and the *repC* gene from pRSF1010$^{4/74}$. Of the eight asRNAs, 6 were further classified into “class I” expression (class I/internal) pattern, and the

remaining 2 asRNAs were categorized as "class II/internal". Northern blot analysis confirmed the presence of one asRNA from each "class I/internal" and "class II/internal" categories (Figure 4.10A & B). Finally, eight asRNAs belonged to the "3' end" category, where the asRNA overlaps the sense coding gene with tail to tail orientation. Further classification in these eight asRNA based on their expression levels compared with the sense coding genes revealed that 3 asRNAs were "class I" and remaining 5 were "class II". The asRNAs opposite to *finO*, *pefB*, *rck* genes from pSLT[4/74] were designated as "class I/3' end".

Table 4.1: List of asRNAs that overlap the 5' UTRs of coding genes and could potentially play a regulatory role.

asRNA ID	Strand	Start	End	Length	Antisense to	Class
474P1AS_50	**+**	**38566**	**38637**	**72**	**STM474_p1041**	**1**
474P1AS_60	+	60912	60980	69	STM474_p1070	1
474P1AS_110	-	85661	85731	71	*pefI*	1
474P1AS_150	-	88177	88278	102	*rck*	1
474P2AS_10	**-**	**296**	**365**	**70**	***repY***	**2**
474P2AS_30	-	13885	13965	81	STM474_215A	2
474P2AS_40	-	24026	24132	107	*ycjA*	1
474P2AS_50	**-**	**30007**	**30109**	**103**	***ydfB***	**1**
474P3AS _30	+	2822	2898	77	*repA*	2

Red coloured asRNAs were validated by Northern blot. Other details as Table S4.3.

4.2.5 *Role of Hfq on plasmids-encoded asRNAs*

Hfq, the Sm-like bacterial RNA chaperone protein is required for the stability of many regulatory sRNAs (Valentin-Hansen et al. 2004, Vogel and Papenfort 2006) in bacteria. Hfq is required by the sRNAs to make short and imperfect base-paring with their target mRNAs (*trans*-acting regulatory role; Sharma et al. 2011, Papenfort et al. 2012). In contrast, asRNA interacts with sense target RNA by making long and perfect complementary base-pairing, which typically does not require the Hfq (*cis*-acting regulatory asRNAs). But, accumulating evidence suggests that some asRNAs do bind to Hfq (Klas et al. 2005). Recently our group found that 18 of the 41 annotated asRNAs from the *Salmonella* chromosome were bound by Hfq (Kröger et al. 2013).

To determine whether the plasmid-encoded asRNA bound to Hfq, we used the Hfq-co-immunoprecipitation coupled with high-throughput sequencing data published by Chao

et al. (2012). Raw sequence data were downloaded from GEO-NCBI (GEO accession number GSE38884) and the Hfq-binding characteristics of the 33 asRNAs were determined. The analysis revealed that none of the asRNAs were enriched by Hfq, suggesting that the plasmid-encoded asRNAs are not bound to Hfq. It is possible that other RNA chaperone proteins play RNA binding role on these asRNAs, e.g., FinO. FinO protects the FinP asRNA from degradation by RNase E (Jerome et al. 1999) and acts as an RNA binding protein that facilitates the FinP-*traJ* interactions (Arthur et al. 2003).

4.2.6 *Transcriptional activity across the plasmids of Salmonella*

All RNA molecules in bacterial cells are synthesized by the complex molecular machine known as RNA polymerase (RNAP). RNAP along with sigma factor, binds to the promoter regions and drives the transcription initiation of a gene. To confirm that the identified TSS were indeed associated with bacterial promoters, the transcriptionally active areas of *Salmonella* plasmids were defined experimentally. ChIP-chip was done with a monoclonal antibody that recognized β-subunit of RNAP (Supplementary Dataset 4.1). The RNAP ChIP-chip data was directly compared with newly predicted TSS of all three plasmids. Combining three plasmids, 82 regions showed dynamic binding of RNAP that covered 28% (52,404 nt) of the plasmid sequences. Our observation found that 76 of the 120 plasmid-encoded TSS (63%) were bound by RNAP (Figure 4.11; Table S4.1).

To identify promoter regions more precisely, the RNAP ChIP-chip experiment was also performed after treatment with rifampicin (Rif). Rifampicin is an inhibitor of transcription elongation (Herring et al. 2005) that effectively anchors RNAP at bacterial promoters. In this experiment, RNAP+rif covered only 6.5% of the plasmid regions and the binding regions overlapped with 41 TSS (34%). A total of 79 TSS associated with RNAP (RNAP and/or RNAP+rif).

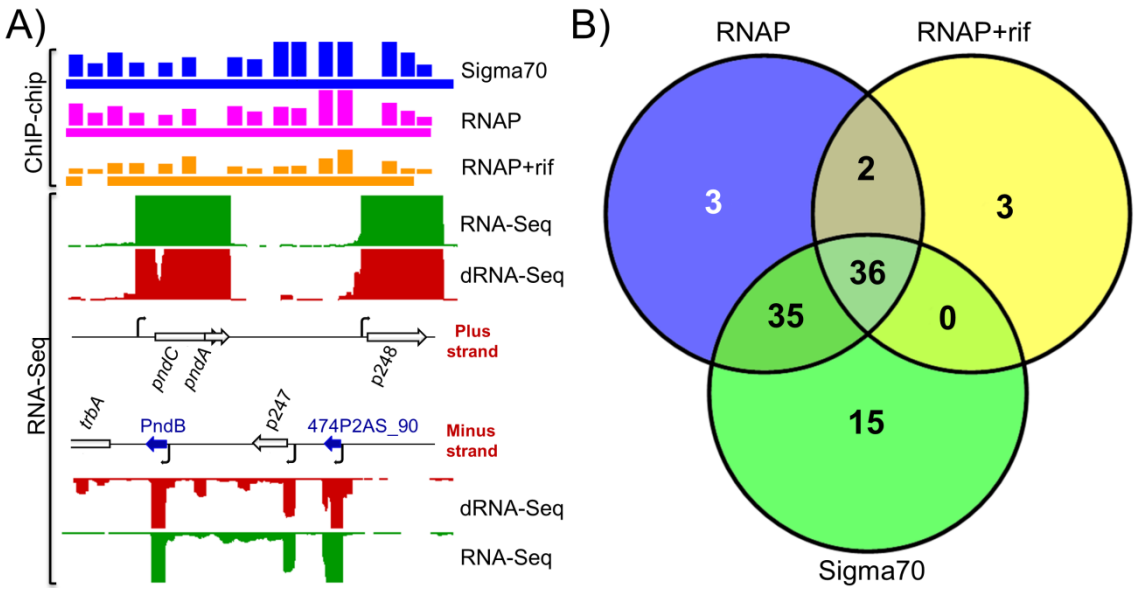


Figure 4.11: Interaction of the transcriptional machinery with transcriptional start sites in *S*. Typhimurium 4/74 plasmids.

A) Visualization of ChIP-Chip results (blue, Sigma70; purple, RNAP; orange, RNAP+rif) along with RNA-Seq datasets using IGB. The RNA-Seq (green) and dRNA-Seq (red) libraries of biological replicate 3 are shown here; vertical scale 0-100 normalised cDNA reads. Each ChIP-chip dataset is presented as quantitative data in top lane and ChIPotle predicted binding sites depicted below each lane as bar. The region has three pTSS and two asTSS. The ChIP-chip data show that these TSS are bound by RNAP, RNAP+rif and Sigma70. B) Venn diagram indicates the number of the TSS that lie within the binding region of the RNAP (blue), RNAP+rif (yellow) and RpoD (green). Approximately 66% of TSS were bound by RNAP and/or RNAP+rif. The number of TSS associated with Sigma70 bound regions are little higher than the TSS lie within the RNAP bound regions. In total, 21% of TSS did not lie within predicted RNAP and/or Sigma70 binding regions. The ChIP-Chip experiment was performed by Dr. Shane C. Dillon, Dorman lab, Trinity College Dublin.

The sigma subunit is the key regulator of the bacterial transcription. Among the different sigma factors, σ^{70} (Sigma70 or RpoD) also known as 'housekeeping sigma factor' plays a major role in transcription initiation during the cell growth (Paget and Helmann 2003). Sigma70 directs RNAP to specific promoter sites that are usually 5-6 bp in length and are centred 10 and 35 bp upstream (-10 and -35 positions) of the transcription initiation site (Gruber and Gross 2003). Recently, our study in *Salmonella* revealed that the transcription of 63% of the chromosomal genes were initiated by RpoD at the ESP condition (Kröger et al. 2012). To determine the relative importance of the σ^{70} sigma factor in the transcription initiation of plasmid-encoded genes, we performed a ChIP-chip experiment with an anti-RpoD monoclonal antibody. This method identified 113 significant σ^{70} binding regions which covered 41% of the plasmid

sequences (Supplementary Dataset 4.1). The ChIP-Chip experiment revealed that 86 TSS (72%) associated with RpoD, which exceeds the number of TSS bound by RNAP. Of these 86 RpoD-bound TSS, 71 TSS also associated with RNAP binding regions. In total, only 35 TSS were located within binding regions of RNAP, RpoD in each of the three experiments (Figure 4.11; Table S4.1). The primary TSS of functionally known genes from pSLT[4/74] including *traJ, traM,* FinP, *tlpA; cib & imm* genes from pCol1B[4/74] and *sulII* from pRSF1010[4/74] were situated within binding regions of both RNAP and RpoD, which suggests that RpoD is critical for expression of these genes at the ESP condition.

It has been previously reported that RpoS regulates the *spv* region of *Salmonella* virulence plasmid (Fang 1992). A clear mapped peak of RpoD within the *spv* locus indicates that this region may also be regulated by RpoD in ESP conditions (data not shown). Further experimental validation will be needed to confirm the role of RpoD in the transcription regulation of the *spv* locus.

4.2.7 Conservation *of σ^{70} promoter motifs between plasmids and chromosome*

Recently we identified the promoter consensus of the *S*. Typhimurium chromosome (Kröger et al. 2012), but promoters of *Salmonella* plasmids have not yet been analysed. To test whether σ^{70} targets the same DNA motifs in plasmids as in the chromosomes of *Salmonella* and *E. coli*, we searched for the conservation of the DNA motifs in TSS that are overlapped by both RNAP and σ^{70} in the ChIP datasets (n = 71). To identify canonical σ^{70} motifs upstream of chosen TSS, an unbiased motif search was performed using the MEME software tool (Bailey et al. 2006) as described (*Section 2.9*). The σ^{70} promoter analysis revealed that the motifs from *Salmonella* chromosome and plasmids are very similar. The extended G at position -3 within the -10 box of the *Salmonella* chromosome (Kröger et al. 2012) also appeared in plasmid-encoded -10 Pribnow motifs (Figure 4.12), which shares extensive homology with the -10 motif of the *E. coli* chromosome. Plasmid-encoded -35 motifs differed from chromosomal -35 motifs by the frequency of G nucleotide at position '1', and the disappearance of a 'C' nucleotide at position '2', but a higher number of TSS is needed to evaluate the significance of this finding. In general, more σ^{70} sigma factor binding sequences will be needed to predict the motifs precisely, and to understand the significance of nucleotides that appear to differentiate the plasmid and chromosome motifs.

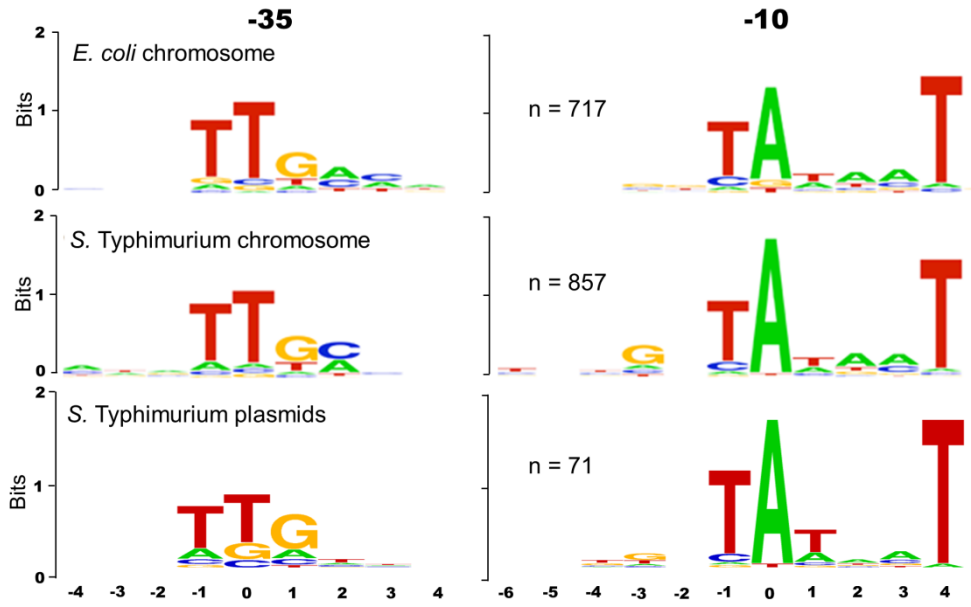


Figure 4.12: Sigma70 binding motif of plasmids is compared with chromosomes of *E. coli* and *Salmonella*.

The Sigma70 binding motif of 71 genes from plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$, is predicted using MEME software tool (Bailey et al. 2006). The motif logo is generated using WebLogo tool (Crooks et al. 2004). *Salmonella* and *E. coli* Sigma70 motifs are adapted from Kröger et al. (2012). The X-axis represents the nucleotide positions as suggested by Shultzaberger et al. (2007) and Y-axis represents the frequency of occurrence of four nucleotides in each position.

4.2.8 *Re-annotation, 5'-UTRs and transcription initiating nucleotides*

The availability of experimental evidence (dRNA-Seq and RNA-Seq) showing TSS and expression of plasmid-encoded genes allowed us to examine whether our data could be used to improve the accuracy of plasmid coding gene (CDS) annotation. We found three examples of TSS that fell within the coding region of the annotated proteins in pSLT$^{4/74}$, suggesting an incorrect translational start site (Table S4.5) in pSLT. In addition, 1 small ORF was found in pCol1B$^{4/74}$, which encoded a DinI-like protein. Reads were visible in the RNA-Seq data, indicating that this sORF is expressed at ESP. Finally, the CDS annotation of plasmids from strains 4/74 D23580 were improved using the published annotation from *S.* Typhimurium SL1344 (Kröger et al. 2012) (Table S4.5 and S4.6).

A 5'-untranslated region (5'-UTR) is defined as the DNA sequence located between the transcriptional and translational start of a gene. The 5'-UTR can be an important regulatory element that carries riboswitches or binding sites for sRNAs (Storz et al.

2011). These regulatory elements cannot be detected computationally. By using the RNA-seq approach, the 5'-UTR of protein-encoded genes of the *Salmonella* plasmids were analysed. The 5'-UTR length of plasmid-encoded genes varied from 12nt (*repC*) to 513nt (*parB* – primary also internal to *parA*). The average length of a plasmid 5'-UTR is between 20 to 35 nt long (Figure 4.13A), which is similar to the 5'-UTRs in the *Salmonella* chromosome (Kröger et al. 2012). Leaderless mRNAs were not identified in any of the plasmids.

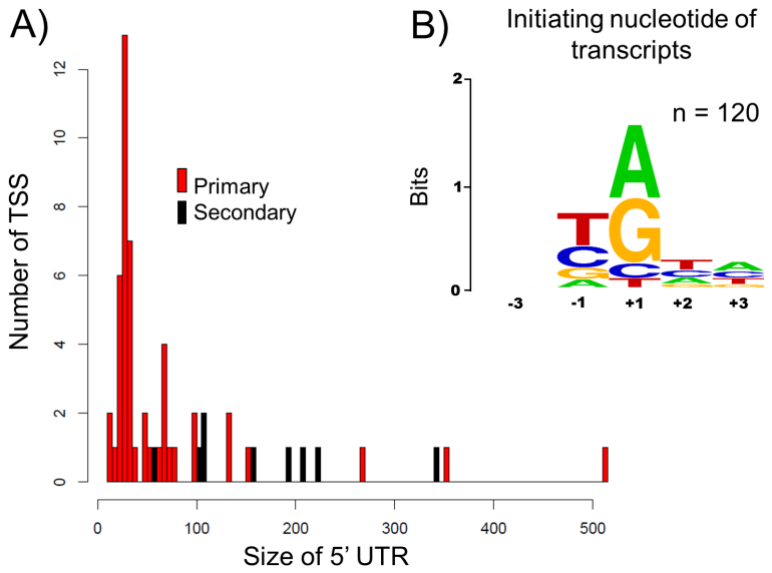


Figure 4.13: Analysis of 5' UTR and the initiating nucleotides of transcripts encoded by three plasmids.

A) Distribution and frequency of the length of primary and secondary 5'-UTRs of plasmid protein coding genes. The 5' UTRs from 50 pTSS (red) and 9 sTSS (black) are used in this study. B) The sequence logo of initiating nucleotides of transcription starts generated from 120 TSS. The X-axis and Y-axis represent the nucleotide position and occurrence of 4 nucleotides in each position

The initiation of transcription imposes some unique mechanistic requirements on RNA polymerase. Most RNAP show a similar preference for the initial nucleotide of the transcript. The RNAP of *E. coli* often initiates transcripts with ATP, although some begin with other NTPs like GTP (Reddy 1994). Of the 120 TSS mapped in all three plasmids, the majority of transcripts initiate with purine nucleotides (ATP = 40% & GTP = 30%) (Figure 4.13B), which resembles the nucleotides used for the transcription initiation of chromosomal genes.

4.3 *Conclusion*

Various technologies have been developed to study the transcriptome of bacteria, including microarrays and Serial Analysis of Gene Expression (SAGE). However, the development of high-throughput sequencing techniques revolutionised the transcriptomic study of bacteria. Here, we have defined the primary element of gene expression on plasmids of *Salmonella*, the promoters. Using dRNA-Seq, we identified 120 promoters including 50 primary promoters of protein coding genes and 36 promoters antisense to coding genes. Our ChIP-chip experiments identified 86 σ^{70}-dependent promoters, which suggests that RpoD plays a dominant role in plasmid gene expression at ESP. It has been previously reported that RpoS regulates the *spv* region of *Salmonella* virulence plasmid (Fang et al. 1992). A clear binding region of RpoD within the *spv* locus indicates that this region may also be controlled by RpoD at ESP. Further experimental validation will be needed to confirm the role of RpoD in the transcription regulation of the *spv* locus.

Significantly more antisense transcripts were identified in *Salmonella* plasmids (17%) than observed on the chromosome (1.5%). However, we note that the reported estimates of antisense transcription vary greatly among bacteria, from 1% to 60% of annotated genes (Georg and Hess 2011, Lasa et al. 2011, Lioliou et al. 2012, Lybecker et al. 2013). Our RNA-Seq based approach identified 37 candidate asRNAs in *Salmonella* plasmids, which includes the asRNAs located opposite to genes involved in plasmid replication, maintenance and virulence. We identified two candidate asRNAs opposite to genes of the *tra* operon. It is known that pSLT from *S*. Typhimurium SL1344 has a very low frequency of conjugation in comparison to pSLT from *S*. Typhimurium LT2 (Ahmer et al. 1999; García-Quintanilla and Casadesús 2011) (note that the pSLT from strains SL1344 and 4/74 are identical). Due to their location antisense to *traI and finO* genes, it is possible that these new asRNAs could be involved in regulation of the *tra* gene transcript levels and might account for the low level of plasmid conjugation in SL1344 (García-Quintanilla and Casadesús 2011). Two candidate asRNAs were identified that were transcribed antisense to *spvA* and *spvA*. We identified seven asRNAs within the *pef* locus, including *srgA, B, C, orf5, 6, 7*, and *rck*, which made this region a hot spot for antisense transcription. The asRNA within the *spv* and *pef* locus are of interest because of the biological importance of their sense genes, and further experimentation is needed to determine any regulatory significance. We identified asRNAs located antisense to genes involved in replication of all three *Salmonella* plasmids. The presence of antisense transcripts overlapping the *rep* genes

(*repA, repC* and *repY*) encoded by *Salmonella* plasmids, raises the possibility that replication of these plasmids is tightly regulated by asRNAs.

The categorization of asRNAs based on their location relevant to mRNAs revealed that 22 asRNAs were antisense to the 5' end (5' ORF and/or 5' UTR) of the sense coding genes. Further analyses revealed that nine asRNAs overlapped the ribosomal binding region of the coding genes. This finding would be consistent with a *cis*-acting regulatory role for these nine asRNAs. Our analysis failed to identify any plasmid-encoded asRNA that bind to Hfq, which suggests that Hfq may not play a role in the stabilisation of plasmid-encoded asRNAs, and that the asRNAs are unlikely to have a *trans*-acting regulatory functions.

Table S4.1: Transcriptional start sites of plasmid-encoded genes of *S*. Typhimurium 4/74 at ESP.

Promoter categorization was carried out as described (*Section 4.2.1*), and the promoter category is indicated by '1'.

The three right-hand columns show if a TSS is present (indicated by '1') within the regions bound by RNA polymerase (RNAP), RNA polymerase + Rifampicin (RNAP+Rif) and Binding factor Sigma70 (RpoD) using ChIP-chip (*Section 4.2.6*).

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	RNAP + Rif	RNAP	RpoD
pSLT$^{4/74}$											
14	+	STM474_p1001	*finO*					1			
616	-	STM474_p1001	*finO*	1					1	1	1
1250	+	STM474_p1002	*traX*					1		1	1
3733	+	STM474_p1003	*traI*					1		1	1
6707	+	STM474_p1005		1					1	1	
10391	-	STM474_p1008	*traT*			1				1	
11280	-	STM474_p1009	*traS*	1					1	1	1
11381	-	STM474_p1009	*traS*		1				1	1	1
16225	-	STM474_p1013	*traQ*	1							
31518		STM474_p1035	*traY*			1					
32265	+	FinP	FinP					1			
32369	-	STM474_p1036	*traJ*	1							
33234	+	STM474_p1038		1						1	1
33360	+	STM474_p1038				1				1	1
36490	-	STM474_p1040	*psiB*			1					1
38566	+	STM474_p1041						1	1	1	1
40507	-	STM474_p1044		1						1	1
46037	+	STM474_p1053	*samA*	1						1	1
49333	-	STM474_p1055	*parB*	1						1	1
50044	-	STM474_p1056	*parA*	1							

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	RNAP + Rif	RNAP	RpoD
50276	+	STM474_p1058			1						1
50415	+	STM474_p1058		1						1	1
51091	+	STM474_p1058				1			1	1	1
54196	-	STM474_p1063	*tlpA*	1							1
54404	+	STM474_p1064		1					1	1	1
54672	+	STM474_p1064				1				1	
55539	+	STM474_p1066		1						1	1
56487	+	STM474_p1067	*rlgA*	1						1	1
60912	+	STM474_p1070						1			
61906	-	STM474_p1072	*spvR*					1			1
62794	-	STM474_p1073	*spvA*					1			
68021	-	STM474_p1081		1						1	1
68060	-	STM474_p1081			1					1	1
68152	+	STM474_p1082		1							
73777	-	STM474_p1094	*ccdA*	1							
74090	-	STM474_p1094	*ccdA*		1						1
74408	+	STM474_p1095		1							1
75958	-	STM474_p1098	*repA2*					1		1	1
76024	+	STM474_p1098	*repA2*		1					1	1
76065	+	STM474_p1098	*repA2*	1						1	1
77167	+	STM474_p1098	*repA2*			1				1	1
78104	-	STM474_p1098A				1					
79600	-	STM474_p1100	*pefB*					1			1
84434	+	STM474_p1104	*orf6*			1				1	1
84664	+	STM474_p1104	*orf6*			1				1	1
84793	+	STM474_p1104	*orf6*			1				1	1
85731	-	STM474_p1105	*pefI*					1			

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	RNAP + Rif	RNAP	RpoD
86331	-	STM474_p1106	*orf7*					1		1	1
87460	-	STM474_p1108	*srgB*					1			1
87967	-	STM474_p1108	*srgB*					1			
88278	-	STM474_p1109	*rck*					1			
88905	-	STM474_p1109	*rck*					1		1	1
89799	-	STM474_p1110	*srgC*					1		1	1
89867	+	STM474_p1110	*srgC*			1				1	1
92010	+	STM474_p1114A	*repA3*					1	1	1	1
92179	-	STM474_p1114A	*repA3*	1					1	1	1
92409	+	STM474_p1115	*repC*					1	1	1	1
92519	-	STM474_p1115	*repC*	1					1	1	1
pCol1B$^{4/74}$											
155	+	STM474_p201	*repY*	1					1	1	1
365	-	STM474_p201	*repY*					1	1	1	1
2370	+	STM474_p202	*yacA*	1					1	1	1
2915	+	STM474_p204	*yacC*	1						1	1
3657	+	STM474_p204	*yacC*			1				1	1
3921	+	STM474_p205	*yadA*	1						1	1
4157	+	STM474_p205	*yadA*			1			1	1	1
4526	+	STM474_p206	*yaeA*	1					1	1	1
4752	-	STM474_p206	*yaeA*					1	1	1	1
6340	-	STM474_p209	*yafB*	1					1		
6634	+	STM474_p210	*yagA*	1							
7482	+	STM474_p210	*yagA*			1					
8227	+	STM474_p211	*cib*	1							
10424	-	STM474_p212	*imm*			1					1

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	RNAP + Rif	RNAP	RpoD
12332	+	STM474_p214	*ybfA*	1							
13965	-	STM474_p215A						1			1
14146	+	STM474_p216	*parA*	1							
14595	+	STM474_p216	*parA*			1					
15338	+	STM474_p218	*parM*	1							
15515	+	STM474_p218	*parM*			1					
17134	-	STM474_p219A	*dinI-like*	1							
24132	-	STM474_p229A	*ycjA*					1	1	1	1
27906	+	STM474_p233	*yddA*	1						1	1
30109	-	STM474_p236	*ydfB*					1	1	1	1
30412	+	STM474_p237		1					1	1	1
30895	+	STM474_p237	*ydgA*			1			1	1	1
32219	-	STM474_p238A	*ydhA*					1		1	1
33282	+	STM474_p239	*ydiA*			1				1	1
33763	-	STM474_p240		1					1	1	1
33868	+	STM474_p240A	*nikA*	1					1	1	1
40030	+	STM474_p243	*trbB*					1		1	1
41736	+	STM474_p245	*pndC*	1					1	1	1
41872	-	STM474_pndB	PndB					1	1	1	1
42429	-	STM474_p247		1					1	1	1
42636	-	STM474_p248						1	1	1	1
42698	+	STM474_p248		-		1			1	1	1
43554	+	STM474_p249		1						1	1
44588	-	STM474_p250	*excA*	1					1		
44622	-	STM474_p250	*excA*		1				1		
45033	-	STM474_p250A	*traY*			1			1	1	
46727	+	STM474_p250A	*traY*					1			1

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	RNAP + Rif	RNAP	RpoD
63809	+	STM474_p266	*traJ*					1		1	1
65701	+	STM474_p268	*traH*					1	1	1	1
65966	-	STM474_p268	*traH*	1					1	1	1
68796	-	STM474_p271	*traE*	1							
83463	-	STM474_p286	*traD*			1				1	1
83475	+	STM474_p286	*traD*					1		1	1
86343	-	STM474_p289	*traA*	1					1	1	1
86485	-	STM474_p290		1					1	1	1
86612	-	STM474_p290			1				1	1	1
pRSF1010[4/74]											
864	+	STM474_p301	*sul2*					1	1	1	1
871	-	STM474_p301	*sul2*	1					1	1	1
947	-	STM474_p301	*sul2*		1				1	1	1
951	-	STM474_p301	*sul2*		1				1	1	1
1570	+	STM474_p302	*repC*					1		1	1
2822	+	STM474_p303	*repA*					1			1
3300	-	STM474_p305		1							
5617	-	STM474_p308	*mobC*					1			1
5641	+	STM474_p308	*mobC*	1							1
6239	+	STM474_p309		1							1
8630	+	STM474_p311	*strA*			1				1	

Table S4.2: Transcriptional start sites in *S*. Typhimurium D23580 pSLT-BT^{D23580} under ESP growth condition.

Promoter category is indicated by '1'.

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense
1042	-	pSLTBT_p001	*srgC*			1		
1110	+	pSLTBT_p001	*srgC*					1
1980	+	pSLTBT_p002	*rck*					1
2607	+	pSLTBT_p002	*rck*					1
2918	+	pSLTBT_p003	*srgB*					1
3425	+	pSLTBT_p003	*srgB*					1
4554	+	pSLTBT_p005	*orf7*					1
5154	+	pSLTBT_p006	*pefI*					1
6092	-	pSLTBT_p007	*orf6*			1		
6221	-	pSLTBT_p007	*orf6*			1		
6448	-	pSLTBT_p007	*orf6*			1		
6451	-	pSLTBT_p007	*orf6*			1		
11206	-	pSLTBT_p011	*pefA*	1				
11285	+	pSLTBT_p012	*pefB*					1
11769	-	pSLTBT_p012	*pefB*	1				
12781	+						1	
13718	-	pSLTBT_p013	*repA2*			1		
14820	-	pSLTBT_p013	*repA2*	1				
14861	-	pSLTBT_p013	*repA2*		1			
14927	+	pSLTBT_p013	*repA2*					1
15664	+						1	
16176	-	pSLTBT_p014				1		
16188	+	pSLTBT_p014						1
16477	-	pSLTBT_p014		1				

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense
16795	+	pSLTBT_p014A	*ccdA*		1			
17108	+	pSLTBT_p014A	*ccdA*	1				
22724	-	pSLTBT_p021		1				
22816	+	pSLTBT_p022			1			
22855	+	pSLTBT_p022		1				
28082	+	pSLTBT_p026	*spvA*					1
28970	+	pSLTBT_p027	*spvR*					1
29964	-	pSLTBT_p028						1
31055	-	pSLTBT_p028						1
36891	+	pSLTBT_p034	*sulII*	1				
36974	-	pSLTBT_p034	*sulII*					1
39543	-	pSLTBT_p037	*int*			1		
40522	+	pSLTBT_p038	*dhfrI*	1				-
40728	-	pSLTBT_p037	*int*	1				
41192	-	pSLTBT_p038	*dhfrI*	-				1
41214	+	pSLTBT_p039	*aadA1*	1		-		
42258	+	pSLTBT_p040	*qacE*	1				
46541	-	pSLTBT_p044	*istA*			1		
46639	-	pSLTBT_p044	*istA*	1				
47997	-	pSLTBT_p047						1
48476	+	pSLTBT_p048	*insB3*			1		-
49482	+	pSLTBT_p050	*cat*	1				
54822	-	pSLTBT_p053	*tnpB*	1				
55754	-	pSLTBT_p054	*blaT*	1				
55879	+	pSLTBT_p055						1
57144	-	pSLTBT_p056		1				
58092	-	pSLTBT_p057		1				

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense
58142		pSLTBT_p075			1			
58959	-	pSLTBT_p058				1		
59227	-	pSLTBT_p058		1				
59435	+	pSLTBT_p059	*tlpA*	1				
62540	-	pSLTBT_p063				1		
63216	-	pSLTBT_p063		1				
63355	-	pSLTBT_p063			1			
63586	+	pSLTBT_p064	*parA*	1				
64298	+	pSLTBT_p065	*parB*	1		-		
67594	-	pSLTBT_p067	*samA*	1				
73124	+	pSLTBT_p075		1				
75076	-	pSLTBT_p078						1
77153	+	pSLTBT_p079	*psiB*			1		
80282	-	pSLTBT_p081				1		
80408	-	pSLTBT_p081		1				
81274	+	pSLTBT_p083	*traJ*	1				
81378	-	FinP	FinP					1
82125	+	pSLTBT_p085	*traA*	1		-		
97418	+	pSLTBT_p106	*traQ*	1				
102262	+	pSLTBT_p110	*traS*		1			
102363	+	pSLTBT_p110	*traS*	1				
102311	-	pSLTBT_p109	*traG*					1
102343	-	pSLTBT_p109	*traG*					1
103252	+	pSLTBT_p111	*traT*			1		
106936	-	pSLTBT_p113B		1				
109910	-	pSLTBT_p114	*traI*					1
112393	-	pSLTBT_p115	*traX*					1

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense
113027	+	pSLTBT_p116	*finO*	1				
113629	-	pSLTBT_p116	*finO*					1
114965	+	pSLTBT_p119	*repC*	1				
115074	-	pSLTBT_p119	*repC*					1
115305	+	pSLTBT_p119A	*repA3*	1				
115454	+	pSLTBT_p119B	*tap*	1				
115474	-	pSLTBT_p119A	*repA3*					1

Table S4.3: List of asRNA candidates identified in *S*. Typhimurium 4/74 plasmids at ESP.

* The length has been determined from the visualisation of RNA-Seq data in the IGB browser.

Gene expression levels were calculated using the TPM approach as described (*Section 3.2.2.6*).

The asRNAs were classified into two categories based on their expression levels compared to sense mRNA (based on TPM values). Class I: asRNA showed higher expression than the sense mRNA; class II: asRNA and mRNA showed similar expression levels.

The asRNAs show in **red** were validated by Northern blot.

asRNA ID	Strand	Start	End	Length*	Antisense to	Class	Location relative to coding gene	Expression values (asRNA / mRNA)
pSLT$^{4/74}$								
474P1AS_10	+	14	96	83	*finO*	2	3'	16 / 16.6
474P1AS_30	+	3733	3809	77	*traI*	1	Internal	1061.4 / 5.2
FinP	+	32266	32341	76	*traJ*	1	5'	1797.4 / 0.9
474P1AS_50	+	38566	38637	72	STM474_p1041	1	5'	5194.4 / 0.6
474P1AS_60	+	60912	60980	69	STM474_p1070	1	5'	34.7 / 1.8
474P1AS_70	-	61830	61906	77	*spvR*	1	Internal	1741 / 1
474P1AS_80	-	62679	62794	116	*spvA*	1	5'	42.8 / 1.5
474P1AS_90	-	75861	75958	98	*repA2*	2	5'	42.7 / 88.9
474P1AS_100	-	79507	79600	94	*pefB*	1	3'	51.4 / 1.5
474P1AS_110	-	85661	85731	71	*pefI*	1	5'	89.1 / 17.2
474P1AS_130	-	87387	87460	74	*srgB*	1	5'	38.5 / 3.2
474P1AS_140	-	87918	87967	50	*srgB*	1	Internal	39.3 / 3.2
474P1AS_150	-	88177	88278	102	*rck*	1	5'	31.5 / 8
474P1AS_160	-	88852	88905	54	*rck*	1	3'	43.6 / 8
474P1AS_170	-	89696	89799	104	*srgC*	2	Internal	35.2 / 14.7
474P1AS_180	+	92010	92103	94	*repA3*	2	Internal	10561.3 / 502.1
474P1AS_190	+	92409	92492	84	*repC*	2	5'	74.3 / 399.5

Continued on next page

asRNA ID	Strand	Start	End	Length*	Antisense to	Class	Location relative to coding gene	Expression values (asRNA / mRNA)
pCol1B$^{4/74}$								
474P2AS_10	-	296	365	70	*repY*	2	5'	74863 / 196.1
474P2AS_20	-	4693	4752	60	*yaeA*	2	3'	115.9 / 113.1
474P2AS_30	-	13885	13965	81	STM474_215A	2	5'	73.8 / 16.6
474P2AS_40	-	24026	24132	107	*ycjA*	1	5'	1222.9 / 1.4
474P2AS_50	-	30007	30109	103	*ydfB*	1	5'	4134 / 2
474P2AS_60	+	40030	40102	73	*trbB*	1	Internal	42.4 / 3
PndB	-	41811	41872	62	*pndC*	2	5'	13651.7 / 1180.4
474P2AS_80	-	42547	42636	90	STM474_p248	2	5'	18546.6 / 7065.5
474P2AS_90	+	46727	46777	51	*traY*	1	5'	2075.5 / 8
474P2AS_100	+	63809	63878	70	*traJ*	1	Internal	54.2 / 5.1
474P2AS_110	+	65701	65760	60	*traH*	1	5'	62.4 / 6
474P2AS_120	+	83475	83575	101	*traD*	2	3'	80.2 / 20.3
pRSF1010$^{4/74}$								
474P3AS_10	+	864	1500	637	*sul2*	2	5'	576.8 / 485
474P3AS _20	+	1570	1632	63	*repC*	1	Internal	527.4 / 8.6
474P3AS _30	+	2822	2898	77	*repA*	2	5'	898.1 / 30.6
474P3AS _40	-	5514	5617	104	*mobC*	1	5'	164.5 / 36.9

Table S4.4: Candidate asRNAs identified in *S*. Typhimurium D23580 plasmid pSLT-BT.

The **blue** coloured asRNAs were identified within the multidrug resistance (MDR) locus. Other details as in the legend to Table S4.3

asRNA ID	Strand	Start	End	Length	Antisense to	Class	Location relative to coding gene	Expression values (asRNA / mRNA)
D23P1AS_10	+	1110	1240	131	*srgC*	1	Internal	30.1 / 8.5
D23P1AS_20	+	1980	2039	60	*rck*	1	3'	**14.5 / 6.3**
D23P1AS_30	+	2607	2686	80	*rck*	1	5'	17.7 / 6.3
D23P1AS_40	+	2918	2996	79	*srgB*	1	Internal	19.9 / 2.9
D23P1AS_50	+	3425	3512	88	*srgB*	1	5'	13.7 / 2.9
D23P1AS_70	+	5154	5279	126	*pefI*	1	5'	31.6 / 8.6
D23P1AS_80	+	11285	11389	105	*pefB*	1	3'	24.2 / 2.9
D23P1AS_90	+	14927	15073	147	*repA2*	2	5'	28 / 77
D23P1AS_110	+	28082	28198	117	*spvA*	1	5'	22.7 / 3.4
D23P1AS_120	+	28970	29057	88	*spvR*	1	Internal	13.3 / 1.2
D23P1AS_130	-	29964	29896	69	pSLTBT_p028	1	5'	1003.3 / 2.4
D23P1AS_140	-	31055	30850	206	pSLTBT_p028	1	3'	20.9 / 2.4
D23P1AS_150	-	36974	36341	634	*sulII*	2	5'	45.1 / 158.5
D23P1AS_160	-	41192	40867	326	*dhfrI*	2	3'	23.9 / 132
D23P1AS_170	-	47997	47741	257	pSLTBT_p047	1	5'	22.7 / 8.7
D23P1AS_180	+	55879	56048	170	pSLTBT_p055	2	3'	13.6 / 44.9
D23P1AS_190	-	75076	74877	200	pSLTBT_p078	1	5'	1936.9 / 0.8
FinP	-	81378	81302	77	*traJ*	1	5'	1249.4 / 1.2
D23P1AS_210	-	102311	102130	182	*traG, traS*	2	5'	10.7 / 147
D23P1AS_220	-	109910	109736	175	*traI*	1	Internal	208.8 / 4.8
D23P1AS_240	-	113629	113555	75	*finO*	2	3'	17.9 / 12.6
D23P1AS_250	-	115074	114958	117	*repC*	2	5'	33.2 / 229.8
D23P1AS_260	-	115474	115381	94	*repA3*	2	Internal	7621.9 / 35.8

Table S4.5: List of gene with their correct co-ordinates in *S*. Typhimurium 4/74 plasmids.

Blue: Annotation of these genes was corrected from the using RNA-Seq.

Red: Newly identified genes by using RNA-Seq technique.

Black: List of genes introduced into the 4/74 plasmid annotation following a comparison with the published *S*. Typhimurium SL1344 annotation (Kröger et al. 2012).

Systemic ID	Common name	Strand	Start	End	Length	Expressed at ESP
pSLT$^{4/74}$						
STM474_p1046A		-	41666	42115	450	no
STM474_p1062	*yacC*	-	52553	52969	417	no
STM474_p1067	*rlgA*	+	56519	57079	561	yes
STM474_p1098	*repA2*	+	76217	77206	990	yes
STM474_p1098A		-	78078	78287	210	yes
STM474_p1114A	*repA3*	-	91997	92119	123	yes
pCol1B$^{4/74}$						
STM474_p201A	*repZ*	+	455	1486	1032	no
STM474_p210	*yagA*	+	6669	8015	1347	yes
STM474_p215A		+	13893	14096	204	no
STM474_p219A	*DinI*-like	-	16922	17101	180	yes
STM474_p219B	*yccA*	+	17236	17481	246	no
STM474_p223A	*ycfA*	+	20205	20981	777	no
STM474_p223B	*ycgA*	+	20956	21426	471	no
STM474_p224	*ycgB*	+	21399	21824	426	no
STM474_p224A	*ycgC*	+	21871	22293	423	no
STM474_p229A	*ycjA*	+	24121	26085	1965	no
STM474_p238A	*ydhA*	+	32007	32342	336	no
STM474_p240A	*nikA*	+	33401	33736	336	yes
STM474_p250A	*traY*	-	44633	46864	2232	yes
STM474_p263A	*sogL*	-	58484	62251	3768	no
STM474_p272A	*shfB*	+	70132	70383	252	no
STM474_p272B		-	70385	70630	246	no
STM474_p272C	*shfC*	+	70653	71006	354	yes
STM474_p272D	*pilV*	-	71003	72295	1293	no

Table S4.6: Nine genes identified in the *S*. Typhimurium D23580 plasmid pSLT-BT^{D23580} by comparative annotation methods.

Systemic ID	Common name	Strand	Start	End	Length	Expressed at ESP
pSLTBT_p013A		-	15162	15458	297	yes
pSLTBT_p014A	*ccdA*	+	17139	17357	219	yes
pSLTBT_p014B	*ccdB*	+	17359	17664	306	yes
pSLTBT_p014C		+	17666	17956	291	yes
pSLTBT_p113A		-	106279	106677	399	yes
pSLTBT_p113B		-	106677	106907	231	yes
pSLTBT_p118A		+	115365	115487	123	yes
pSLTBT_p118B		+	115492	115569	78	yes
pSLTBT_p120A		+	116443	116658	216	no

Listed genes were introduced into the pSLT-BT^{D23580} annotation following comparison with the published *S*. Typhimurium SL1344 annotation (Kröger et al. 2012).

CHAPTER 5

INVESTIGATING THE TRANSCRIPTOMIC PROFILE OF PLASMID-ENCODED GENES OF *SALMONELLA*

5.1 *Introduction*

Transcriptomic analysis is widely used to unravel regulatory mechanisms that control cellular processes in cells. Microarrays have been the classical investigating tool of choice for large-scale studies of gene expression levels, although this hybridization-based technology is largely restricted to known genes, and has a limited dynamic range (Allison et al. 2006). In recent years, RNA-Seq (RNA-sequencing) has emerged as an alternative to microarrays for measuring the global expression of genes at a single base resolution. Several comparative studies of RNA-Seq and microarray data demonstrated a general concordance between these two methods (Bradford et al. 2010; Agarwal et al. 2010; Mortazavi et al. 2008). These comparisons also revealed that sequencing-based approaches are more sensitive and offer a larger dynamic range of quantification and reduced technical variability compared with microarray (Mariono et al. 2008). RNA-Seq is an accurate, strand-specific method that extends the possibilities of transcriptome studies to the analysis of previously unidentified genes and antisense transcripts, 5' untranslated regions and can be used to define operon relationships (Sharma et al. 2010; Kröger et al. 2012). These advantages, coupled with the decreasing cost of sequencing, make RNA-Seq an increasingly attractive method for whole-genome expression studies in many biological systems.

In this study, we used a suite of RNA-Seq data to generate a comprehensive picture of differential gene expression of both coding and non-coding genes encoded by the three plasmids of S. Typhimurium 4/74, namely pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$. The key findings are that the expression profile *spv* locus showed similarities to SPI-2 genes, and *strAB* genes showed a similar expression profile to the *rpsM* gene. We analysed the relationship between SPI-2 and the *spv* locus at the regulatory levels. Other interesting findings are the similarities in expression profile between plasmid and chromosomal-encoded genes and some potential cross-talk between chromosome and plasmids at the regulatory levels.

5.2 *Results and discussion*

5.2.1 *RNA-Seq and dRNA-Seq libraries used in this study*

S. Typhimurium is an adaptable micro-organism that thrives in a variety of locations, including the gastrointestinal tract of mammals, and the intracellular milieu of eukaryotic cells. The ability to survive within an infected host and fast adaptability of its transcriptional program when it encounters a new environment (Rolfe et al. 2012) makes *Salmonella* a successful pathogen. It is clear from the literature that the gene expression of *Salmonella* is highly sensitive to environmental changes, particularly when the bacteria infect the mammalian host (Hébrard et al. 2011). To determine and analyse the transcriptional expression profiles of *Salmonella* plasmid-encoded genes during the process of infection, we used 29 RNA-Seq datasets of strain 4/74 generated from 22 *in vitro* infection-relevant growth situations. These situations represent aspects of different environments encountered by *Salmonella* during pathogenesis. The details of these growth conditions are given in Supplementary Dataset 3.1 (except InSPI1 and MMA conditions). RNA-Seq libraries (two biological replicates) generated from *Salmonella* grown within murine macrophages (MAC) were also included here to show the expression of plasmid-encoded genes during infection of mammalian cells (*in vivo*). These RNA-Seq data were generated by Dr. Carsten Kröger and Dr. Magali Hébrard, Hinton lab, Trinity College Dublin, Ireland.

Salmonella inhabit various ecological niches and have evolved multiple strategies to adapt their metabolism in response to internal and external signals. Transcriptional regulation plays a prominent role in adaptation, because it controls the cellular response to environmental changes through the activity of DNA-binding regulatory proteins. To determine the role of chromosomal regulatory proteins on plasmids-encoded genes, we used a further 18 RNA-Seq libraries generated from 17 mutant stains that lacked particular regulatory proteins including global regulators (Dam and Fur), transcription factors (RpoE and RpoS) SPI-1 associated regulators (HilD, FliZ, BarA/SirA, and Pnp), and the SPI-2 associated transcription regulators (SlyA, SsrA, SsrB, PhoP/Q, EnvZ/OmpR). The mutants and growth conditions used for these experiments are listed in Table S5.1. These RNA-Seq data were generated as by Ms. Aoife Colgan. I processed the data, including mapping and calculating gene expression values (TPM), and analysed the plasmid results.

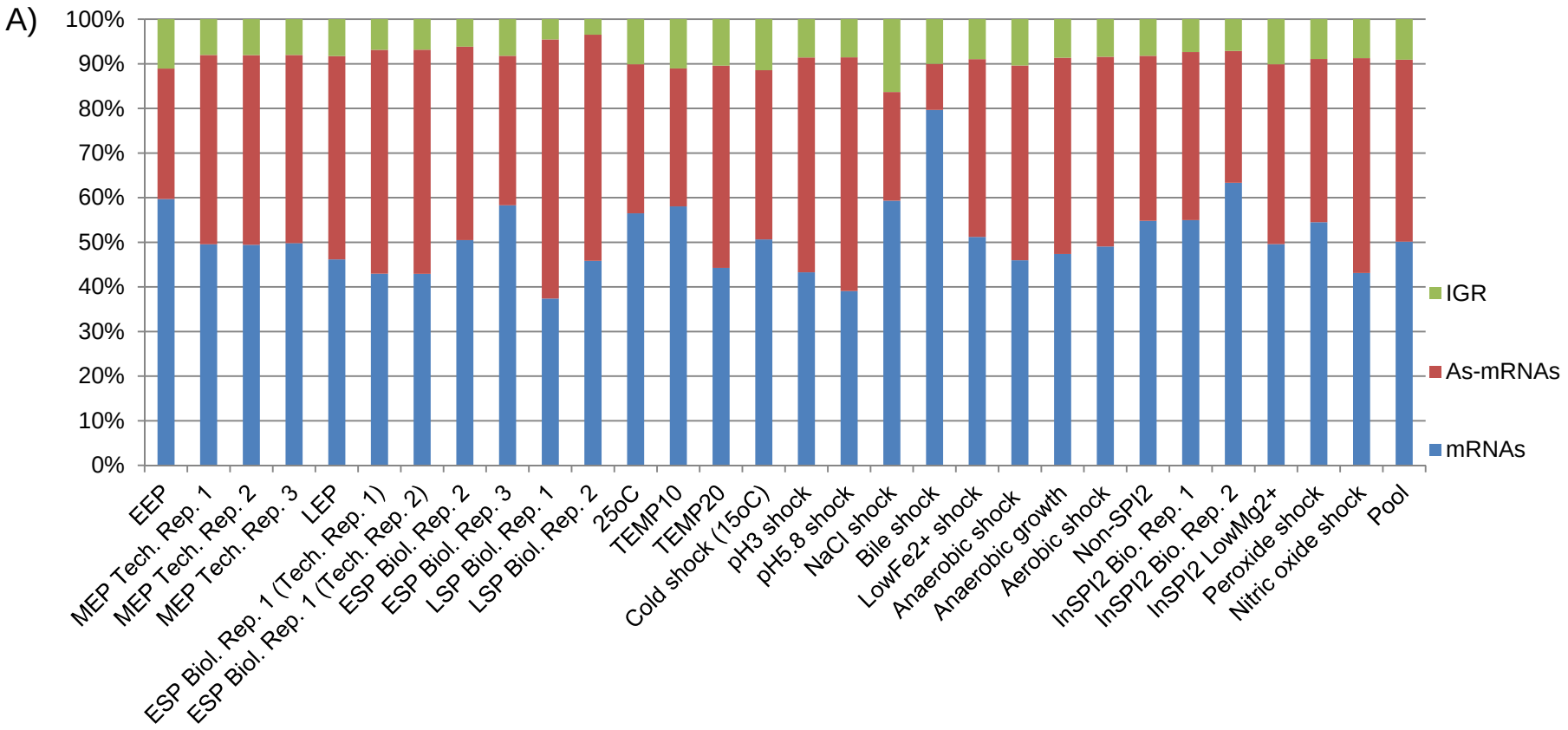
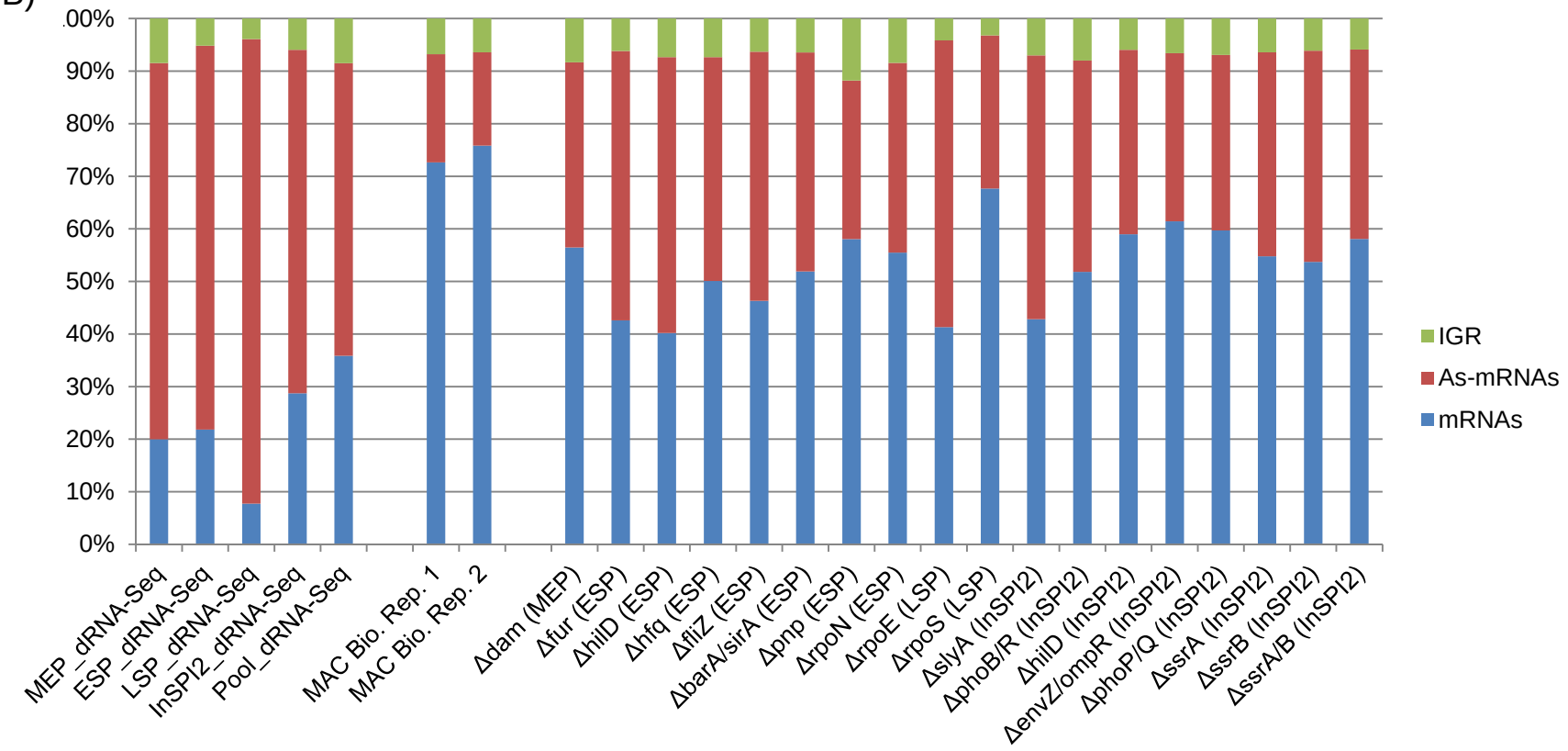


Figure 5.1: List of RNA-Seq and dRNA-Seq libraries Distribution of uniquely mapped cDNA reads to different locations on the plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.

A and B) The relative levels of uniquely mapped cDNA reads mapped to coding regions (mRNA), antisense to coding regions (As-mRNA) and intergenic regions (IGR) are given here. All RNA-Seq and dRNA-Seq libraries used in this study are listed here.

Continued on next page

B)



Total RNA extraction from different growth conditions was done by Dr. Carsten Kröger, Ms. Aoife Colgan and Dr. Magali Hébrard, Hinton lab, Trinity College Dublin (for detail see Supplementary Dataset 3.2). I mapped the cDNA reads to *Salmonella* genome and analysed the cDNA reads that mapped to *Salmonella* plasmids alone. Read distribution was calculated as described in *Section 3.2.2.4*.

The accurate identification of promoter regions and TSS is an important step for understanding the mechanisms of transcription regulation. Previously, we identified 120 TSS in three plasmids of *Salmonella* using three biological replicates of dRNA-Seq libraries at the ESP condition (*Section 4.2.1*). However, promoters of plasmid virulence genes, including *spv* and *pef* loci, were not identified and only 40% of plasmid genes were expressed at ESP. To identify the promoters of plasmid-encoded genes that were not expressed at ESP, we used 5 dRNA-Seq libraries from different growth conditions. Three libraries were generated from three distinct growth conditions in rich medium (MEP, ESP [biological replicate 4] and LSP; LB broth) and a fourth was produced from InSPI2 (*Salmonella* Pathogenicity Island-2 inducing) condition. The fifth dRNA-Seq library was constructed from an RNA Pool (see detail below).

A total of 50 RNA-Seq and 5 dRNA-Seq libraries were used in this study. The distribution of cDNA reads mapped uniquely to different regions of the plasmids, including ORFs, antisense to ORFs and IGRs are shown in Figure 5.1A and 5.1B.

The relative proportion of cDNA reads that mapped to each region (ORFs, antisense to ORFs and IGRs) varied depends on growth conditions. For example, the proportions of cDNA reads that mapped antisense to ORFs were increased from EEP to LSP growth conditions. A small proportion (10%) of antisense transcripts was observed in Bile shock and the intracellular condition (MAC – 18%). In contrast, a maximum of 54% of the reads were antisense during growth in LSP. The proportion of cDNA reads that mapped to the coding regions (ORFs) varied from 38% to 80%. Approximately 8-12% of the cDNA reads mapped to the IGRs of the *Salmonella* plasmids.

5.2.2 *Identification of Transcriptional Start Sites*

To identify TSS from five different environmental conditions, we used a rigorous approach (*Section 3.2.2.5*). This analysis increased the number of TSS for the *Salmonella* plasmids from 120 derived from one condition (ESP) to 162 (Table S5.2). All the TSS described in *Section 4.2.1* matched exactly with the ones reported here. Our dRNA-Seq library constructed from mimicking intracellular (InSPI2) conditions revealed two primary TSS within the *spv* locus (*spvR* & *A*; Figure 5.2). The *spv* locus consists of the regulatory gene *spvR* followed by four virulence genes, *spvA, B, C & D,* which are expressed as a polycistronic operon. The primary TSS of *spvR* was found 17

nt upstream of the *spvR* start codon, which matches exactly with the TSS identified by Kowarz et al. (1996). In contrast, for the *spvA* gene, we observed a 102 nt long 5' UTR. The asRNA 474P1AS_80 and the gene *spvA* were expressed divergently from a single promoter region.

Of the 162 TSS, 59 were classified as primary and 16 as secondary promoters (Figure 5.3A). Exactly 50% of the TSS were categorized as antisense and internal promoters. Interestingly, our TSS analysis identified four orphan promoters within the *Salmonella* plasmids. ORF prediction analysis within these orphan regions did not identify any ORFs downstream of the TSS, which may suggest that these RNA transcripts have a regulatory role. Further experiments are needed to investigate whether these transcripts have a biological function and whether they should be annotated as new genes.

We observed that the promoters of certain genes were activated at particular environmental conditions, for example, the *imm* gene was expressed at MEP and LSP. To determine if the promoter was active in a particular growth condition, a set of TPM values were calculated for the first ten nucleotides of each transcript extracted from individual RNA-Seq experiments (MEP, ESP, LSP and InSPI2). A promoter was only regarded as activated when the TPM value was ≥10 (Table S5.2; Kröger et al. 2013). This analysis revealed that 118 TSS (~73% of all identified promoters) were initiated at all four growth conditions (Figure 5.3B), which include the promoters of plasmid structural genes involved in plasmid replication, partitioning, mobilization and stabilization. The promoters of auxiliary genes which encode host-beneficial traits including *spvR & A* (InSPI2) and *pefB* (MEP) were expressed in specific growth conditions. The majority of antisense promoters (~67% of asTSS) were activated at all four growth conditions, including promoters antisense to *traJ, traI, pefI, srgB, rck* and *sul2*. In some cases, the expression of asTSS in a single growth condition was also observed, for example, an asTSS opposite to *nuc* (InSPI2), *nikB* (ESP) and *ccdA* (LSP) genes (see Table S5.2 for details).

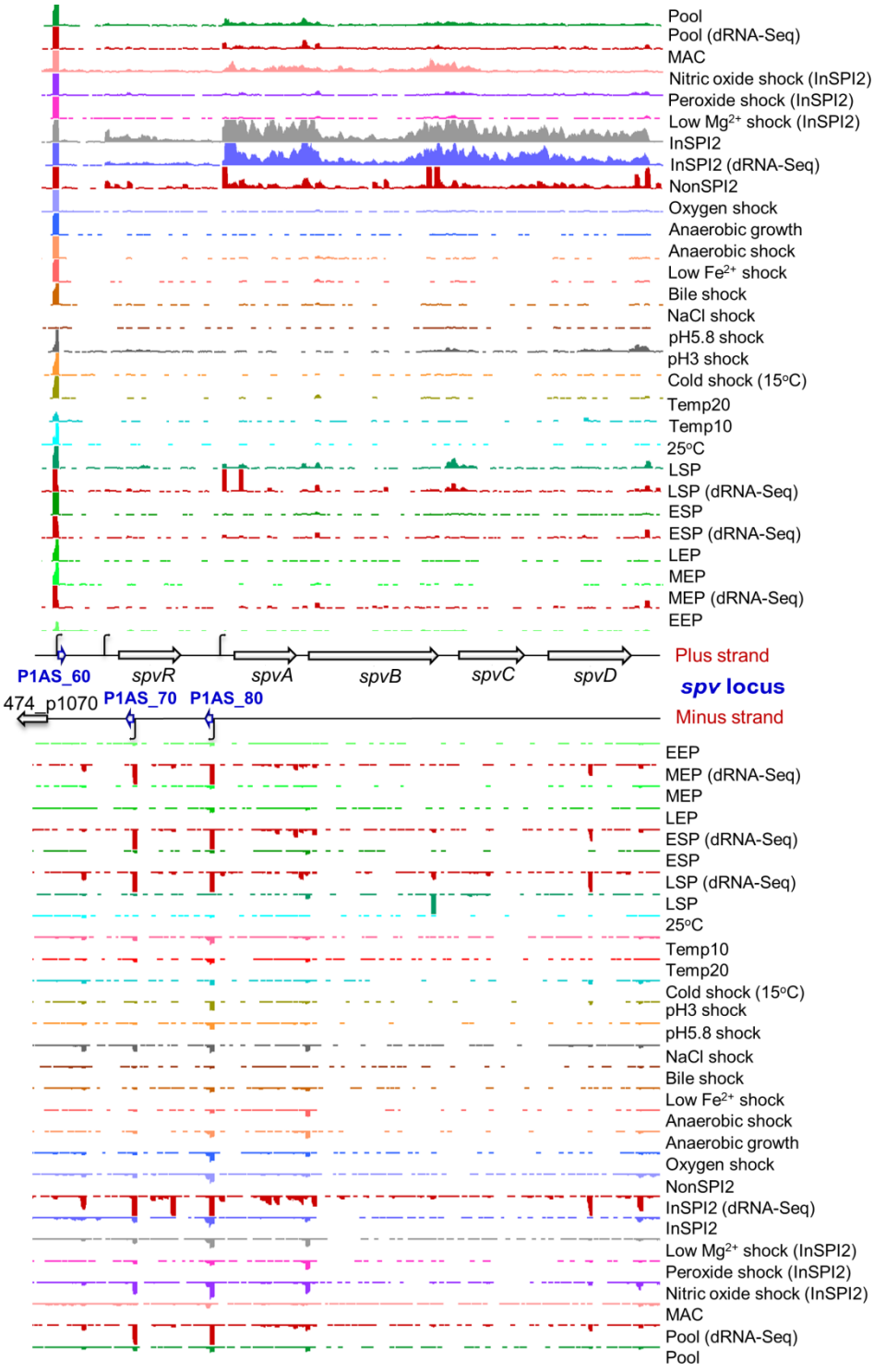


Figure 5.2: Visualization of mapped cDNA reads in the *spv* locus of pSLT$^{4/74}$.
Protein coding gene names are labelled with black, asRNA gene names in blue and TSS are shown by curved arrows. The vertical scale is 0-100 normalized sequenced reads for each library. The *spv* locus is only expressed in the InSPI2 condition and inside macrophages. The growth conditions are detailed in Table 2.3 (*Section 2.1.3*)

The dRNA-Seq and RNA-Seq libraries of a "pooled" RNA sample (RNA Pool) were constructed by combining the total RNA extracted from 22 environmental conditions

(Table 2.3, except MAC; *Section 2.1.3*) prior to RNA-Seq, to generate a complete picture of the *Salmonella* transcriptome under these conditions (Kröger et al. 2013). In Kröger et al. (2013), we established the value of the RNA Pool for the identification of TSS of chromosomal genes. We showed that the pooling of RNA from multiple conditions allowed the reliable identification of promoter regions for the vast majority (96% of all identified TSS) of chromosomal genes. To determine the success rate of TSS identification in the RNA Pool sample for the *Salmonella* plasmids, we compared the TSS found in other growth conditions with the TSS of the RNA Pool. This revealed that ~98% of plasmid-associated TSS (158 of 162) could be identified in the RNA Pool. Interestingly, the TSS of hypothetical protein STM474_p1058 (p1058) from pSLT$^{4/74}$ was only identified in the RNA Pool. A clear -10 box upstream of this TSS indicated that this TSS has been used as a secondary TSS for the transcription of the p1058 gene. In summary, the pooling of RNA was a successful strategy for the identification of the vast majority of promoter regions in *Salmonella* plasmids.

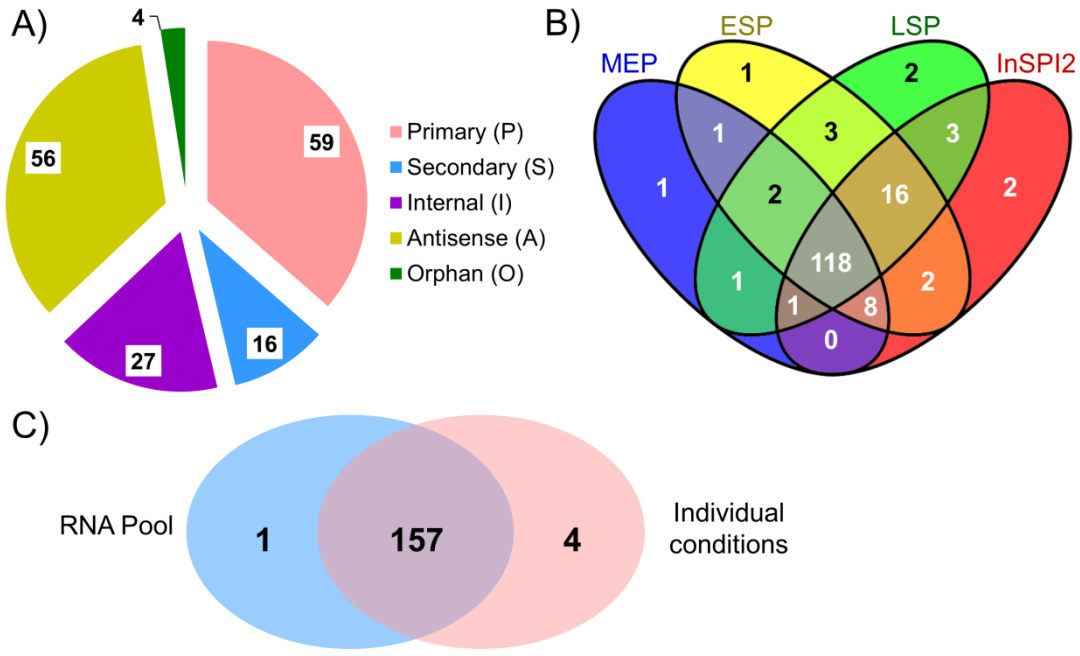


Figure 5.3: Identification and categorization of TSS from *Salmonella* plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.

A) Categorization of 162 TSS of the three plasmids identified in this study, of which 50% are antisense or internal relative to coding genes (Table S4.2). Explanation of the genomic context of the TSS localization is described in *Section 4.2.1*. B) Promoter distribution from four dRNA-Seq experiments. See also Table 5.1. C) The TSS identified from four environmental conditions were compared with the RNA Pool.

5.2.3 *The majority of core functional plasmid genes are expressed in most environmental conditions*

To relate mapped sequence reads to relative transcript abundance, the absolute levels of gene expression were quantified from the RNA-Seq data by the transcripts per million (TPM) measurement (*Section 3.2.2.6*) (Li et al. 2010, Wagner et al. 2012, 2013). We determined the number of plasmid-encoded genes that were expressed in at least one sample condition. A gene was only regarded as expressed when the TPM value was ≥10 (Supplementary Dataset 5.1; Kröger et al. 2013). On average, 43% of genes were expressed in each condition. This was slightly lower than the average expression of chromosomal-encoded genes expressed in a single growth condition, which was 63% (Kröger et al. 2013). Similar to the chromosome, plasmids also expressed a much lower proportion of genes at LSP (23%). Interestingly, the highest level of coding gene expression (~57%) was observed in the intracellular growth condition (MAC). A small set of genes were only expressed in the MAC condition, including 7 conjugal transfer (*tra*) genes, SOS inhibitors (*psiA* & *psiB*), resolvase (*resA*), and 4 hypothetical proteins. However, the expression level of these genes was extremely low (TPM value between 10 and 25).

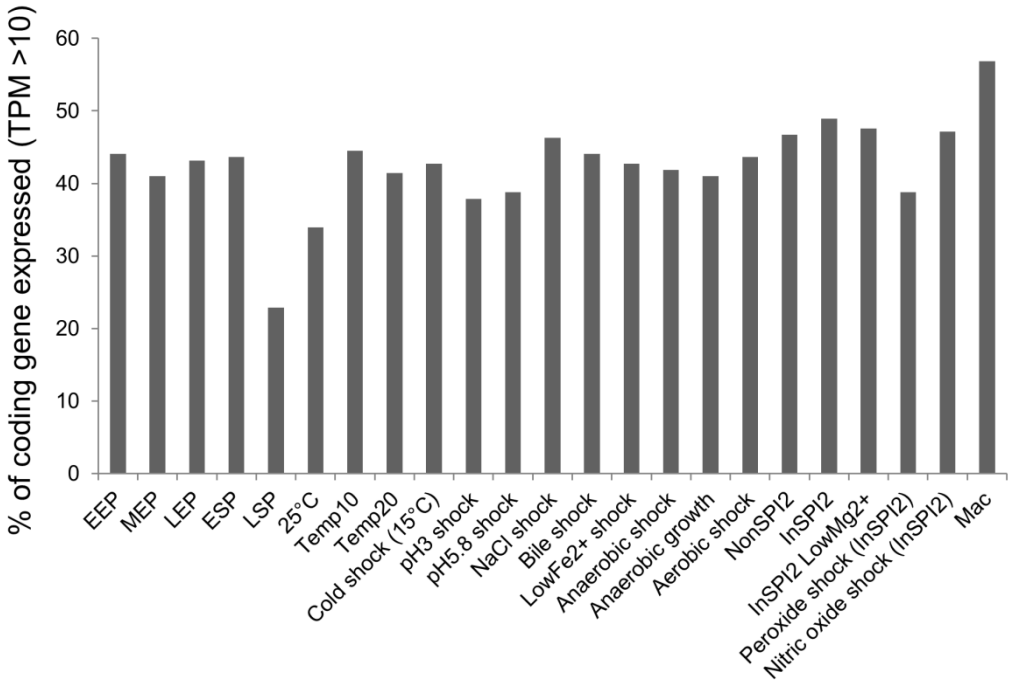


Figure 5.4: Expression analysis of coding genes of *Salmonella* plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.

Percentage of protein coding genes expressed in each environmental condition. See also Supplementary Dataset 5.1.

Approximately 68% of genes were expressed in at least one of the suite of 23 conditions. Of the 154 genes expressed in at least one of the environmental conditions, 92 have an annotated function. The genes that encode plasmid core functions including replication, maintenance and stability, were expressed in the majority of the environmental conditions. Further analysis revealed that the 16 genes which encode plasmid core functions were expressed in all 23 environmental conditions, including genes encoding the plasmid partitioning system (*parA/B* in pSLT$^{4/74}$ and *stbA/B* in pCol1B$^{4/74}$), maintenance (*ccdA/B* in pSLT$^{4/74}$ and *pndA/C* in pCol1B$^{4/74}$) and replication (*repA* and *repC*). Three genes that encode for non-core functions like drug resistance (*strA, strB* and *sul2*) were also expressed in all conditions showing that the transcription of these antibiotic resistance genes is not environmentally controlled. Of the 38 genes expressed in all conditions, 19 are functionally uncharacterised (FUN; Hinton 1997). However, expression of these FUN genes in all growth conditions raises the possibility that they encode novel plasmid core functions.

We identified 72 genes that were not expressed in the suite of growth conditions (TPM value <10; Supplementary Dataset 5.1). More than half of these genes (56%) are functionally characterised. Certain well-characterised genes, including the *pef* genes (*pefABCD*), were not expressed in any of the growth conditions. The CsrA-mediated activation of the *pef* genes is suppressed by the expression of chromosomal type I fimbrial genes (Sterzenbach et al. 2013), suggesting that this could be the reason for suppression of the *pef* genes in all conditions. Each environment generates a distinct transcriptional signature that reflects the mechanisms used by *Salmonella* to adapt to stressful environments. The genes that encode for host-beneficial traits were only expressed in a subset of certain conditions, including the up-regulation of the *spv* locus in SPI-2 inducing conditions.

In summary, the genes that encode plasmid core functions, including replication, partitioning and maintenance, were expressed in most growth conditions. This is because the expression of these genes is required by plasmids to replicate and partition in each bacterial generation. Even though conjugational genes are considered to be plasmid core functional genes, most of the *tra* and *pil* genes were not expressed in any of the growth and stress conditions. This could be because the expression of these genes is not required for plasmid maintenance in a homogeneous bacterial population. In the *tra* locus, the only gene that encodes for

serum resistance *traT* (Glöckner et al. 1997) and adjacent *traS* were expressed in most environmental conditions. We observed that the genes encoding host-beneficial traits were only expressed in growth conditions where they could affect the survival of the bacterium, such as the *spv* locus (see below).

Table 5.1: Number of plasmid-encoded CDS* that are >3-fold up- or down-regulated during each environmental condition.

Comparator conditions	Environmental test Conditions	Differentially expressed gene		
		Up	Down	Total
EEP	MEP	0	1	1
	LEP	1	0	1
	ESP	7	4	11
	LSP	4	38	42
MEP	Cold shock (15°C)	4	0	4
	pH3 shock	1	9	10
	pH5.8 shock	0	3	3
	NaCl shock	31	1	32
	Bile shock	3	3	6
	LowFe^{2+} shock	1	0	1
	Anaerobic shock	1	4	5
Anaerobic growth	Aerobic shock	3	0	3
InSPI2	InSPI2 LowMg^{2+}	1	2	3
	Peroxide shock	14	15	29
	Nitric oxide shock	4	6	10
NonSPI2	InSPI2	5	1	6
ESP	MAC	46	2	48
25°C	Temp10	21	0	21
	Temp20	5	0	5

* CDS encoded by plasmids $pSLT^{4/74}$, $pCol1B^{4/74}$ and $pRSF1010^{4/74}$.

To investigate transcriptional changes in the plasmids of *S*. Typhimurium, we determined the number of genes that were up- and down-regulated in each condition. The relative expression of a gene was calculated by comparing the TPM values of the gene between test and comparator conditions, as described (*Section 3.2.2.6*). Nineteen test conditions and their appropriate comparator conditions are listed in Table 5.1, which also shows the number of genes that were >3-fold up- and down-regulated during each condition. The largest changes in gene expression

occurred in the intracellular (MAC, 21%) growth condition (Supplementary Dataset S5.1). As expected (Guiney et al. 1995), the pSLT$^{4/74}$-encoded *spv* locus was upregulated in MAC (see below). Relative expression analyses revealed that 38 plasmid coding genes were downregulated in LSP, including genes involved in plasmid replication and conferring drug resistance (both sulphonamide and streptomycin). Among the different environmental shocks compared with MEP, the largest changes in gene differential expression occurred upon NaCl shock. Individual environments caused the up- or down-regulation of specific sets of genes. However, it was not possible to investigate the functional similarity between these genes due to low number of genes were differentially expressed in each growth condition.

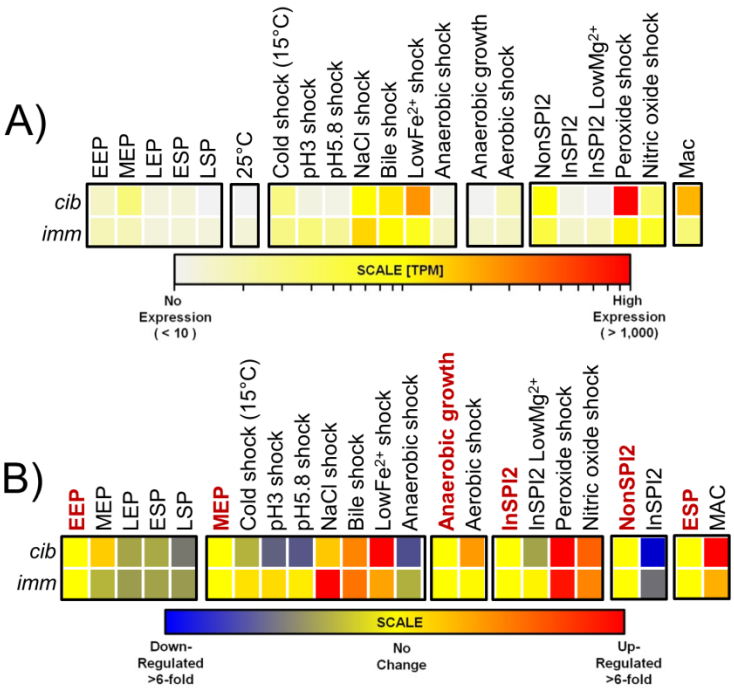


Figure 5.5: Absolute and relative expression of colicin toxin and immunity genes from pCol1B$^{4/74}$ plasmid.

The heat map shows the absolute (A) and relative (B) gene expression of the *cib* and *imm* genes from the colicinogenic plasmid of *Salmonella*. The abbreviations for each condition are shown in Table 2.3 (*Section 2.1.3*). Each gene is represented by a horizontal series of blocks, while individual environmental conditions are shown in vertical columns. The conditions highlighted in red (vertical columns) were used for normalisation against other growth, stress and infection-relevant condition to produce the profile of the relative expression pattern of individual genes.

Interestingly, we observed that each plasmid-encoded gene generated a unique signature in its expression profile across the 23 environmental conditions. Figure 5.5 depicts the signature expression profile of colicin producing gene (*cib*) and its immune protein gene (*imm*). The *cib* gene was significantly up-regulated in certain environmental shocks including Bile, Low-iron, nitric oxide and peroxide. As expected, the *imm* gene was upregulated in the same shocks as the *cib* gene, with minor differences. Overall, the expression patterns of these two genes showed strong similarities, reflecting their involvement in the same cellular process.

5.2.4 *Similarities in expression of partitioning system genes in plasmids pSLT$^{4/74}$ and pCol1B$^{4/74}$*

The differences between growth in various environments can be reflected in changes in particular cellular pathways. Underpinning these pathways are complex transcriptional networks that drive the expression of subsets of the genes to achieve a specialised function. The set of genes required for any cellular function must share transcriptional regulation, so that their products are available in the correct place at the right time. Based on this concept, a new approach has been coined by John Quackenbush (2003) known as 'Guilt by Association'. It hypothesised that the genes associated with particular biological function are co-expressed, or up- or down-regulated, in a similar way (co-expression). It has been predicted that conserved functions of groups of genes showed similar pattern of gene expression between species (Stuart et al. 2003).

To investigate the genes which have similar expression profiles, we determined the mutual relationship between expression patterns of plasmid-encoded genes by calculating the correlation coefficient using the procedure described in *Section 2.10*. Two different analyses were performed; first to identify the plasmid genes that shared an expression pattern within the same plasmid or between plasmids; and second to identify plasmid-encoded genes that shared an expression pattern with chromosomal genes. The relationship between profiles was considered to show co-expression when the correlation coefficient cut-off value was > 0.7. Relationships between many expression profiles generated by plasmid-encoded genes were identified (Table S5.3). The interesting findings include similarities between plasmid genes, and between chromosomal and plasmid-encoded genes (see below).

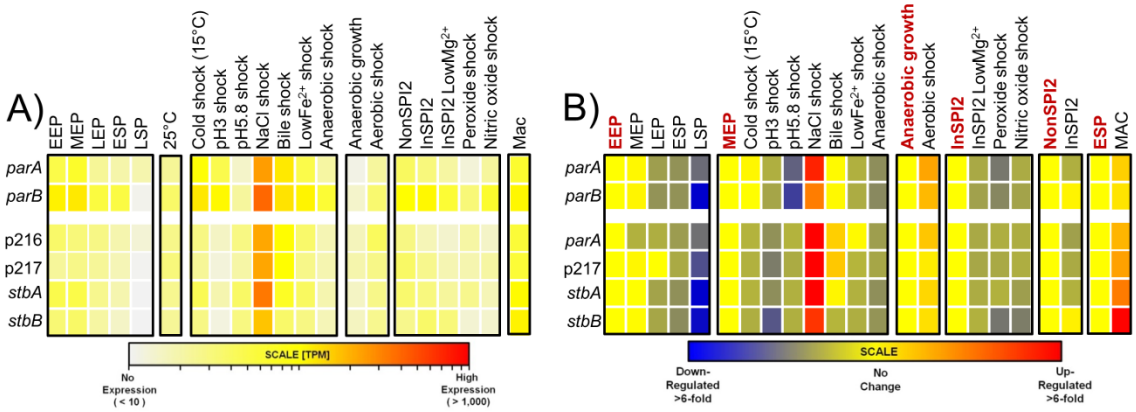


Figure 5.6: Absolute and relative expression of partitioning systems utilised by *Salmonella* plasmids pSLT$^{4/74}$ and pCol1B$^{4/74}$ during segregation.
Other details as the legend to Figure 5.5.

Plasmid partitioning cassettes are involved in the control of plasmid segregation during cell division (stability). The virulence plasmid of *Salmonella* encodes a Walker-type ATPase (*parA*/*B*) partitioning system, and the colicinogenic plasmid encodes an Actin-like ATPase (*stbA*/*B*) system (see *Section 1.2.1.2*). Both partitioning systems showed similar expression profiles, although their functional mechanisms do vary (Figure 5.6). The correlation coefficient between these two systems over the 23 environmental conditions was 0.75. Interestingly, the expression profile of two genes located adjacent to the *stbA* gene resembled the partitioning cassette (STM474_p216 and p217; correlation coefficient 0.78 and 0.73 respectively). It is not surprising that the expression pattern of the p216 gene mirrored the partitioning genes, because a BLAST search to find domains on the p216-encoded protein revealed that it encodes a ParA homolog. For the p217 gene, we failed to identify protein domains related to a partitioning system or any other known protein domains. But the expression profile and location of p217 in pCol1B$^{4/74}$ (in between p217 and *stbA* genes) suggests that this protein could play a role in plasmid partitioning during segregation. Further experimentation is required to deduce the function of the p217 gene.

5.2.5 *Levels of expression of plasmid toxin and antitoxin systems vary between pSLT$^{4/74}$ and pCol1B$^{4/74}$*

Toxin and antitoxin (T/A) mechanisms favour the growth of plasmid-retaining cells by reducing the competitiveness of their plasmid-free counterparts, thereby ensuring the retention of the plasmids in the population (Hayes 2003). The *Salmonella* virulence

plasmid uses Type I T/A gene pairs (*ccdB*/*ccdA*; toxin/antitoxin) to ensure its persistence during bacterial replication. In contrast, pCol1B$^{4/74}$ uses type II T/A coding/non-coding RNA pairs (*pndA*/PndB) for post-segregational cell killing (for details see *section 1.2.1.3*). Both T/A cassettes were expressed in all conditions, but the Type II cassette showed higher expression in absolute levels than the Type I (Figure 5.7A). The analysis to determine the mutual relationships revealed negative correlation between these two T/A systems (Correlation coefficient -0.2). The *ccdB*/*A* system was slightly up- or down-regulated in most environment conditions. A little variation was observed between two cassettes in their relative expression patterns, including Bile shock, aerobic shock, InSPI2 and *in vivo* growth conditions (figure 5.7B).

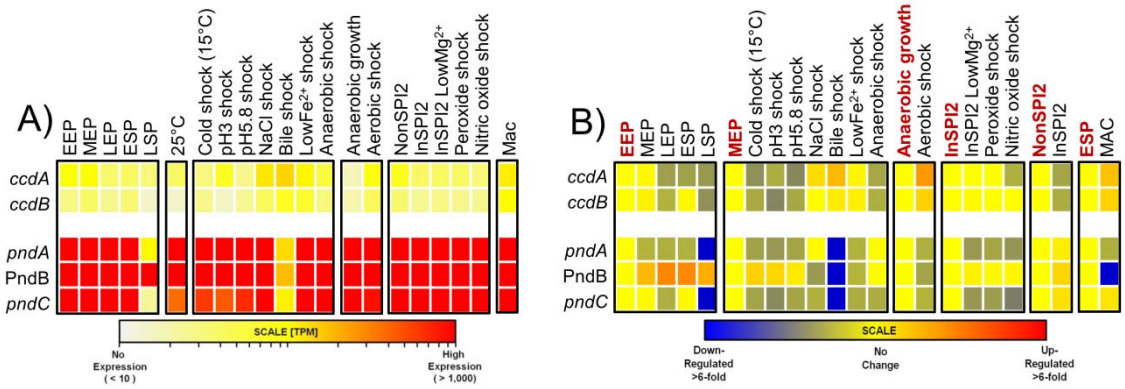


Figure 5.7: Absolute and relative expression of T/A systems used by *Salmonella* plasmids pSLT$^{4/74}$ and pCol1B$^{4/74}$.
The heat map shows the absolute (A) and relative (B) gene expression of the toxin and antitoxin (T/A) cassettes utilised by virulence and colicinogenic plasmids of *S.* Typhimurium. Other details as the legend to Figure 5.5.

Interestingly, the expression differences between toxin and antitoxin were observed in the Type II system. The killer gene *pndA* of the Type II system was slightly downregulated by *Salmonella* during growth (EEP to LSP) conditions, and in shocks including cold and acid. In contrast, the inhibitor non-coding RNA, PndB showed little upregulation in the above mentioned environments. In summary, both T/A cassettes showed distinct gene expression profiles, even though they perform a similar function in plasmid maintenance.

5.2.6 *Similarities between expression profiles of the spv locus and SPI-2*

The *Salmonella* pathogenicity Island 2 (SPI-2) is a type III secretion system that is essential for *Salmonella* to survive and replicate inside host cells, and contributes to enteric disease in animal models (Shea et al. 1999, Bispham et al. 2001). We searched for a SPI-2 pattern of gene expression in plasmid-encoded genes. The *ssaG* gene was used as an archetypical gene that represents the SPI-2 pattern of expression (Kröger et al. 2013), and encodes a membrane-located structural component of the SPI-2 type three secretion apparatus. The cluster-based analysis (Pearson correlation; *Section 2.10*) between the SPI-2 gene and the plasmid genes revealed that the *spv* locus has a similar expression profile to SPI-2 (Figure 5.8A). The correlation coefficient between the expression of *ssaG* and *spvA* over the 23 environmental conditions was 0.85, showing a high level of concordance between the expression profiles of the two genes (Table S5.3).

The *spv* locus was highly expressed in SPI-2 inducing and LowMg^{2+} (InSPI2) conditions, which resembles the SPI-2 gene expression profile. The expression of the *spv* locus was downregulated on peroxide and nitric oxide shock in SPI-2-inducing conditions. One difference between the *spv* locus and the SPI-2 gene *ssaG* is that the *spv* locus showed slight upregulation in LowMg^{2+} compared to InSPI2, whereas no up-regulation of *ssaG* in LowMg^{2+} was observed. Similarly, the expression of *ssaG* was upregulated by *Salmonella* in ESP (Bustamante et al. 2008), but this pattern was not found for the *spv* locus.

As expected, the pSLT$^{4/74}$-encoded *spv* locus which is essential for systemic infection but does not affect the ability of *S*. Typhimurium to infection epithelial cells in tissue culture (Guiney et al. 1995) was upregulated in MAC condition. The four-gene operon *spvABCD* operon is controlled by the upstream transcriptional regulatory protein *spvR*. Both *spvR* and the alternative sigma factor RpoS are required for efficient expression of the *spv* operon (Fang et al. 1992) in late stationary phase. The essential role of the SPI-2 island and *spv* genes in apoptosis of intestinal epithelial cells (Paesold et al. 2002) indicates that the genes from the SPI-2 locus and *spv* from pSLT$^{4/74}$ may be controlled by the same regulators.

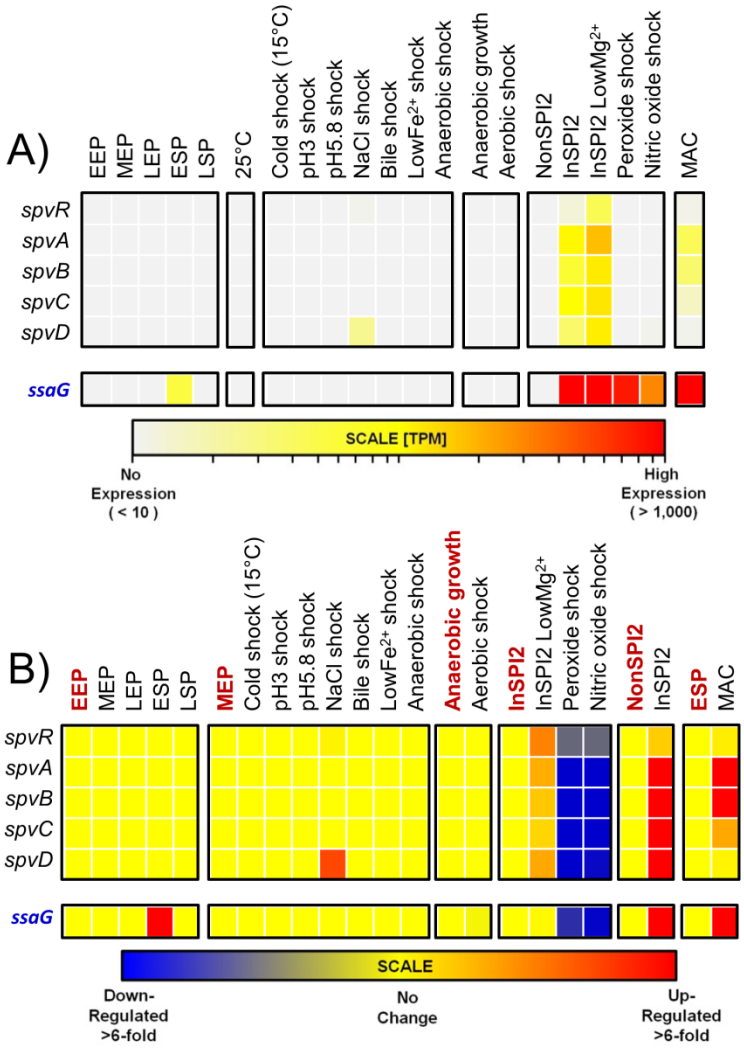


Figure 5.8: Absolute and relative expression of *spv* operon shows similar expression profiles to SPI-2 pathogenicity island (Global expression profile).
The heat map shows the absolute (A) and relative (B) gene expression of the *spv* locus. Other details in the legend to Figure 5.5. The *ssaG* gene was used as an archetypical gene that represents the SPI-2 pattern of expression. Only genes that show a correlation coefficient with *ssaG* of > 0.7 are shown.

To identify the relationship between *spv* and SPI-2 loci at the regulatory level, we compared the expression profiles of these two loci in a series of mutants that lack known regulatory factors. Cluster analysis of regulatory mutant gene expression profiles with plasmid-encoded genes revealed that the *spv* and SPI-2 loci showed same regulatory pattern (Figure 5.9; Supplementary Dataset 5.2). Higher expression of the *spv* locus in the *fur* mutant at early stationary phase suggests that the Fur protein represses this locus. Fur is a global iron regulator, which represses gene transcription by binding with high affinity to a DNA sequence known as a FUR box (Escolar et al. 1999). No FUR box was identified upstream of the *spvR* and *spvA* genes, suggesting

that the Fur-dependent repression of the *spv* locus is indirect. Reduced expression of the *spv* locus in deletion mutants lacking SlyA, PhoP/Q, EnvZ/OmpR, SsrA and SsrB strongly suggests that the *spv* locus is activated by SPI-2-relevant regulatory factors in the same manner as the SPI-2 locus itself. However, no changes in expression of *spv* locus were observed in the *rpoS* mutant in LSP condition.

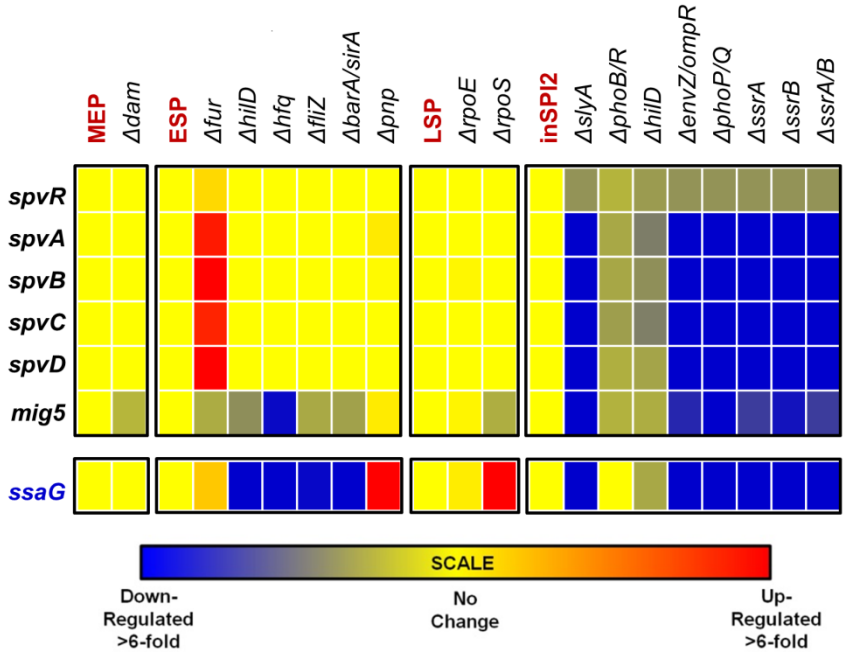


Figure 5.9: Cluster analysis showing the link between the regulation of the SPI-2 locus and plasmid-encoded genes. The profiles show gene expression in different regulatory mutants (Mutant expression profiles).
The cluster analysis groups the genes based on their expression levels in 16 different regulatory mutants grown in appropriate growth conditions by the Pearson correlation. The abbreviations for each mutants and growth conditions are shown in Supplementary Dataset 3.2. Other details as the legend to Figures 5.5 and 5.6.

We discovered that the STM474_p1066 gene (correlation coefficient between *ssaG* and p1066 – 0.79) showed similarities to the SPI-2 expression pattern in the mutant expression profile, raising the possibility that these genes may also play a role in *Salmonella* virulence. Valdivia et al. (1997) showed higher expression of this gene within the macrophage and named it '*mig5*' (macrophage-inducible gene 5). Similarly, we found that the *mig5* gene is the most upregulated gene in intracellular condition (~28-fold; see also Supplementary Dataset 5.2). The Mig5 protein, of 246 aa, has a transmembrane region and is predicted to contain a 'carbonic anhydrase' domain (Pro_CA, Pfam PF00484). The function of *mig5* remains unknown. The differential

expression pattern of *mig5* gene in Peroxide and Nitric oxide shocks (also MAC and InSPI2 growth conditions) showed similarities with *spv* and SPI-2 loci, suggesting that *mig5* may play a role in virulence. The reduction in expression of *mig5* in deletion mutants that lack SPI-2-relevant regulatory factors strongly suggests that, as for *spv* this gene is associated with SPI-2. Further experiments would be needed to investigate the role of *mig5* in *Salmonella* pathogenesis. Gonzalo-Asensio et al. (2013) identified a lengthy (597 nt) asRNA known as lesR-1 and located antisense to the 5' end of the *mig5* gene in $pSLT^{SL1344}$. It is particularly induced during persistence inside the host cell. However, the expression of the lesR-1 asRNA was not observed in any the 23 environmental conditions, suggesting that lesR-1 expresses only at intracellular conditions particularly when *Salmonella* persists inside the host cell.

Table 5.2: List of regulators involved in *spv* gene expression.

Regulators	References
Positive regulators	
SpvR	Caldwell et al. 1991; Wilson and Gulig 1998
RpoS	Fang et al. 1992
PhoP/Q	Libby et al. 1997; Matsui et al. 2000; This study
IHF	Marshall et al. 1999
ppGpp	Pizarro-Cerda and Tedin 2004
RcsB	Mariscotti et al. 2009
SsrA/B	This study
PhoB/R	This study
EnvZ/OmpR	This study
Negative regulators	
cAMP-CRP	O'Byrne and Dorman 1994
Dam	Heithoff et al. 1999 (*spvB*)
H-NS	O'Byrne and Dorman 1994
Lrp	Marshall et al. 1999
Pnp	Ygberg et al. 2006; This study
Fur	This study

The genetic control of *spv* locus expression is complex and requires specific environmental conditions, as well as interplay between several transcriptional regulators. Expression of *spv* genes is clearly induced when the bacteria reside within mammalian cells (Guiney et al. 1995) and *in vitro,* when the bacteria are grown in minimal medium mimicking the intravacuolar environment (up-regulation of *spv* locus by *Salmonella* in InSPI2 and $LowMg^{2+}$ conditions). Expression of *spv* can also be achieved, at least to some extent, as the bacteria enter the stationary phase of growth

in rich medium (Fang et al. 1992; Kowarz et al. 1994; Robbe-Saule et al. 1997). Apart from SpvR-mediated induction, there are several other regulators that play a role in the regulation of *spv* genes. Table 5.2 gives an overview of regulatory proteins which control *spv* gene expression in different growth and stress conditions.

5.2.7 *Drug resistance genes show similar expression pattern to the rpsM gene*

The *rpsM* gene encodes the 30S ribosomal protein, and is highly expressed in early exponential phase, with reduced expression at later growth stages. Interestingly, the genes responsible for streptomycin resistance (*strA* and *strB*) showed similar expression patterns in both absolute and relative expression profiles. The *strA and strB* genes showed a similar profile to *rpsM* at different growth phases, and also after salt and peroxide shocks. The correlation coefficient between the expression profile of *rpsM* and *strA* over the 23 environmental conditions was 0.88, showing a high level of concordance between the expression of the two genes. Figure 5.10 shows the genes that shared a correlation coefficient of more than 0.85 with the *rpsM* profile. It is possible that the function of StrA and StrB is related in some way to the ribosome, as chromosomally-encoded streptomycin resistance is known to be mediated thorugh the RpsL-encoded ribosomal protein S12 (Fernandez and Anton 1987).

There are significant similarities between the expression of the *nikA*/*B* genes from pCol1B$^{4/74}$ and *rpsM* (correlation co-efficient 0.77 and 0.91 for *nikA* and *nikB*, respectively). The *nikA/B* genes, required for conjugational DNA processing, were essential for the recombination (Pansegrau et al. 1990). NikA and TraJ proteins share amino acid sequence homology, as do the NikB and TraI proteins (Pansegrau et al. 1993). NikA and NikB are required for form of relaxation complex during the conjugation of R64 plasmids like pCol1B$^{4/74}$ (Pansegrau et al. 1990; Furuya et al. 1991). Even though the *nikA*/*B* gene pairs are key to bacterial conjugation, their expression profiles varied from other conjugational genes. The difference in the *nikA/B* expression profile raises the possibility that these genes encode other functions, like the *traT* gene which encodes for serum resistance (Glöckner et al. 1997).

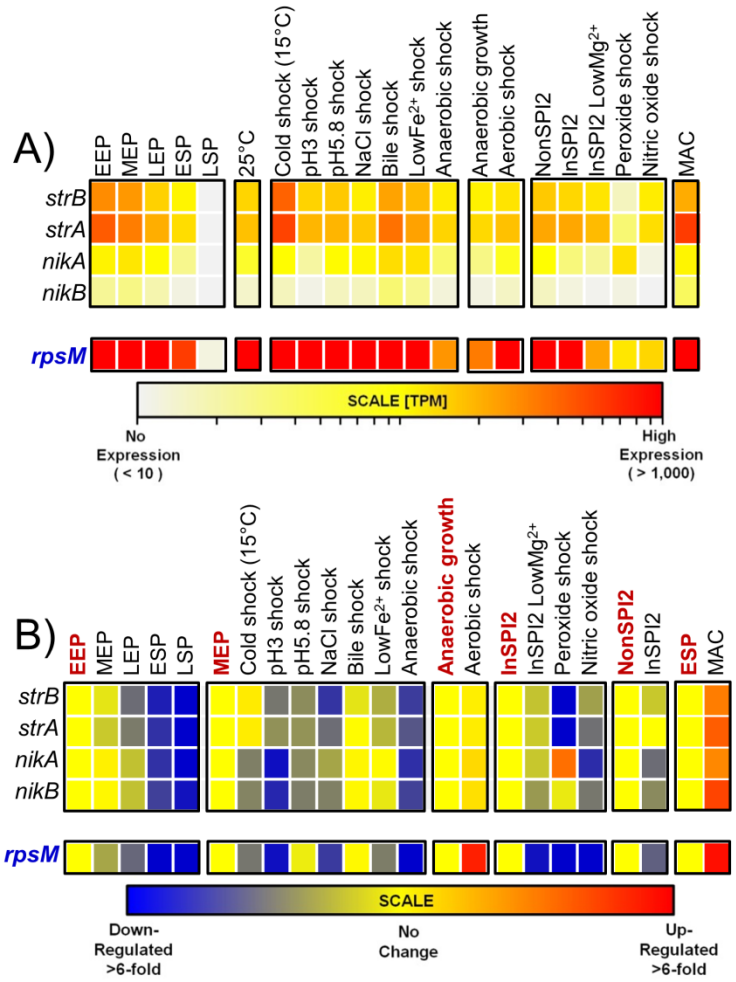


Figure 5.10: Identification of plasmid-encoded genes with similar expression profiles to *rpsM*.

Details as the legend to Figure 5.6.

Our cluster analysis identified similarities in the expression of genes which encode core functions including *repA, repZ, stbB* and *rsdB*, and the profile of the *rpsM* gene (Table S5.3). The *stbB* (*parR*) gene encodes a partitioning function which is similar to *parB* gene. *rpsM* encodes a chromosomal core function, which may be the possible explanation for an expression profile similar to plasmid core genes. Interestingly, the expression profile of two genes located within the *pef* locus (*orf6* and *srgA*) and a thermoregulatory protein (*tlpA*) also showed similarities to the *rpsM* gene. Further experimental validation would be needed to study the relationship between these genes.

5.2.8 *Many asRNAs are down-regulated by Bile shock*

In general, asRNAs inhibit the translation of complementary mRNA by base pairing and physical occlusion of the translation machinery. The upregulation of asRNA would affect the expression of sense coding gene in the same condition, which generates opposite expression patterns between asRNA/mRNA pairs. The RNA-Seq based transcriptomic study based upon the ESP growth condition identified 33 asRNAs in all three *Salmonella* plasmids (*Section 4.2.4*). The expression profile of each asRNA was determined and analysed to identify positive or negative relationship between asRNAs and coding genes. To determine the negative relationship, we calculated correlation coefficients between expression profiles of asRNA/mRNA (as described in *Section 2.10*). However, the analysis did not reveal strong correlation between any asRNA and its cognate sense mRNA. Even functionally well-characterised pairs including FinP/*traJ* and PndB/*pndA* did not show any negative or positive correlation.

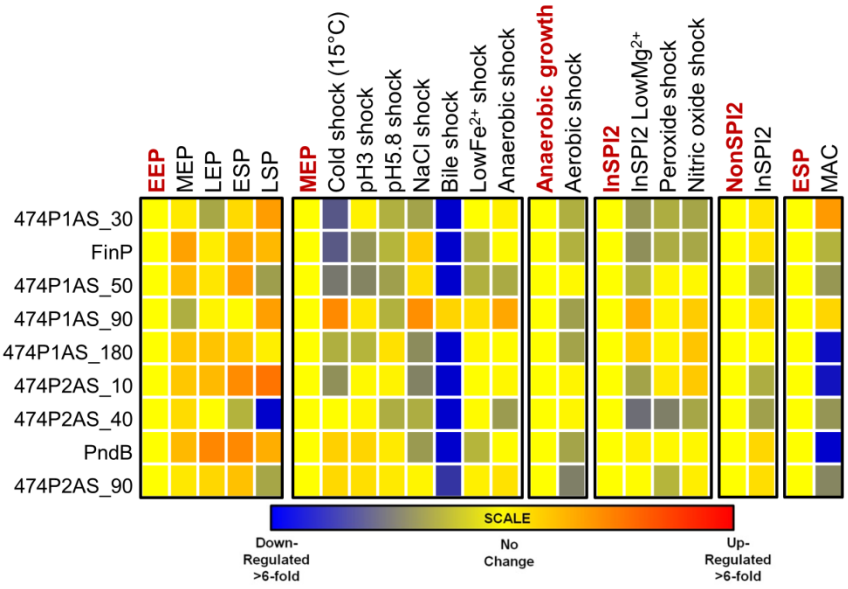


Figure 5.11: Relative expression profile of plasmid-encoded asRNAs.
Newly predicted asRNAs from plasmids of *Salmonella* showed similar expression pattern, which includes the down-regulation of asRNAs by bile shock in MEP growth condition. Other details as the legend to Fig. 5.5.

We observed that most of the asRNAs were expressed in most of the environmental conditions except 474ASP1_10, 474ASP2_60 and 474ASP2_100 which were only expressed in certain conditions (See Supplementary Data 5.1). Interestingly, the relative expression profiles revealed that most of the asRNAs were downregulated by bile shock (Dataset S5.1). Figure 5.11 illustrates the relative expression profiles of

asRNAs which were validated by Northern blot (*Section 4.2.4*). Bile is a fluid synthesised by liver cells and contains salt, cholesterol and a variety of proteins and electrolytes (Hofmann et al. 2008). *Salmonella* is a bile-resistant pathogen and it has been suggested that the ability of *Salmonella* to colonise gall stones and form biofilms in the gall bladder is one aspect of its pathogenic lifestyle that is relevant to chronic carriage (Hernández et al. 2009). The asRNAs that were downregulated by bile shock under the MEP growth condition might suggest that at least some of these asRNAs may have a role in regulating some cell-surface *Salmonella* proteins. The asRNAs antisense to replication genes including 474P1AS_190 and 474P1AS_180 were upregulated by low magnesium and nitric oxide shock. Further experimental validation is required to determine if these asRNAs have any function related to *Salmonella* virulence. Because of the higher expression of 474P1AS_30 in the intracellular condition, it is tempting to speculate that this asRNA could play a role in *Salmonella* virulence.

5.2.9 *Cross-talk between chromosome and plasmids of Salmonella at the regulatory levels*

Regulatory factors play a key role in the control of bacterial gene expression. To investigate the role of chromosomal-encoded regulatory proteins on plasmid-encoded genes, we determined the genes that were ≥ 3 fold up- or down-regulated in RNA-Seq data generated from a series of mutant strains (Table S5.1). The results revealed that a small subset of genes were differentially expressed in certain regulatory mutants, including up-regulation of 23 coding genes and 7 asRNAs in the *pnp* mutant (Figure 5.12). The analysis to identify the similar expression patterns in plasmid genes did not reveal any similarities, except the genes showing an SPI-2 pattern (*Section 5.2.6*). However, the pSLT$^{4/74}$-encoded *rlgA* and two hypothetical proteins (STM474_p1068 and STM474_p1069) were > 3-fold down-regulated by *Salmonella* in SPI-2 regulator, mutants that lack PhoP/Q and SsrA/B (note that rlgA, STM474_p1068 and STM474_p1069 genes and *mig5* are adjacent genes; see Supplementary Dataset 5.2).

As expected, a down-regulation (> 6-fold) of FinP was observed in the *dam* mutant (Torreblance et al. 1999; Camacho et al. 2005) at MEP condition (Figure 5.12). Similarly, the pSLT4/74-encoded asRNA 474P1AS_100 was 3-fold down-regulated by *Salmonella* in the *dam* mutant. In the FinP/*traJ* pair, Dam methylation involves two independent, concerted actions; activation of FinP synthesis and repression of *traJ*

transcription thereby inhibiting the plasmid conjugation. The 474P1AS_100 is located antisense to 3' end of the *pefB* gene and showed Class I expression pattern (*Section 4.2.4*; Table S4.3). In 474P1AS_100/*pefB* pair, Dam activates *pef* transcription by methylation of the *pef* GATC site located downstream to the *pefB* gene (Nicolson and Low 2000). The expression pattern (Class I) and potential Dam-mediated regulation of 474P1AS_100, suggests that this asRNA should be studied at the experimental level.

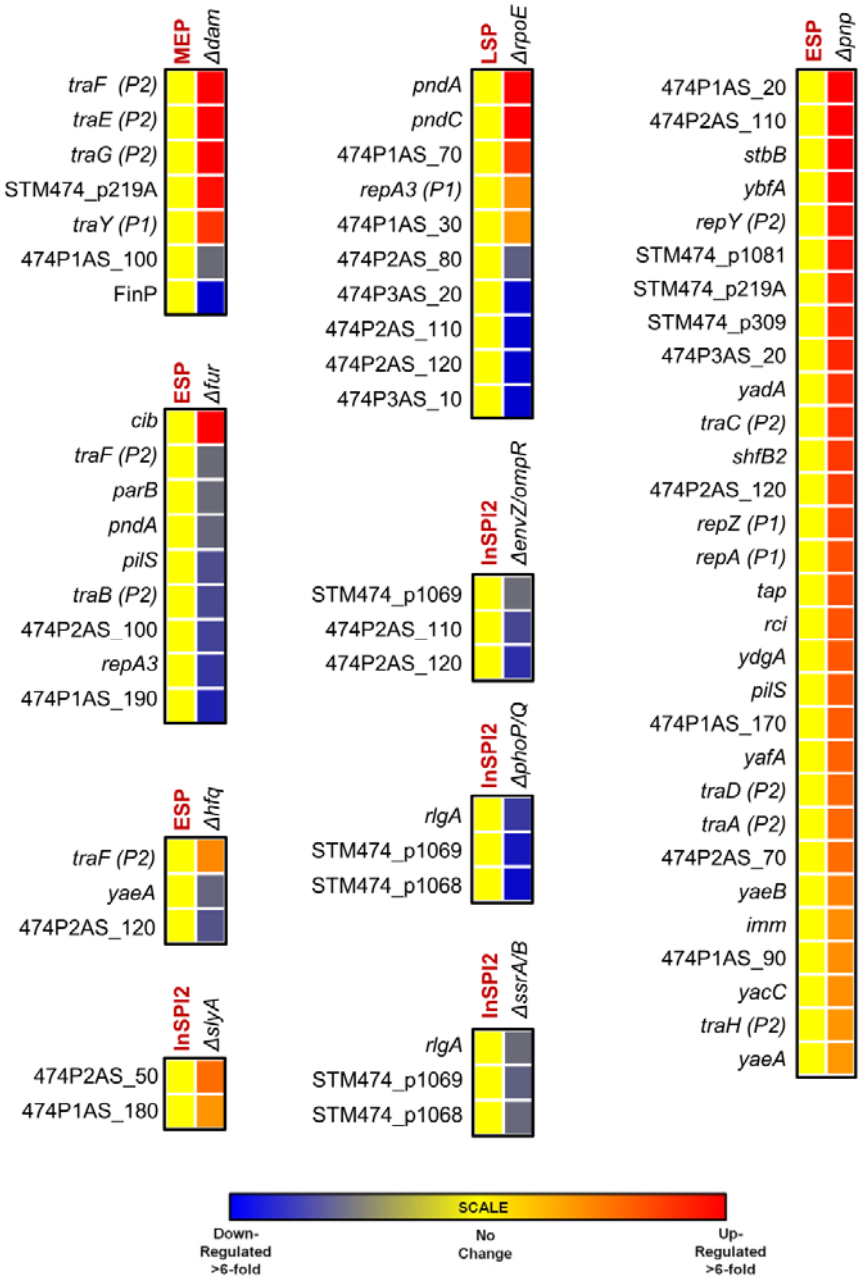


Figure 5.12: The plasmid-encoded genes that are ≥ 3-fold up- or down-regulated by *Salmonella* in different regulatory mutants.

The genes showed in Figure 5.9 are not included here. P1: pSLT$^{4/74}$, and P2: pCol1B$^{4/74}$. Other details as in the legend to Fig. 5.9.

5.3 *Conclusion*

Using dRNA-Seq data generated from 4 different growth conditions, we have identified 162 TSS in three plasmids of *S*. Typhimurium 4/74, including 56 antisense TSS and 27 internal TSS. Of the 56 antisense TSS, 33 were already annotated as asRNAs (*Section 4.2.4*). We found a secondary TSS for one asRNA, 474P2AS_10 (Table S5.2). Interestingly, our dRNA-Seq approach revealed two orphan TSS in *Salmonella* plasmids pSLT$^{4/74}$ and two in pCol1B$^{4/74}$. Further experimentation including Northern blot validation is required to confirm the presence of these orphan transcripts. Finally, our dRNA-Seq data on 'RNA Pool' revealed that 98% of the plasmid-associated TSS could be identified, suggesting that RNA Pool is a successful strategy to identify the vast majority of promoters in a single step from multiple growth conditions.

The transcriptome reveals the status and functional mechanism of a cell as it responds to external stimuli. Many RNA-Seq analysis methods have been used to determine the functional mechanism of cell responses to external stimuli by identifying the list of genes that are differentially expressed (Caimano et al. 2014; Kröger et al. 2013; Osmundson et al. 2013). In addition, evidence has shown that genes with similar expression patterns are often regulated through similar mechanisms (Puniya et al. 2014; Fang et al. 2013; Stuart et al. 2003). Here, we globally clustered genes in *Salmonella* plasmids according to their expression patterns across 23 environmental conditions. The results showed that the genes with similar functions showed similar expression patterns including partitioning systems from pSLT$^{4/74}$ (*parA/B*) and pCol1B$^{4/74}$ (*stbA/B*), and the *spv* locus clustered with SPI-2 genes. In contrast, the genes involved in plasmid maintenance (T/A system) including *ccdA/B* from pSLT$^{4/74}$ and *pndA*/B encoded by pCol1B$^{4/74}$ showed a negative correlation, even though they carry out similar functions in plasmids. Finally, the experimentally validated plasmid-encoded asRNAs of *Salmonella* (nine asRNAs; *Section 4.2.4*) showed similar expression patterns (see Supplementary Dataset 5.1). Differential expression profiles of these asRNAs revealed that most of them were down-regulated in Bile shock.

This chapter develops hypotheses for the function of several individual plasmid-encoded coding genes or antisense transcripts that should be tested experimentally in the future, and these are summarised in Table 5.3. Individual studies revealed the role of certain regulatory proteins in controlling expression of plasmid genes of *Salmonella* (Nicholson and Low 2000; Fang et al. 1992; Ahmer et al. 1998). Here, the cross-talk

between plasmids and chromosome of *Salmonella* at the regulatory level was investigated using a series of mutants that lack known regulatory factors, including transcription factors, global regulators, SPI-1 and SPI-2 associated regulators. The results unveiled the highly complex regulatory network between the chromosome and plasmids of *Salmonella*.

Table 5.3: List of plasmid-encoded genes worthy of future experimental study.

*Reason for the choice of further experimental studies.

#The relevant Sections or tables where the genes are discussed.

Plasmid-encoded asRNAs are shown in **red**.

Gene or asRNA	Reason*	Section#
474P1AS_70	Antisense to *spvR*	5.2.2
474P1AS_80	Antisense to *spvA*	5.2.2
STM474_p217	Showed similar expression profile with *stbA/B* partitioning system (correlation co-efficient 0.73)	5.2.4
spv locus	Expression was altered by global regulator Fur	5.2.6
mig5	Showed similar expression profile with SPI-2 (correlation co-efficient 0.79)	5.2.6
474P1AS_190	Up-regulated by LowMg^{2+} and nitric oxide shock	5.2.8
474P1AS_30	Up-regulated in intracellular condition	5.2.8
STM474_p1068	Expression was altered by the SPI-2-associated regulators PhoP/Q and SsrA/B	5.2.9
STM474_p1069	Expression was altered by the SPI-2-associated regulators PhoP/Q and SsrA/B	5.2.9
474P1AS_100	Overlaps the 3' end of the *pefB* gene and expression was regulated by Dam	5.2.9
474P2AS_60	Showed similar expression profile with SPI-2 (correlation co-efficient 0.80)	Table S5.3
STM474_p1095	Showed similar expression profile with SPI-2 (correlation co-efficient 0.79)	Table S5.3
STM474_p1098A	Showed similar expression profile with SPI-2 (correlation co-efficient 0.79)	Table S5.3
STM474_p1059	Showed similar expression profile with *rpsM* (correlation co-efficient 0.86)	Table S5.3
STM474_p1091	Showed similar expression profile with *rpsM* (correlation co-efficient 0.86)	Table S5.3
474P2AS_100	Showed similar expression profile with SPI-1 (correlation co-efficient 0.79)	Table S5.3

Table S5.1: Number of plasmid-encoded CDS* that are >3-fold up- or down-regulated in different regulatory mutants.

Absolute expression (TPM) values of plasmid-encoded genes were calculated using TPM approach as described in *Section 3.2.2.6*.

Relative expression was calculated by comparing the TPM values from mutant RNA-Seq data with wild-type RNA-Seq data generated under the same growth condition.

Growth conditions are detailed in Table 2.3 (*Section 2.1.3*).

* CDS encoded by plasmids pSLT[4/74], pCol1B[4/74] and pRSF1010[4/74].

Growth conditions	Mutant strains	Differentially expressed gene		
		Up	Down	Total
MEP	Δ*dam*	5	0	5
ESP	Δ*fur*	5	6	11
	Δ*hilD*	0	0	0
	Δ*hfq*	1	2	3
	Δ*fliZ*	0	0	0
	Δ*barA/sirA*	0	0	0
	Δ*pnp*	23	0	23
LSP	Δ*rpoE*	0	0	0
	Δ*rpoS*	3	0	3
InSPI2	Δ*slyA*	0	5	5
	Δ*phoB/R*	0	0	0
	Δ*hilD*	0	0	0
	Δ*ompR/envZ*	0	6	6
	Δ*phoP/Q*	0	9	9
	Δ*ssrA*	0	5	5
	Δ*ssrB*	0	7	7
	Δ*ssrA/B*	0	8	8

Table S5.2: Transcriptional start sites of *Salmonella* plasmids pSLT$^{4/74}$, pCol1B$^{4/74}$ and pRSF1010$^{4/74}$.

Promoter categorization was carried out as described in *Section 3.2.2.5*. In columns five to nine, the promoter category is indicated by '1'.

The five right-hand columns show if a TSS is present (TPM ≥ 10; coloured in black) or absent in the respective condition (TPM < 10; coloured in red). Cells containing black coloured values are considered to be expressed in the relevant condition(s) (TPM ≥ 10).

Growth conditions are detailed in Table 2.3 (*Section 2.1.3*). Newly identified asRNAs are shown in **blue**.

4/74 Plasmid Transcription Start Sites									TPM (First 10 nt)				
TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
pSLT$^{4/74}$													
14	+	**474P1AS_10**						1	9	14	19	19	12
616	-	STM474_p1001	*finO*	1					57	58	36	43	47
1250	+	**474P1AS_20**						1	84	15	58	105	76
3733	+	**474P1AS_30**						1	30	57	233	197	32
6707	+	STM474_p1005		1					94	107	578	90	136
10391	-	STM474_p1008	*traT*			1			41	179	1030	479	141
11280	-	STM474_p1009	*traS*	1					107	645	975	348	437
11300	+							1	0	21	104	17	14
11332	+							1	5	25	123	19	24
11381	-	STM474_p1009	*traS*		1				29	258	856	90	148
11398	-	STM474_p1009	*traS*		1				8	44	488	68	66
16225	-	STM474_p1013	*traQ*	1					2	16	61	11	6
31518	-	STM474_p1035	*traY*			1			14	27	55	74	41
32265	+	**FinP**	FinP					1	135	244	968	375	224
32369	-	STM474_p1036	*traJ*	1					27	38	108	40	50
32872	-	STM474_p1037	*traM*	1					2	22	55	6	12

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
33234	+	STM474_p1038		1					5	17	29	20	18
33360	+	STM474_p1038				1			37	112	112	92	97
36490	-	STM474_p1040	*psiB*			1			26	27	23	22	28
38566	+	474P1AS_50						1	246	664	822	490	259
40507	-	STM474_p1044		1					10	25	25	31	22
41117	+							1	2	16	202	0	14
46037	+	STM474_p1053	*samA*	1					31	50	211	45	89
48057	-	STM474_p1055	*parB*			1			81	48	186	139	100
49333	-	STM474_p1055	*parB*	1					59	113	68	49	49
50044	-	STM474_p1056	*parA*	1					133	78	94	62	193
50276	+	STM474_p1058			1				0	10	0	4	11
50415	+	STM474_p1058		1					70	32	28	49	47
51091	+	STM474_p1058				1			151	223	182	90	206
54196	-	STM474_p1063	*tlpA*	1					142	60	142	37	88
54404	+	STM474_p1064		1					52	19	81	19	29
54672	+	STM474_p1064				1			43	47	85	121	87
55539	+	STM474_p1066		1					190	362	290	1533	512
56487	+	STM474_p1067	*rlgA*	1					226	756	1175	686	682
59821	+	**474P1AS_60**						1	14	60	3	43	39
60912	+	**474P1AS_70**						1	112	460	315	580	157
61502	+	STM474_p1072	*spvR*	1					0	0	8	17	17
61828	+	STM474_p1075	*spvA*	1					2	1	14	13	7
61906	-	**474P1AS_80**						1	9	38	55	43	50
62794	-						1		35	37	40	62	66
68021	-	STM474_p1081	*IS200-like*	1					109	94	252	46	86

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
68060	-	STM474_p1081	*IS200-like*		1				29	49	142	15	35
68152	+	STM474_p1082	*STM474_p1082*	1					2791	3032	6265	1492	2519
73777	-	STM474_p1094	*ccdA*	1					188	157	295	118	183
74022	+							1	7	8	399	3	9
74090	-	STM474_p1094	*ccdA*		1				56	50	59	15	56
74408	+	STM474_p1095	*STM474_p1095*	1					7	18	25	55	29
75958	-	**474P1AS_90**						1	19	55	458	86	64
76014	-							1	24	30	173	56	41
76024	+	STM474_p1098	*repA2*		1				168	100	348	77	195
76065	+	STM474_p1098	*repA2*	1					484	896	3757	462	657
77167	+	STM474_p1098	*repA2*			1			81	109	34	75	69
78104	-	STM474_p1098A				1			7	34	165	84	32
78483	-	STM474_p1098A		1					7	18	11	101	30
78588	+						1		31	6	3	69	49
79116	+	STM474_p1099	*pefB*	1					12	8	0	7	11
79600	-	**474P1AS_100**						1	51	44	521	37	65
84331	-						1		27	16	23	273	64
84434	+	STM474_p1104				1			50	47	82	32	53
84664	+	STM474_p1104				1			81	108	590	156	123
84793	+	STM474_p1104				1			206	109	77	208	182
84940	-							1	61	16	6	50	49
85731	-	**474P1AS_110**						1	97	78	103	114	107
86331	-	**474P1AS_120**						1	0	40	25	27	24
87460	-	**474P1AS_130**						1	19	33	77	34	44
87967	-	**474P1AS_140**						1	20	42	125	76	41

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
88278	-	**474P1AS_150**						1	23	48	12	66	31
88905	-	**474P1AS_160**						1	33	64	35	45	52
89533	-							1	9	13	231	19	17
89799	-	**474P1AS_170**						1	99	43	88	76	95
89867	+	STM474_p1110	*srgC*			1			38	118	138	74	86
90649	+							1	11	30	28	18	20
92010	+	**474P1AS_180**	*repA3*					1	11025	8397	24959	4828	9500
92179	-	STM474_p1114A	*repA3*	1					894	1015	706	918	1113
92409	+	**474P1AS_190**						1	17	71	18	22	27
92519	-	STM474_p1115	*repC*	1					385	488	940	299	392
93835	-	STM474_p1117		1					8	19	50	4	12
pCol1B$^{4/74}$													
155	+	STM474_p201	*repY*	1					2308	1884	1753	2780	2702
365	-	**474P2AS_10**						1	16520	36336	72530	22082	18345
461	-				1				30	9	14	33	29
2370	+	STM474_p202	*yacA*	1					124	294	1272	124	220
2915	+	STM474_p204	*yacC*	1					8	18	243	39	30
3657	+	STM474_p204	*yacC*			1			32	62	40	103	74
3921	+	STM474_p205	*yadA*	1					204	178	540	177	200
4157	+	STM474_p205	*yadA*			1			14	39	109	53	26
4526	+	STM474_p206	*yaeA*	1					159	272	4012	131	263
4540	-							1	16	33	32	55	78
4752	-	**474P2AS_20**						1	12	90	383	48	40
4796	-				1				8	8	57	15	11
6340	-	STM474_p209	*yafB*	1					88	170	959	43	96

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
6634	+	STM474_p210	*yagA*	1					21	86	70	32	44
7482	+	STM474_p210	*yagA*			1			27	48	92	60	40
8090	-							1	50	31	64	46	54
8162	-				1				7	6	218	7	24
8227	+	STM474_p211	*cib*	1					50	15	108	9	88
10424	-	STM474_p212	*imm*			1			37	41	71	28	56
10591	+							1	93	38	287	9	96
10593	-	STM474_p212	*imm*	1					23	2	74	3	31
11064	-						1		24	21	52	59	60
11938	-	STM474_p213	*ydfA*	1					55	54	36	28	47
11974	-	STM474_p213	*ydfA*		1				24	15	6	7	16
12332	+	STM474_p214	*ybfA*	1					20	42	254	15	39
13899	-							1	23	36	541	13	23
13965	-	**474P2AS_30**						1	49	35	59	13	59
14146	+	STM474_p216	*parA*	1					34	66	236	16	58
14595	+	STM474_p216	*parA*			1			63	38	17	27	64
15338	+	STM474_p218	*stbA*	1					253	88	76	58	173
15515	+	STM474_p218	*stbA*			1			111	59	120	112	99
17134	-	STM474_p219A		1					73	51	100	60	130
24132	-	**474P2AS_40**						1	746	943	524	1322	1510
26523	-							1	7	18	22	43	17
27906	+	STM474_p233	*yddA*	1					40	101	232	114	66
30109	-	**474P2AS_50**						1	8337	3741	1982	2581	5926
30412	+	STM474_p237		1					62	33	6	136	52
30895	+	STM474_p237	*ydgA*			1			50	77	137	110	56

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
32217	-							1	48	11	8	78	63
33282	+	STM474_p239	*ydiA*			1			7	24	4	126	19
33763	-	STM474_p240		1					608	379	90	350	601
33868	+	STM474_p240A	*nikA*	1					68	63	203	53	92
36919	-				1				24	26	19	116	51
36970	-	**474P2AS_60**						1	5	13	0	9	5
40030	+	**474P2AS_70**						1	76	50	53	96	72
40208	+							1	26	24	36	67	47
41736	+	STM474_p245	*pndC*	1					135	355	100	294	147
41872	-	**PndB**	PndB					1	7735	18119	38203	6574	7564
42429	-	STM474_p247		1					73	199	72	123	98
42636	-	**474P2AS_90**						1	682	1808	1625	818	781
42698	+	STM474_p248		-		1			2158	6030	3662	6245	2331
43554	+	STM474_p249		1					20	104	31	40	69
44588	-	STM474_p250	*excA*	1					556	658	855	521	816
44622	-	STM474_p250	*excA*		1				14	32	52	31	32
44692	-	STM474_p250	*excA*		1				44	24	60	135	54
45033	-	STM474_p250A	*traY*			1			13	57	355	26	57
46727	+	**474P2AS_100**						1	2	44	24	12	12
47037	+							1	26	47	8	144	45
62426	+							1	2	5	6	72	14
63809	+	**474P2AS_110**						1	12	49	235	101	54
65701	+	**474P2AS_120**						1	2	68	308	27	36
65966	-	STM474_p268	*traH*	1					2727	2267	4720	1912	3551
68796	-	STM474_p271	*traE*	1					104	172	19	87	86

Continued on next page

TSS	Strand	Systemic ID	Common name	Primary	Secondary	Internal	Orphan	Antisense	MEP	ESP	LSP	InSPI2	Pool
70531	-	STM474_p272B	*shfB2*			1			2	41	120	29	23
70993	-							1	9	23	7	24	22
71313	-	STM474_p272C	*pilV*			1			14	18	55	41	21
80922	+							1	38	18	3	39	17
83463	-	STM474_p286	*traD*			1			41	35	17	39	44
83475	+	**474P2AS_130**						1	21	74	62	63	63
84161	-	STM474_p282	*traD*	1					38	29	9	68	53
85835	-						1		12	40	20	32	42
86343	-	STM474_p289	*traA*	1					3962	10073	35363	6927	5877
86485	-	STM474_p290		1					72	210	282	212	118
86612	-	STM474_p290			1				26	148	34	19	43
pRSF1010$^{4/74}$													
864	+	**474P3AS_10**						1	733	3122	28591	1283	2034
871	-	STM474_p301		1					1983	1100	588	1336	1542
947	-	STM474_p301			1				171	110	41	236	170
951	-	STM474_p301			1				147	93	33	216	148
1570	+	**474P3AS_20**						1	145	423	968	290	323
2822	+	**474P3AS_30**						1	35	86	609	111	55
3300	-	STM474_p305		1					407	286	335	309	466
5617	-	**474P3AS_40**						1	245	224	154	117	381
5641	+	STM474_p308		1					79	43	11	40	67
6239	+	STM474_p309		1					39	114	440	69	88
7191	-	STM474_p310	*strB*			1			523	229	64	429	347

Table S5.3: List of plasmid-encoded genes that showed similarities with expression profile of chromosomal genes under 23 environmental conditions.

Expression profiles of plasmid-encoded genes that showed similarities with chromosomal genes (Pearson correlation coefficient score > 0.7) are listed here.

The eight representative chromosomal genes were selected based on their biological importance during Salmonella infection and the data are described in *Section 5.2.6* and *5.2.7*.

Chromosomal genes	Systemic ID	Common name	Correlation coefficient
ssaG [STM474_1412]	STM474_p1075	*spvA*	0.85
	STM474_p1077	*spvB*	0.84
	STM474_p1008	*traT*	0.83
	STM474_p237	*ydgA*	0.80
	474P2AS_60	474P2AS_60	0.80
	STM474_p1078	*spvC*	0.77
	474P2AS_110	474P2AS_110	0.75
	STM474_p1009	*traS*	0.75
	STM474_p1072	*spvR*	0.74
	STM474_p1095		0.74
	STM474_p1098A		0.73
	STM474_p1080	*spvD*	0.71
	STM474_p250A	*traY*	0.70
prgH [STM474_3013]	474P2AS_100	474P2AS_100	0.79
	STM474_p274	*pilT*	0.77
	STM474_p258	*traQ*	0.72
siiA [STM474_4452]	474P2AS_100	474P2AS_100	0.75
	STM474_p275	*pilS*	0.73
	STM474_p258	*traQ*	0.70
asr [STM474_1494]	STM474_p1078	*spvC*	0.85
	STM474_p1077	*spvB*	0.82
	STM474_p1075	*spvA*	0.79
	STM474_p239	*ydiA*	0.79
	STM474_p1098A		0.78
	STM474_p1095		0.78
	STM474_p1072	*spvR*	0.76
	STM474_p1080	*spvD*	0.76
	STM474_p1066		0.75
	474P2AS_60	474P2AS_60	0.72

Continued on next page

Chromosomal genes	Systemic ID	Common name	Correlation coefficient
pagC [STM474_1244]	STM474_p263A	*sogL*	0.75
	STM474_p1077	*spvB*	0.73
	STM474_p237	*ydgA*	0.72
	STM474_p1075	*spvA*	0.71
	474P2AS_110	474P2AS_110	0.71
pipC [STM474_1082]	STM474_p281	*pilM*	0.84
	STM474_p258	*traQ*	0.82
	STM474_p275	*pilS*	0.82
	STM474_p276	*pilR*	0.81
	STM474_p274	*pilT*	0.78
	STM474_p278	*pilP*	0.76
	474P2AS_110	474P2AS_110	0.75
	STM474_p279	*pilQ*	0.70
entA [STM474_0618]	STM474_p1090	*rsdB*	0.79
	STM474_p1091		0.78
	STM474_p301	*sul2*	0.75
	STM474_p1092		0.74
	STM474_p1088		0.73
	STM474_p241	*nikB*	0.72
	STM474_p310	*strB*	0.71
rpsM [STM474_3583]	STM474_p241	*nikB*	0.91
	STM474_p310	*strB*	0.88
	STM474_p311	*strA*	0.88
	STM474_p1059		0.86
	STM474_p1091		0.86
	STM474_p1090	*rsdB*	0.85
	STM474_p1092		0.84
	STM474_p1088		0.84
	STM474_p1062		0.82
	STM474_p1060		0.82
	STM474_p201A	*repZ*	0.79
	STM474_p1107	*srgA*	0.79
	STM474_p240A	*nikA*	0.77
	STM474_p301	*sul2*	0.77
	STM474_p219	*stbB*	0.75
	STM474_p1098	*repA2*	0.74
	STM474_p1104	*orf6*	0.73
	STM474_p215A		0.73
	STM474_p1117		0.72
	STM474_p1063	*tlpA*	0.72
	474P1AS_40	474P1AS_40	0.70
	STM474_p213	*ybaA*	0.70

CHAPTER 6

FUNCTIONAL CHARACTERISATION OF PLASMID-ENCODED ANTISENSE RNAs AND SMALL RNAs

6.1 *Introduction*

Bacterial sRNAs are an emerging class of bacterial post-transcriptional regulators (Storz et al. 2011). Apart from protein binding sRNAs (e.g., CsrA and CsrB), others directly interact with the mRNA target at the post-transcriptional level (Liu et al. 2010). The sRNAs can be grouped into two categories; *cis*-encoded asRNAs and *trans*-encoded sRNAs. In general, *cis*-encoded asRNAs have a regulatory role on cognate coding genes located on the sense strand mediated by perfect base pairing (Thomason and Storz 2010; Georg and Hess 2011). In contrast, *trans*-encoded sRNAs have a regulatory role that involves genes that are located on distinct regions of the chromosome (Sharma et al. 2007; 2011). Most challenging in terms of target identification are the *trans*-encoded sRNAs, because the genomic locations of their targets are unknown unlike asRNAs. *Trans*-encoded sRNAs are a fast-acting and versatile class of regulators. Target mRNA abundance is significantly altered within 10 min of sRNA expression (Papenfort et al. 2006). The experiment described by Papenfort et al. (2006) proved the successful identification of *trans*-encoded sRNA targets. Subsequently, a study by Urban and Vogel (2007) showed that the direct interaction between an sRNA and its target mRNA could be validated using a GFP-based two-plasmid assay. Corcoran et al. (2012) improved the two-plasmid assay by translationally fusing the 5'UTR and start of the CDS of the putative target mRNA to the superfolder-GFP (sf-GFP) protein.

Using the experiments mentioned above, I investigated the regulatory role of plasmid-encoded sRNA and asRNAs. The key findings are putative roles of sRNA STncP1-1 in *Salmonella* virulence and regulation of chromosomal-encoded mRNA gene expression. The putative role of plasmid-encoded antisense RNAs on chromosomal-encoded genes were studied using pulse-overexpression followed by microarray experiments.

6.2 *Results and discussion*

6.2.1 *STncP1-1 a plasmid-encoded small RNA*

6.2.1.1 *Identification of a novel small RNA in Salmonella virulence plasmid*

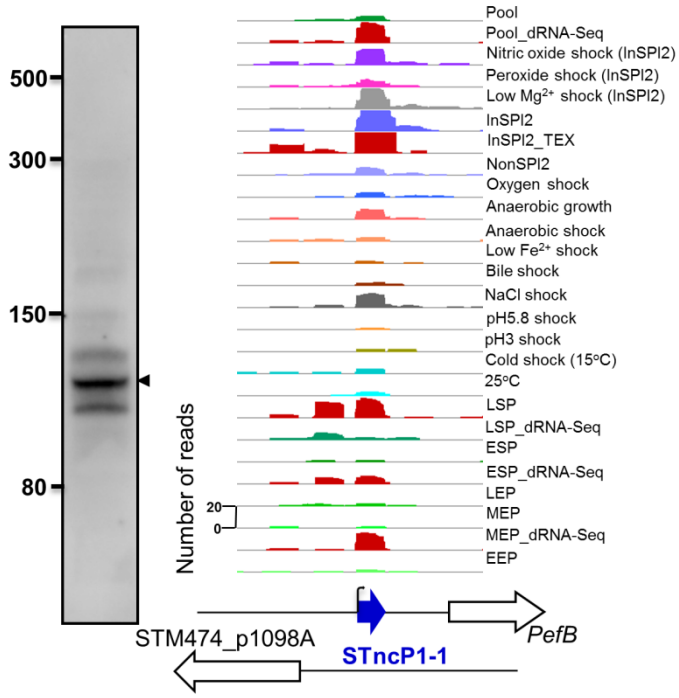


Figure 6.1: RNA-Seq based identification of plasmid trans-encoded sRNA followed by Northern blot validation.

STncP1-1 is visualised in IGB (right panel) and validated by Northern blot (left panel). In the IGB screenshot each lane indicates a different library (growth conditions – Table 2.3). The vertical axis in each library indicates the number of normalized cDNA reads (scale 0 – 20 reads). The orientation of the newly predicted sRNA (blue) and its adjacent flanking genes (black) are also indicated as arrow marks. The bent arrow indicates the transcriptional start site of STncP1-1.

RNA-Seq based transcriptome analysis (*Section 5.2.2*) of *S.* Typhimurium 4/74 revealed expression of a transcript located in the intergenic region between STM474_p1098A and STM474_p1099 (Figure 6.1). These adjacent genes are transcribed divergently and separated by 892 bp. STM474_p1098A encodes a hypothetical (69 aa) protein of unknown function, and does not contain any protein domain that is conserved within *Salmonella* serovars (data not shown). The STM474_p1098 gene, also known as *pefB*, encodes a 100 aa length protein which is required for plasmid-encoded fimbrial production (Bäumler and Heffron 1995). *pefB* is the first gene of the *pef* operon (*pefBACDI*) which mediates adhesion to murine intestinal epithelial cells and is critical for fluid accumulation in the mouse (Bäumler et

al. 1996). An ORF prediction search was performed on the transcribed region located between STM_p1098A and *pefB* using NCBI ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The search results did not identify any ORF, suggesting that the transcript could play a regulatory role as a non-coding small RNA. The expression of the newly identified sRNA STncP1-1 was experimentally validated by Northern blot (Figure 6.1) in the LowMg^{2+} growth condition.

STncP1-1 sRNA is transcribed in conditions mimicking stress faced by the bacterium during the infection process including LowMg^{2+}, SPI2 inducing conditions, and temperature shift from 25^{o}C to 37^{o}C (Supplementary Dataset 5.1). A significant reduction in expression was observed during the growth phases EEP, MEP, LEP, ESP & LSP in nutrient rich medium (Figure 6.1). The coordinates of STncP1-1 and its flanking genes are presented in Table 6.1.

Table 6.1: Coordinates of STncP1-1 and its adjacent flanking genes.

Gene ID	Strand	Start	End	Length
STncP1-1	+	78,588	78,694	107*
Flanking genes				
STM474_p1098A	-	78,078	78,287	210
STM474_p1099 (*pefB*)	+	79,178	79,480	303

To investigate the conservation of STncP1-1 at the sequence level in other *Salmonella* serovars, we performed BLAST (NCBI BLAST) search using STncP1-1 against the NCBI non-redundant database. BLAST results revealed that the sequence of STncP1-1 was only conserved in the serovars of *Salmonella enterica,* including Bovismorbificans, Paratyphi, Choleraesuis and Enteritidis which harbour the *Salmonella* virulence plasmid. A multiple sequence analysis (MSA) was conducted using BLAST results of STncP1-1 (Figure 6.2). The MSA results revealed the absence of the Pribnow box (-10) in other serovars, suggesting that this sRNA may only be transcribed in *S*. Typhimurium.

Figure 6.2: Sequence conservation of STncP1-1.
Alignment of STncP1-1 sRNA Identified in different *Salmonella enterica* serovars using computer-based searches. Predicted -10 and -35 boxes are coloured with yellow and green, respectively. The left and right bent arrows indicate the start and end coordinates of STncP1-1. The predicted Rho-independent terminator is underlined in black. pSLT – *S*. Typhimurium 474; pSLT-BT – *S*. Typhimurium D23580; pSLA5 – *S*. Enteritidis LA5; pVIRBov – *S*. Bovismorbificans 3114; pSCV50 – *S*. Choleraesuis SC-B67; pSCV – *S*. Paratyphi C RKS4594.

TraDIS (transposon-directed insertion-site sequencing) is an exhaustive transposon based mutagenesis assay for identification of bacterial gene function and virulence-attenuation phenotypes in animals (Langridge et al. 2009). Recently, the TraDIS technique was used to identify the genes of *S*. Typhimurium 4/74 that are required for colonisation in the gastro intestinal tract of chicken, cattle, pigs and mouse (Chaudhuri et al. 2013). This TraDIS global mutagenesis experiments showed virulence-attenuated phenotypes for transposons that mapped within STncP1-1, which prompted us to study the virulence and regulatory roles of this sRNA. The transposon insertion sites within STncP1-1 and fitness scores of attenuated phenotypes in different animals are presented in Table 6.2.

Table 6.2: TraDIS significant virulence-attenuated phenotypes of STncP1-1.

Transposon insertion sites		Attenuated phenotypes fitness scores					
Position	Strand	Chicken		Pig		Calf	
		scores	P-value	score	P-value	score	P-value
78,588	+	-2.67	0.01	-7.30	$2.16E^{-10}$	-15	$3.19E^{-25}$
78,624	+	-15	$2.88E^{-12}$	---	---	-15	$6.44E^{-25}$

The cut-off for TraDIS fitness scores that show significant attenuation are chicken < -2, pig < -2, and calf < -3 (Chaudhuri et al. 2013).

6.2.1.2 *Putative role of STncP1-1 in Salmonella virulence*

Epithelial invasion, fold replication in macrophages and competitive inhibition in mice model are different procedures to test virulence phenotypes in *Salmonella* (Beuzón and Holden 2001; Fàbrega and Vila 2013). Different sRNA mutants were shown to have reduced virulence phenotypes in the BALB/c mice model and during intra-macrophage infection (Santiviago et al. 2009; Gong et al. 2011). Santiviago et al. (2009) showed that the virulence of mutants of IstR and SroA was attenuated within BALB/c mice. The IsrM sRNA is required for epithelial invasion, intra-macrophage replication and growth in the spleen/ileum of mice (Gong et al. 2011). Due to attenuated phenotypic results from TraDIS, It was of interest to determine if a strain that lacked the plasmid-borne sRNA STncP1-1 was attenuated for virulence in macrophage and mice. These experiments were done in collaboration with Dr. Shabarinath Srikumar, Hinton lab, TCD, Ireland and Prof. Daoguo Zhou, Purdue University, USA.

An in-frame deletion mutant of STncP1-1 in ST4/74 was created using the protocol described (*Section 2.5*). To determine the role of STncP1-1in the intra-macrophage survival of *Salmonella,* we conducted a Gentamycin protection assay in macrophage (RAW 264.7) as explained in Eriksson et al. (2003) with minor modifications. The level of intracellular replication was measured using the fold replication which is defined as the number of intracellular bacteria that grow during the infection experiment. The fold replication of the 4/74-ΔP1-1 strain was tested within macrophages in comparison with the wild type 4/74 as a positive control and 4/74-Δ*rpoE* as a negative control (Humphreys et al 1999). After 15.30 hours of intracellular growth, the wild type and Δ*rpoE* had fold replication of 48 and 8 respectively. The STncP1-1 mutant showed a

hyper-virulent phenotype, with a fold replication of 66, higher than the wild-type (Figure 6.3).

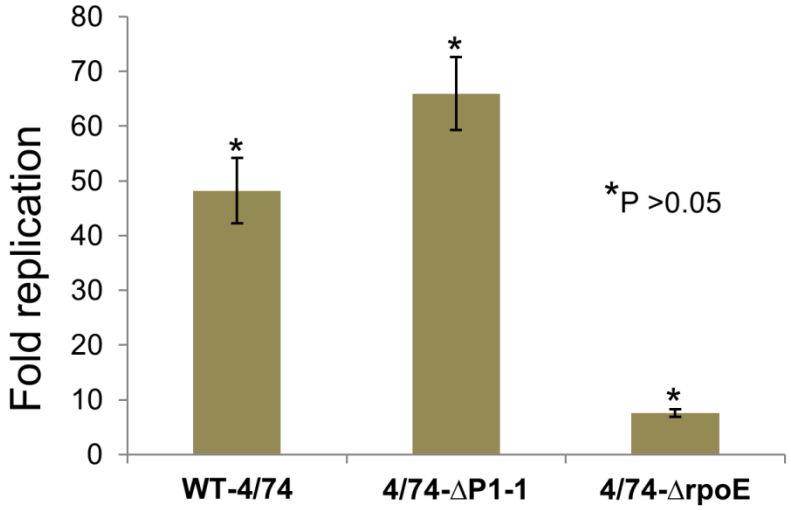


Figure 6.3: Intra-cellular proliferation of 4/74-ΔP1-1 in comparison to wild type.
The fold replication of 4/74-ΔP1-1 is calculated by a gentamycin protection assay. Wild type 4/74 was used as a positive control and ST4/74-*ΔrpoE* as a negative control. Macrophage experiments were done by Dr. Shabarinath Srikumar, Hinton lab, TCD, Irelan.

To study the virulence role of STncP1-1 in *Salmonella*, we infected BALB/c mice intra-peritoneally (5 X 10^3 CFU ml^{-1} per mouse) and considered the mean time to death as the determinant of virulence. The mice (n=3) were infected with 4/74ΔP1-1 (test strain), wild type 4/74 (positive control) and STZP18, a *S.* Typhimurium *ssaV* mutant (negative control) (Shea et al. 1999). All animals injected with wild-type 4/74 died within 5 days of injection while the STZP18 did not kill the mice (Figure 6.4). Mice infected with 4/74ΔP1-1 killed all the infected mice within 3 days of infection, suggesting that this sRNA mutant could be hyper-virulent. The result produced by the mouse experiment was consistent with the intra-macrophage assay. To confirm these more-virulent phenotypes, genetic complementation experiments will be required. However, time limitations prevented the construction of a STncP1-1 complemented strain.

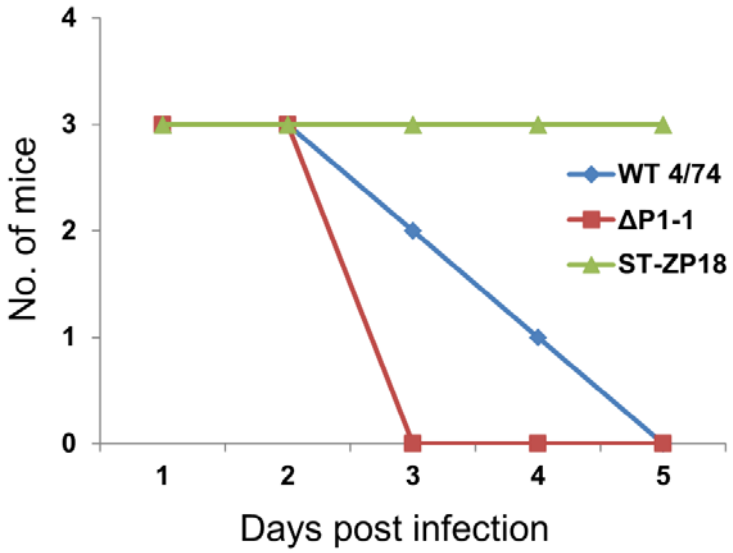


Figure 6.4: Virulence phenotype of STncP1-1 in murine model.
The virulence of STncP1-1 was calculated by taking mean time to death of mice. Wild-type 4/74 and ST-ZP18 (Δ*ssaV*) used as the positive and negative controls respectively. Mouse experiments were done by Prof. Daoguo Zhou, Purdue University, USA.

Hyper-virulent sRNA mutants have been observed earlier in *Salmonella.* For example, an OxyS mutant was shown to be hyper-virulent in mice (Santiviago et al. 2009). OxyS coordinates oxidative stress is induced by peroxide stress (Zhang et al. 1997) and is highly up-regulated in macrophages (Dr. Shabarinath Srikumar per. comm., Hinton lab, TCD). Two other recently identified *Salmonella* sRNAs; STnc1940 and STnc3080, (Kröger et al. 2013), were highly up-regulated in macrophages (RNA-Seq data) and were also shown to be more-virulent in macrophages (Dr. Shabarinath Srikumar per. comm., Hinton lab, TCD).

Gonzalo-Asensio et al. (2013) identified a pSLT^{SL1344}-encoded asRNA lesR-1, expressed only when *Salmonella* persist inside the host cell. Authors showed that *Salmonella* that lack the lesR-1 asRNA showed ~3-fold increased intracellular proliferation in human (BJ-5ta) fibroblast cells. Although a mechanism is not determined yet, it seems plausible that expression of STncP1-1 prevents the bacterium from hyper-replicating and the consequent threat to the host cell viability. However, much more detailed analysis such as plasmid-based over-expression of the sRNA within an intra-macrophage environments, is required to establish a viable model for the sRNA based control of hyper-replication.

6.2.1.3 *Identification of STncP1-1 mRNA targets using pulse-overexpression experiments*

In general, sRNAs play regulatory roles in bacteria by either repressing or stabilizing the target mRNAs (Li et al. 2012; Desnoyers et al. 2012). The repression or stabilization of target mRNA by sRNAs is aided by direct binding between sRNA and target mRNA, and the lack of an sRNA can abolish the functional role of an sRNA on target mRNA (Sharma et al. 2007; 2011). For example, in the case of SgrS-mediated activation of the *yigL* gene (Papenfort et al. 2012), the activation of *yigL* was arrested by deletion of SgrS and restored by re-introducing the pBAD (expression vector) containing SgrS. The putative hyper-virulence of 4/74-ΔP1-1 in the macrophage and murine assays prompted us to study the functional role of this sRNA. To identify the mRNA targets of STncP1-1 in *Salmonella*, we used a pulse over-expression approach (*Section 2.6*) and observed global changes in mRNA abundance with microarrays after transient over-expression of the pBAD containing STncP1-1 (pBAD-P1-1) in a 4/74-ΔP1-1 mutant. Figure 6.5A depicts the overview of pulse-overexpression followed by microarray procedure. The RNA abundance of the 4/74-ΔP1-1 mutant containing pBAD-P1-1 was compared with 4/74-ΔP1-1 containing an empty control vector (pKP-8-35; which expresses a nonsense sRNA).

Furthermore, from the published studies, sRNA expression in wild-type bacteria is induced under conditions where their regulatory function was required (Durand et al. 2010). Incorporating the RNA-Seq data of 22 environmental conditions (*Chapter 5*; except MAC), the target identification analysis was conducted in the two conditions where highest wild type expression of STncP1-1 was observed (Supplementary Dataset 5.1). The strains were grown in LowMg^{2+} and following a temperature shift from 25^{o}C to 37^{o}C for 10 min growth conditions (Temp10). The expression of sRNA was induced by adding L-arabinose for 10 minutes. Total RNA was extracted, labelled cDNA created and hybridised to the SALSIFY array (*Section 2.6*) along with labelled genomic cDNA as common reference. The Microarray data were normalised and compared with controls as described in *Section 2.6.2*.

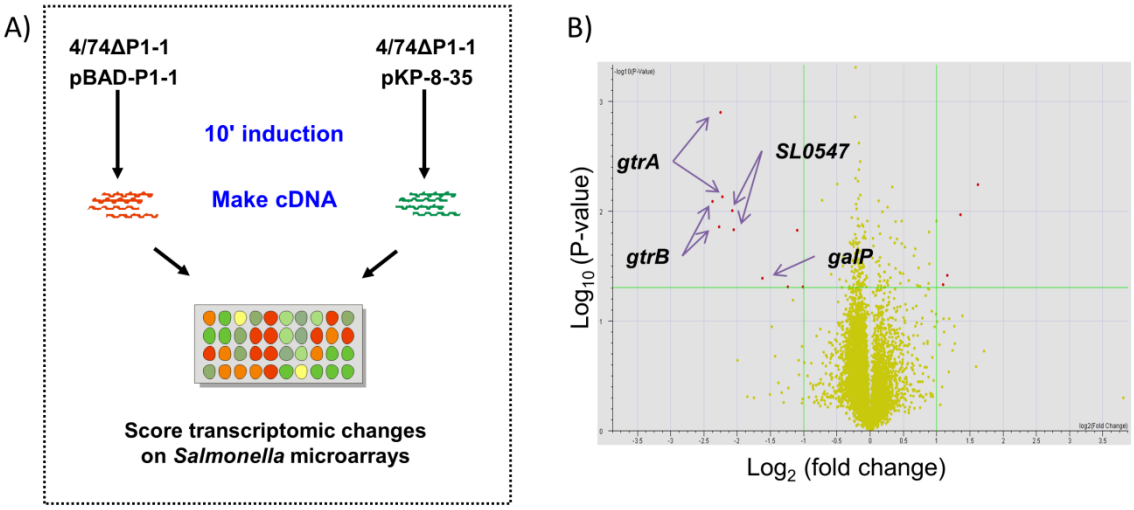


Figure 6.5: Identification of STncP1-1 targets using pulse-overexpression followed by microarray experiments.

A) Overview of pulse-over expression methods. B) Volcano plot showing targets of STncP1-1 in LowMg^{2+} growth conditions. Fold change ≥ 2 and P-value < 0.05 set as a threshold; green vertical bars indicate the fold change thresholds and the green horizontal bar indicates the P-value threshold. The yellow dots represent probes that lie within at least of the threshold (P-value) and red spots represent probes that lie outside the thresholds (fold change) and thus represent candidate target mRNAs (n = 2).

To identify genes that showed statistically significant fold change in their mRNA levels a Volcano plot analysis was performed using GeneSpring 7.3 (Agilent) (Figure 6.5B). The P-value reflects the extent of the fold change as well as the reproducibility between biological replicates and thus the significance of the sRNA expression. Probes that satisfied thresholds, specifically fold change > 2 in mRNA abundance and P-value ≤ 0.05 were considered as candidate target transcripts. The putative mRNA targets of STncP1-1 identified are listed in Table 6.3.

Global transcriptomic analysis showed that the *gtrA, gtrB* and STM474_0576 genes are down-regulated by 4-fold upon overexpression of STncP1-1 in LowMg^{2+} growth conditions. These three genes are located with the Salmonella Pathogenicity Island 16 (SPI-16) in *Salmonella* Typhimurium (Vernikos and Parkhill 2004). The GtrA and GtrB proteins have been previously described to be involved in serotype conversion through O-antigen glycosylation mediated by bacteriophages (Guan et al. 1999; Mavris et al. 1997). The STM474_0576 gene encodes a putative inner membrane protein which is critical for faecal shedding and intestinal persistence of *S*. Typhimurium (Bogomolnaya et al. 2008). The *galP* and *glxR* genes are ~2 fold down-regulated while overexpresing the STncP1-1 in 4/74-ΔP1-1 mutant in LowMg^{2+} and Temp10 growth conditions,

respectively. The GalP protein is involved in the galactose transport system (Postma 1977) and a study by Park et al. (2010) showed that GlxR plays a role as a global regulator controlling both carbon catabolite repression and acetate metabolism in *Cyanobacteria*. The transcription elongation factor *greB* was ~2 fold down-regulated in the Temp10 growth condition. The *greB* and *greA* (both are present in *Salmonella*) are paralogous genes that are involved in the cleavage of backtracked (reverse translocation at paused site of the elongation RNAP complex) RNAP complexes (Borkhov et al. 1993) in both *in vitro* and *in vivo* (Marr and Roberts 2000) conditions. In our study, we did not observe any up- or down-regulation of *greA* by STncP1-1. Detailed experimental verifications are needed to conclude the STncP1-1 regulation on *greB* and difference between *greB* and *greA* regulation by STncP1-1.

Table 6.3: Targets of STncP1-1identified in two different growth conditions.

Growth conditions	Target genes	Fold change[#]	Description
LowMg^{2+}	*ybbV*	3.08	hypothetical protein
	ygdB	2.62	competence gene
	STM474_2858	2.56	hypothetical protein
	STM474_2796	2.29	hypothetical protein
	ygcY	2.24	probable glucarate dehydratase
	gtrA	-4.53	bactoprenol-linked glucose translocase
	gtrB	-4.78	bactoprenol glucosyl transferase
	STM474_0576	-4.48	hypothetical membrane protein
	galP	-2.04	galactose-proton symport
Temp10	STM474_0992	4.09	hypothetical protein
	yciH	2.09	hypothetical translation initiaiton factor
	aceF	2.08	pyruvate dehydrogenase multienzyme
	yebN	-2.36	hypothetical membrane protein
	greB	-2.59	transcription elongation factor GreB
	glxR	-2.70	Tartronic semialadehyde reductase
	STM474_1806	-3.01	hypothetical protein

[#]Fold change of transcript abundance upon transient over-expression of STncP1-1, compared to transient induction of nonsense RNA, ≥ 2-fold and P-value < 0.05 used as a cut-off.

Of the 8 up-regulated genes in both growth conditions, 6 are annotated as hypothetical proteins (unknown function). The genes with known function including *ygdB* and *aceF,* were 2-fold up-regulated in either one of the growth conditions. The gene *ygdB* is co-expressed with *ppdABC* as an operon and is involved in assembly and transport of the

type IV fimbriae (Whitchurch and Mattick 1994). But the differential expression of the *ppdA, B, and C* genes was not observed in both conditions, suggesting the up-regulation of the *ygbB* gene by STncP1-1 may be an artefact. The *aceF* gene encodes one of the proteins of the pyruvate dehydrogenase complex (Texter et al. 1988). Further validation is needed to confirm the *aceF* regulation by STncP1-1.

6.2.1.4 *Identification of candidate mRNA targets with an RNA-Seq approach*

For transcriptomic analyses, the RNA-Seq approach has distinct advantages over microarray based techniques. RNA-Seq inherently provides single nucleotide resolution, while microarrays exhibit cross-hybridization and have a limited dynamic range of detection (Nagalakshmi et al. 2008). Recent evidence showed that RNA-Seq increases the specificity and sensitivity in differential gene expression studies (Agarwal et al. 2010). Sixteen putative mRNA targets of STncP1-1 were already identified using microarray based transcriptomic studies. To identify additional candidate targets of STncP1-1 which were not detected by microarray, we conducted an RNA-Seq based experiment. This approach has proved successful for the identification of sRNA target in other system (Sharma et al. 2011). The transcriptome of the 4/74-ΔP1-1 mutant was compared with the 4/74 wild-type transcriptome. Figure 6.6 explains the overview of RNA-Seq based target identification. Total RNA was extracted from the 4/74-ΔP1-1 and wild-type bacteria grown in InSPI2 condition (PCN; Löber et al. 2006) as described in *Section 2.8*.

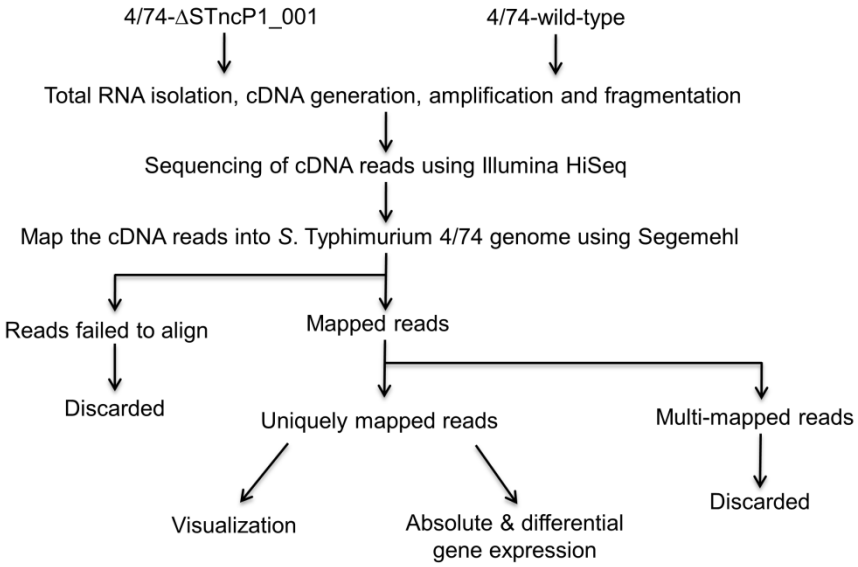


Figure 6.6: Workflow of RNA-Seq based target identification of STncP1-1.
The flow chart explains the STncP1-1 target identification methods using an RNA-Seq approach.

A total of ~61 million cDNA reads were sequenced (combining both samples) with a read length of 101 bases (Supplementary Dataset 3.2; libraries name 'Wild-type (InSPI2)' and 'ΔSTncP1-1 (InSPI2)' for wild-type 4/74 and 4/74-ΔP1-1, respectively). The sequence quality of cDNA reads was checked using FastQC as described (*Section 3.2.2.1*). Both datasets had good quality values for the first 60 bases at the 5' end, and then a reduction in quality for the remaining bases (data not shown). The 41 bases at the 3' end of the cDNA reads were truncated in one step and the rest aligned to the reference genome (*S.* Typhimurium 4/74) as described (*Section 3.2.2.2*). The reads that failed to map were sequentially trimmed by two nucleotides from the 3' end of the read and then mapped again (Step-wise methods; *Section 3.2.2.3*). Approximately 95% of cDNA reads were mapped to the *Salmonella* genome and the mapping statistics are presented in Figure 6.7. The reads that mapped uniquely to the genome were formatted as visualisation files for IGB (*Section 3.2.2.5*). No cDNA reads were observed within the transcribed region of STncP1-1 in the 4/74-ΔP1-1 RNA-Seq library, confirming that the expression of STncP1-1 was ablated by deletion of the sRNA (86 reads at this location in wild-type). The relative expression of genes was calculated by comparing absolute expression values (TPM) between mutant and wild-type as described (*Section 3.2.2.6*).

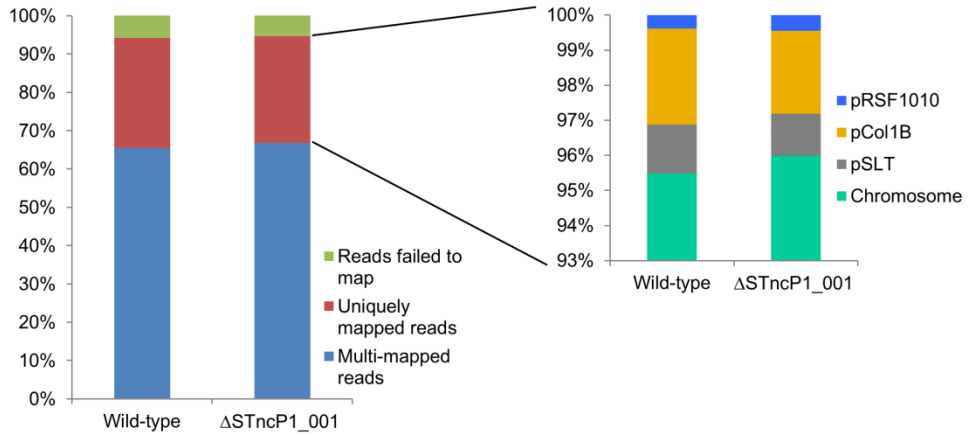


Figure 6.7: Mapping statistics of cDNA reads from 4/74-wild-type and 4/74-ΔP1-1 RNA-Seq libraries.

The bar charts showing percentage of multi and uniquely mapped cDNA reads from both RNA-Seq libraries. The cDNA reads mapping uniquely to the genome were further categorized as reads mapping to the chromosome and three plasmids.

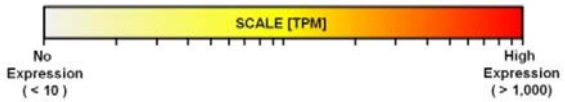
The RNA-Seq experiment identified 46 genes (both coding and non-coding) that were differentially-expressed (≥ 3-fold) in the 4/74-ΔP1-1 mutant compared with wild-type 4/74 (Table S6.1). The $pSLT^{4/74}$ plasmid-encoded gene *spvR* was > 4-fold down-

regulated in the 4/74-ΔP1-1 mutant. SpvR is the LysR family regulatory protein which positively regulates the expression of the *spv* operon (*spvABCD*) (Krause et al. 1995). However, differential expression of structural members of the *spv* operon (*spvABC*) was not observed, suggesting that further experimental verification is needed to confirm the down-regulation of *spvA* in the 4/74-ΔP1-1 mutant. Two cold shock proteins, *capA* and *capB,* were up-regulated (18-fold and 7-fold, respectively) by the 4/74-ΔP1-1 mutant, but differential expression of other *cap* genes was not observed. Five sRNAs including a pCol1B$^{4/74}$-encoded asRNA (474P2AS_120) were > 3-fold down-regulated in the 4/74-ΔP1-1 mutant (Table S6.1). It should be noted that we can not assign any significance values to the results from the RNA-Seq experiment because there were no biological replicates.

Table 6.4: Up-regulation of *gtr* locus in the 4/74-ΔP1-1 mutant identified by RNA-Seq.

Systemic ID	Common name	Absolute expression values (TPM)		Fold change (ΔP1-1/wt)
		Wild-type	4/74-ΔP1-1	
STM474_0576		21.1	116.6	5.9
STM474_0578	*gtrAa*	6.7	58.0	5.8
STM474_0577	*gtrBa*	7.1	35.3	3.5

Absolute expression columns are coloured according to expression levels.



Many of the differentially-expressed genes from the RNA-Seq experiment could be the results of indirect effects. To determine the most likely mRNA target, I searched for correlation between RNA-Seq and pulse over-expression data. Interestingly, the RNA-Seq based target identification results for the *gtr* locus (Table 6.4) were consistent with the pulse over-expression based target predictions. In pulse over-expression, the *gtr* locus showed ~4-fold down-regulation in the pBAD-STncP1-1 over-expressed sample compared to teh pKP-8-35 sample (empty control vector) in the 4/74-ΔP1-1 mutant strain. The RNA-Seq data showed a ~5-fold up-regulation of the *gtr* locus in the 4/74-ΔP1-1 strain compared with the 4/74-wild-type strain.

In summary, the results from both target identification methods suggest that the STncP1-1 sRNA represses the expression of the *gtr* locus in infection-relevant growth conditions (InSPI2 and LowMg^{2+}). The *gtr* locus (*gtrA, gtrB* and STM474_p0576; SPI-16) expresses as a polycistronic operon (Guan et al. 1999). In *Shigella flexneri* bacteriophage X (SFX), the *gtrA* and *gtrB* genes encode proteins responsible for the serotype conversion of the infected bacterium by adding glucose residues to repeating O-antigen subunits of lipopolysaccharide (LPS) (Guan et al. 1999; Mavris et al. 1997). The STM0557 gene (orthologue of STM474_0576; Sequence identity 100%) encodes a putative inner membrane protein in *S*. Typhimurium LT2 and is required for intestinal persistence in mice (Bogomolnaya et al. 2008).

The putative functional role of STncP1-1 on the *gtr* locus could explain the hyper-virulent phenotype of 4/74-ΔP1-1 in virulence experiments (Figure 6.3 and 6.4). Bogomolnaya et al. (2008) showed that the *gtr* locus was involved in the O-antigen glucosylation. The *gtr* locus in *S. flexneri,* is shown to modify the LPS, influencing virulence by enhancing the function of the type-III secretion system (West et al. 2005). The Δ*gtr* mutant (deletion of all three genes) of *S*. Typhimurium and *S*. flexneri showed defective colonization of epithelial cells (West et al. 2005, Bogomolnaya et al. 2008). The data produced in *Section 6.2.1* support the hypothesis that the deletion of STncP1-1 enhances the modification of O-antigen by increasing the expression of the *gtr* locus and causing the hyper-virulence phenotype.

The genome of *S*. Typhimurium 4/74 encodes a second cluster of the *gtr* locus (paralogous genes, STM474_4396, 4397 and 4398; Richardson et al. 2011). The expression of this cluster was not altered in either target identification experiment, suggesting that the regulation of STncP1-1 is specific to the first cluster of the *gtr* locus. This specific role of STncP1-1 and the virulence-related function of the *gtr* locus promoted further experimentation to look for direct binding between *gtrA* (first gene of the *gtr* operon) and STncP1-1.

6.2.1.5 *Target validation using GFP reporter fusion assay*

Most validated sRNA-mRNA interactions occur in close proximity to the start codon or SD sequence of the target mRNA (Sharma et al. 2007; 2011). The direct binding between sRNA and target mRNA often leads to the blockage of the translation process and induction of transcript degradation (Bandyra et al. 2012). To validate the direct interaction with the putative targets (*gtrA*) of STnaP1-1, we performed an sf-GFP-based two-plasmid assay in a heterologous system (*E. coli* TOP10 cells; *recA*$^-$) as described (*Section 2.7*). The entire 5' untranslated region (5'UTR) and 99 bases (33 aa) of the coding sequence (CDS) of *gtrA* was translationally-fused to the sf-GFP protein (Figure 6.8A). The plasmid pXG10-sf with *gtrA::gfp* along with either plasmid pP_L containing STncP1-1 or pJV300 (control) was transformed into *E. coli* TOP10 cells. Any degradation of target mRNA, initiated by binding of the sRNA with the target region can be observed by a reduction in level of fluorescence. The nutrient agar plates were incubated overnight and fluorescence was visualised (Figure 6.8B). The Colony fluorescence assay revealed that the *gtrA::gfp* fusion showed relatively low level of fluorescence. It is not possible to observe the reduction in fluorescent levels when the *gtrA::gfp* fusion showed such a low level of fluorescence.

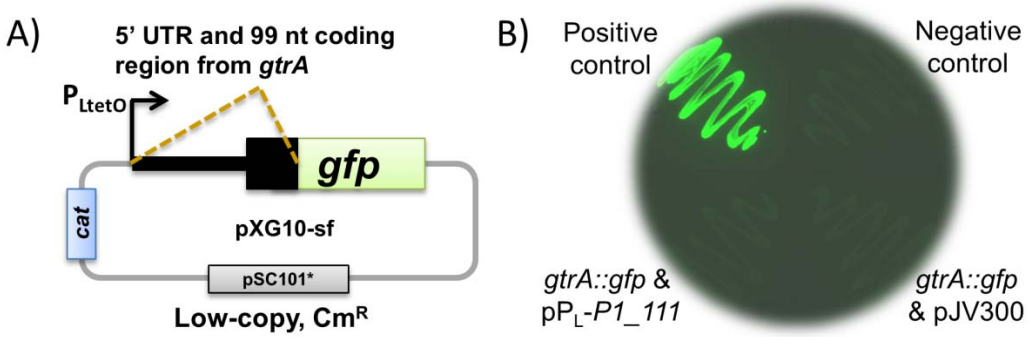


Figure 6.8: Validation of STncP1-1 putative target *gtrA* using GFP reporter fusion assay. A) Construction of pXG10-sf *gtrA*::*gfp* fusion plasmid. B) Colony fluorescence imaging assay on nutrient agar plates. Assay of colony-fluorescence (two-plasmid assay) of E. coli TOP10 cells carrying *gtrA::gfp* fusion plasmid and each fusion with either empty control vector (pJV300) or STncP1-1 constitutive expression plasmid (pP_L-STncP1-1). Positive control - *fliC::gfp* and negative control – *sspH2::gfp*.

Since the fluorescence level was not high enough for visual detection, flow cytometric analysis was conducted to provide a higher level of sensitivity. *E. coli* strains carrying *gtrA::gfp* fusion plasmids in combination with the control vector pJV300 or STncP1-1 constitutive expression vector (pP_L-STncP1-1) were grown to stationary phase and subjected to flow cytometric analysis. However, FACS experiments failed to show

altered fluorescence levels between *gtrA*::*gfp* with pP$_L$-STncP1-1 or with pJV300, because *gtrA::gfp* fusions showed very low levels of fluorescence (Figure 6.9).

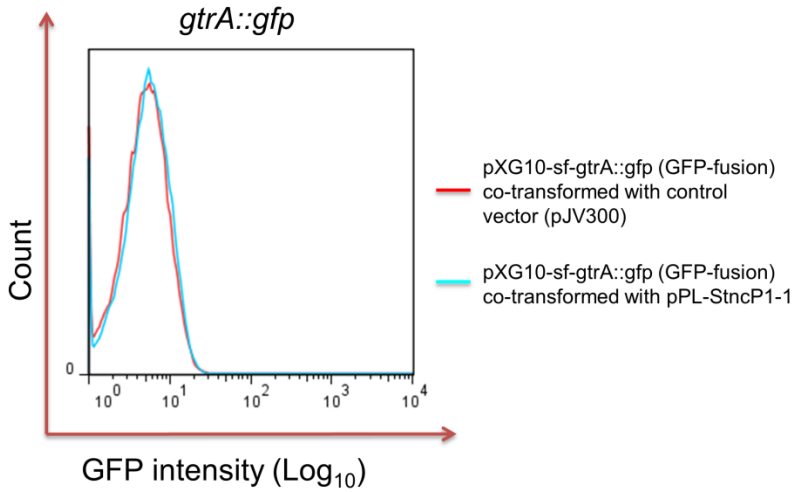


Figure 6.9: FACS-based reporter assay.
Flow cytometric analysis of *E. coli* (TOP10) *recA*$^-$ strains carrying *gtrA::gfp* fusion plasmids in combination with control vector pJV300 or STncP1-1 constitutive expression vector (pP$_L$-STncP1-1). The events measured (20,000 events) from the assay are plotted in a histogram (vertical axis). Cellular fluorescence is given in arbitrary units (horizontal axis).

The low level of GFP fluorescence could be due to impaired folding of GFP protein (Corcoran et al. 2012). If the GFP failed to fluoresce, then the level of GFP protein can be determined by Western blot experiment. The western blotting with an anti-GFP antibody provides a more sensitive measure for the quantification of fusion protein levels (Urban and Vogel 2007). However, time limitations prevented the validation of the STncP1-1 candidate target (*gtrA*) using Western blot.

6.2.2 *Cross talk between plasmid-encoded asRNAs and genes from chromosome*

The SALSIFY microarrays were used to identify the potential mRNA targets of additional plasmid encoded antisense RNAs. Four asRNAs expressed from the plasmid pSLT$^{4/74}$ were chosen (474P1AS_50, 80, 110 and 180) for target identification, which required a new strategy. The optimal way to identify the mRNA targets is to over-express a sRNA in a mutant background that lacks the sRNA. However, the deletion of an asRNA would also prevent expression of the sense coding gene (polar effect). An alternative strategy was used that involved the pulse over-expression of each asRNA in a growth condition associated with low asRNA expression. The lowest expression

levels of each asRNA were identified from our RNA-Seq data for 4/74 from 22 different environmental conditions. The asRNA pulse over-expression, followed by microarray and data analysis was done by the same approach as the experiments of STncP1-1 (*Section 6.2.1.3*). Probes satisfying thresholds, including fold change of ≥ 2 in mRNA abundant and P-value of ≤ 0.1 were considered as candidate target transcripts. Candidate targets of four asRNAs were identified and are listed in Table S6.1.

A number of chromosomally-encoded mRNAs were identified as candidate targets of plasmid-encoded antisense RNAs using pulse-over expression experiments. For example, the genes involved in the glutamine transport system, including *glnQ, glnH* and *glnP* were 3-fold up-regulated by the plasmid-encoded asRNA 474P1AS_50 in growth at 25 ^{o}C, and genes involved in the K^{+} transport system, including *kdpA*, *kdpB* and *kdpC* were 2-fold up-regulated by the same asRNA at LowMg^{2+}. Further experimentation, including a two-plasmid assay is required to validate these candidate mRNA targets of 474P1AS_50. Interestingly, microarray analysis revealed that the genes involved in flagellin production, such as *fliC, fliH, fliD, & fliS* were 2-fold up-regulated by the 474P1AS_80 asRNA which is transcribed antisense to *spvA*. In the future an experiment is needed to determine whether 474P1AS_80 asRNA does impact on the function of the *fli* operon. Kim and Kwon (2013) identified an increased motility phenotype in RyhB-2 deletion mutants and validated direct interaction of RyhB-2 with transcripts of motility (*fliF* and *fliJ*) and a chemotaxis gene (*cheY*). However, the experiments to observe the motility capability of Salmonella when over-expressing 474P1AS_80 did not reveal any swimming differences suggested that the 474P1AS_80-mediated role of motility genes could be an indirect effect (data not shown).

As another example of a plasmid-encoded asRNA, was recently discovered to regulate chromosomal genes (Prof. Josep Casadesus Per. Comm., University of Seville, Spain). FinP asRNA is complementary to the RBS of the *traJ* mRNA which encodes an activator for the conjugal (*tra*) operon. Thus, binding of FinP blocks the *traJ* mRNA translation. Recently, it was found that FinP also targets chromosomally-encoded mRNAs, such as *cysD, ygaE, oppB* and a putative cytoplasmic protein *STM4302* (Prof. Josep Casadesus pers. comm., University of Seville, Spain).

6.3. *Conclusion*

Using the RNA-Seq technique, we have identified more antisense RNAs and sRNA than ever reported before for plasmids (*Section 4.2.4* and *5.2.2*). The putative virulence phenotype of STncP1-1 was identified using macrophage infection and mice model experiments. The candidate mRNA target of STncP1-1, the *gtr* locus, was identified using a pulse-over expression experiment and confirmed using an RNA-Seq approach. The putative regulatory function on chromosomally-encoded genes indicates that these newly identified sRNA/asRNAs could play important regulatory roles in *Salmonella* adaptation, survival and virulence. However, the experiments presented in this chapter are not definitive and have not proven that any plasmid-encoded sRNA or asRNAs do regulate chromosomal gene expression.

Crosstalk between the *Salmonella* core genome and the *Salmonella* genes acquired by horizontal gene transfer (HGT) has been seen before. One of the classical examples is the interaction between InvR and *ompD*. The InvR sRNA is transcribed within the *Salmonella* pathogenicity island-1 (SPI-1), and represses the synthesis of the abundant OmpD porin protein encoded by the *Salmonella* core genome (Pfeiffer et al. 2007). InvR binds the RNA chaperone Hfq with high affinity, and Hfq stabilises the interaction between InvR and *ompD*. Recently, Papenfort et al. (2012) discovered that the SgrS sRNA encoded by the *Salmonella* core genome reduces the expression of the horizontally acquired gene that encodes the virulence factor SopD. These precedents suggest that the control of chromosomal genes by plasmid sRNA/asRNAs are worthy of further study. It is hoped that future work will definitively identify a plasmid-encoded sRNA/asRNA that controls a transcriptional network of chromosomal genes, and define its regulatory mechanism.

Table S6.1: Putative targets of STncP1-1 in InSPI2 growth conditions identified by RNA-Seq.

The absolute expression (TPM) and relative expression values were calculated as described (*Section 3.2.2.6*).

If the TPM value of the gene was less than 10, first adjusted to 10 (cut-off value), then the relative expression was calculated (*Section 3.2.2.6*).

Systemic ID	Common name	Absolute expression values (TPM)		Fold change (∆P1-1 /wt)	Descriptions
		Wild-type	∆P1-1		
Up-regulated genes					
STM474_3821	*cspA*	229.7	4221.1	18.4	cold shock protein
STM474_2028	*cspB*	18.6	133.8	7.2	cold shock protein
STM474_3438A	*yrbN*	50.0	300.6	6.0	conserved sORF
STM474_0576		12.1	71.4	5.9	hypothetical membrane protein
STM474_0578	*gtrAa*	6.7	58.0	5.8	bactoprenol-linked glucose translocase
STM474_1084	*orfX*	8.2	57.4	5.7	conserved hypothetical protein
STM474_3438	*deaD*	21.1	116.6	5.5	ATP-dependent RNA helicase (dead-box protein)
STM474_3006	*hilC*	34.8	156.4	4.5	AraC-family transcriptional regulator
STM474_2759	*yfiD*	27.8	123.3	4.4	conserved hypothetical protein
STM474_0577	*gtrBa*	7.1	35.3	3.5	bactoprenol glucosyl transferase
STM474_2156	*hisG*	35.7	123.0	3.4	ATP phosphoribosyltransferase
STM474_2157	*hisD*	22.4	74.7	3.3	histidinol dehydrogenase
STM474_2426		26.2	85.1	3.3	hypothetical protein
STM474_4056	*asnA*	244.5	757.2	3.1	asparagine synthetase A
Down-regulated genes					
STM474_4371	*metA*	409.9	58.4	-7.0	homoserine O-succinyltransferase
STM474_1745	*yciG*	47.5	9.2	-4.7	conserved hypothetical protein
STM474_4460	*soxS*	67.2	14.4	-4.7	regulatory protein SoxS
STM474_3820	*yiaG*	218.9	47.1	-4.7	hypothetical HTH-type transcriptional regulator
STM474_1251		233.6	51.7	-4.5	hypothetical heat shock protein
STM474_0507	*htpG*	103.5	22.9	-4.5	heat shock protein HtpG
STM474_4142	*metR*	50.5	11.8	-4.3	trans-activator of metE and metH
STnc3120	STnc3120	45.3	10.6	-4.3	
STnc890	STnc890	450.7	109.7	-4.1	

Continued on next page

Systemic ID	Common name	Absolute expression values (TPM)		Fold change (ΔP1-1 /wt)	Descriptions
		Wild-type	ΔP1-1		
STM474_p1072	*spvR*	43.9	10.8	-4.1	Salmonella plasmid virulence lysR family regulator
474P2AS_120	474P2AS_120	40.5	9.9	-4.0	
STM474_1576	*yddX*	56.7	14.5	-3.9	conserved hypothetical protein
Tpke11	Tpke11	122.8	31.8	-3.9	
STM474_0013	*dnaK*	243.7	63.2	-3.9	Chaperone protein dnaK
STM474_0856	*dps*	344.2	89.9	-3.8	DNA protection during starvation protein
STM474_0623	*ybdL*	121.7	32.7	-3.7	hypothetical aminotransferase
STnc3210	STnc3210	155.2	41.8	-3.7	
STM474_2939	*ygaM*	174.9	50.9	-3.4	conserved hypothetical protein
STM474_4284	*metB*	153.0	45.2	-3.4	cystathionine gamma-synthase
STM474_3129	*ygdI*	300.6	90.5	-3.3	possible lipoprotein
STM474_1268		1173.9	358.6	-3.3	hypothetical protein
STM474_1577	*rpsV*	958.0	294.2	-3.3	30S ribosomal protein S22
STM474_0246	*rof*	395.4	121.7	-3.2	ROF protein
STnc3720	STnc3720	39.2	12.1	-3.2	
STM474_1276	*yeaQ*	396.3	122.3	-3.2	hypothetical inner membrane protein
STM474_1275		290.3	89.6	-3.2	conserved hypothetical protein
STM474_1148	*msyB*	131.1	40.8	-3.2	acidic protein MsyB; multicopy suppressor of SecY
STnc1220	STnc1220	33.3	10.4	-3.2	
STM474_4036	*pstS*	81.4	25.9	-3.1	periplasmic phosphate-binding protein
STM474_0622	*ybdH*	110.8	36.0	-3.1	hypothetical oxidoreductase
STM474_3985	*hslT*	68.7	23.0	-3.0	heat shock protein A
STM474_1830		174.0	58.3	-3.0	conserved hypothetical protein

Table S6.2: Candidate target mRNAs of 474P1AS_50, 80, 110 and 180 identified in two different growth conditions.

Candidate mRNA targets of plasmid-encoded asRNAs were identified using pulse-over expression experiment as described (*Section 2.6*).

Fold change ± 2-fold and P-value < 0.01 used as a cut-off.

*Growth conditions described in Table 2.3 (*Section 2.1.3*).

#Fold change of transcript abundance upon transient over-expression of asRNA (e.g., 474P1AS_50), compared to transient induction of nonsense RNA (*Section 2.6*)

asRNAs	Growth conditions*	Target genes	Fold change#	Description
474P1AS_50	LowMg^{2+}	STM474_2904	29.14	hypothetical protein
		leuB	3.99	3-isopropylmalate dehydrogenase
		leuA	3.68	2-isopropylmalate synthase
		ilvB	3.51	acetohydroxy acid synthase I, small subunit
		kdpC	3.09	K^+-transporting ATPase C chain
		kdpB	3.06	K^+-transporting ATPase B chain
		kdpA	2.81	K^+-transporting ATPase A chain
		ilvA	2.37	threonine deaminase
		STM474_1911	-3.92	hypothetical protein
		STM474_4392	-4.53	hypothetical protein
		STM474_2823	-4.57	hypothetical protein
	25^{o}C	STM474_2713	32.89	hypothetical protein
		STM474_2706	9.12	hypothetical protein
		glnQ	4.12	glutamine transport ATP-binding protein.
		glnH	3.48	glutamine-binding periplasmic precursor
		glnP	3.02	glutamine transport system permease protein.
		sucB	-2.41	dihydrolipoamide succinyltransferase com.
		sucC	-3.33	succinyl-CoA synthetase beta chain
		sucD	-3.07	succinyl-CoA synthetase alpha chain
		IsrA	-3.81	hypothetical protein
		lamB	-4.84	maltoporin precursor

Continued on next page

asRNAs	Growth conditions*	Target genes	Fold change#	Description
474P1AS_80	LowMg^{2+}	*fliC*	2.18	flagellin
		fliH	2.09	flagellar assembly protein
		fliD	2.05	flagellar hook associated protein
		fliS	2.02	flagellar protein FliS
		flgB	2.12	flagellar basal-body rod protein
		flgC	2.33	flagellar basal-body rod protein
		glpT	-2.27	glycerol-3-phosphate transporter
		pduU	-3.39	propanediol utilization protein
	ESP	*yiaM*	11.03	hypothetical membrane protein
		STM474_3715	9.53	hypothetical protein
		sthB	8.17	outer membrane fimbrial protein.
		STM474_2713	7.12	hypothetical protein
		allD	4.18	ureidoglycolate dehydrogenase
		dppB	4	dipeptide transport system permease protein.
		ydhC	3.89	hypothetical integral membrane transport protein.
		ugpC	-4.31	Glycerol-3-phosphate transport ATP-binding protein.
		rfc	-3.73	O-antigen polymerase
		rfbM	-4.04	mannose-1-phosphate guanylyltransferase GDP
		rfbX	-3.01	putative O-antigen transporter
474P1AS_110	LowMg^{2+}	STM474_2904	20.68	hypothetical protein
		gmd	5.48	GDP-mannose 4,6-dehydratase
		STnc140	-2.19	sRNA
		STM474_4392	-2.68	hypothetical protein
	ESP	*sthB*	6.36	Outer membrane fimbrial usher protein.
		allD	4.63	ureidoglycolate dehydrogenase
		bcfB	3.84	fimbrial chaperone
		dppB	3.01	dipeptide transport system permease protein.
		yebN	3.01	hypothetical membrane protein
		glpT	-4.38	glycerol-3-phosphate transporter
		glpQ	-4.24	glycerophosphoryl diester phosphodiesterase
		secG	-2.75	protein-export membrane protein.

Continued on next page

asRNAs	Growth conditions*	Target genes	Fold change#	Description
474P1AS_180	LowMg^{2+}	*nifJ*	2.75	probable pyruvate-flavodoxin oxidoreductase
		amtB	2.54	probable ammonium transporter
		ego	-2.93	hypothetical ABC transporter ATP-binding protein.
		wzxC	-3.12	hypothetical transmembrane transport protein.
		tctD	-3.37	transcriptional regulator
		pduT	-3.23	hypothetical propanediol utilization protein.
		ygcB	-4.79	hypothetical protein
	25^{0}C	STM474_2713	18.22	hypothetical protein
		STM474_2706	16.68	hypothetical protein
		STM474_4610	10.88	hypothetical protein
		gtrA	6.29	bactoprenol-linked glucose translocase
		gtrB	5.64	bactoprenol glucosyl transferase
		STM474_0576	5.22	hypothetical membrane protein
		ompS	3.49	outer membrane protein
		SL0290	-3.51	hypothetical protein
		STM474_2704	-3.7	hypothetical protein
		STM474_2834	-3.8	hypothetical protein
		citF2	-4.48	citrate lyase alpha chain

CHAPTER 7

GENERAL DISCUSSION

Salmonella is one of the most important foodborne zoonotic pathogens, with significant health and economic impact for both humans and animals. The genome of *Salmonella* is characterised by a high genomic plasticity, and mobile elements such as plasmids play an important role in the evolution of *Salmonella* (Moreno Switt et al. 2012). The genes carried by plasmids of *Salmonella* such as antibiotic resistance, toxins and virulence factors, are particularly important in the context of infection biology. In this study, the transcriptional landscape and RNA biology of three plasmids of *S.* Typhimurium 4/74 was investigated using a next-generation sequencing approach.

The thesis describes a bioinformatic-based work flow that was developed for the analysis and visualisation of RNA-Seq data. The pipeline was successfully used to analyse the RNA-Seq data produced by the Hinton lab and other high-throughput data from bacteria other than *Salmonella,* including *Leptospira*, *Acinetobacter*, and *Staphylococcus*. Therefore, the pipeline is applicable for a wide range of RNA-Seq data sets. The modular architecture of the pipeline allows it to be adapted to more advanced read-mapping tools that are developed in the future.

A unique expression profile of plasmid-encoded genes was identified, across 23 environmental conditions. The plasmid core-functional genes were expressed in most growth conditions, including genes involved in replication, partitioning and stability of the plasmids (*Section 5.2.3*). This makes biological sense because the expression of core-functional genes is required for plasmid replication and partitioning in each bacterial generation. The genes encoding virulence traits, such as the *spv* locus, were only expressed in growth conditions relevant to the intracellular survival of the bacterium such as SPI-2 inducing conditions (*Section 5.2.6*). It was clear that certain genes which shared a particular function had similar expression profiles, such as the partitioning systems from pSLT$^{4/74}$ and pCol1B$^{4/74}$ (*Section 5.2.4*). However, the expression patterns of the plasmid post-segregational killing systems encoded by pSLT$^{4/74}$ and pCol1B$^{4/74}$ were significantly different, even though they perform a similar function in both plasmids (*Section 5.2.5*).

Comprehensive dRNA-seq-based mapping identified 162 plasmid-encoded transcriptional start sites (TSS; *Section 4.2.1* and *5.2.2*). During growth at ESP, transcription was initiated at 72% of the TSS by σ^{70} (RpoD; *Section 4.2.6*), the chromosomally-encoded 'housekeeping' sigma factor. This is consistent with the transcription of plasmid-encoded genes being highly expressed under most environmental conditions. An interaction between the chromosome and plasmids of *Salmonella* was identified at a regulatory level. In infection-relevant growth conditions, it was discovered that the expression of the *spv* locus was activated by SPI-2 associated regulators, including OmpR/EnvZ, PhoPQ, SlyA and SsrAB (*Section 5.2.6*).

RNA-Seq of the *S.* Typhimurium plasmids revealed that half of the transcriptional activity involved antisense transcripts (*Section 4.2.2*). We found that all antisense transcripts encoded by pSLT in strain 4/74 were also encoded by the pSLT plasmid in a different sequence type of *S.* Typhimurium, strain D23580 (*Section 4.2.2*). We suggest that pervasive antisense transcription is a common feature of plasmids.

It was found that the plasmids of *Salmonella* encode 39 asRNAs during growth in the various environmental and infection-relevant conditions (*Section 4.2.4*), and the existence of nine asRNAs was confirmed by northern blot. Antisense RNAs were transcribed from the negative strand of genes involved in plasmid replication, stabilization, maintenance and conjugation, suggesting that the core-functions of the plasmid could be regulated by asRNAs. Identification of seven asRNAs within the *pef* locus of $pSLT^{4/74}$, make this region a hot-spot for antisense transcription. The functionality or physiological impacts of asRNAs remains to be determined. Our data show that varying growth conditions modulate the expression of asRNAs, consistent with a regulatory role for some of the asRNAs.

The large number of plasmid-encoded asRNAs raises the question of whether *cis*-asRNA-based gene regulation is more appropriate for the control of plasmid gene expression than the regulatory proteins and *trans*-acting sRNAs that generally modulate chromosomal gene expression. One feature of asRNA-mediated gene regulation is that, by definition, asRNAs have perfect complementarity with their target mRNA. This complementarity could yield stable RNA duplexes, without the involvement of RNA chaperone proteins such as Hfq.

We hypothesize that asRNA-mediated regulation could be a preferred mechanism for the post-transcriptional control of plasmids. It is known that the dysregulation of plasmid-encoded genes imposes a fitness burden upon *Salmonella* cells (Doyle et al. 2007). asRNA-mediated control of plasmid gene regulation would offer the advantage that the regulatory elements are plasmid-encoded, and do not require input from chromosomally-encoded factors.

BIBLIOGRAPHY

Agarwal A, Koppstein D, Rozowsky J, Sboner A, Habegger L, Hillier LW, Sasidharan R, Reinke V, Waterston RH and Gerstein M (2010). Comparison and calibration of transcriptome data from RNA-Seq and tiling arrays. *BMC Genomics*. **11**:383.

Agbaje M, Begum RH, Oyekunle MA, Ojo OE and Adenubi OT (2011). Evolution of Salmonella nomenclature: a critical note. *Folia Microbiol (Praha)*. **56**:497-503.

Ahmer BM, Tran M and Heffron F(1999). The virulence plasmid of *Salmonella* Typhimurium is self-transmissible. *J Bacteriol*. **181**:1364-8.

Ahmer BMM, van Reeuwijk J, Timmers CD, Valentine PJ and Heffron F (1998). *Salmonella* Typhimurium encodes an SdiA homolog, a putative quorum sensor of the LuxR family that regulates genes on the virulence plasmid. *J Bacteriol.* **180**:1185–1193.

Albrecht M, Sharma CM, Dittrich MT, Müller T, Reinhardt R, Vogel J and Rudel T (2011). The transcriptional landscape of *Chlamydia pneumonia*. *Genome Biol*. **12**: R98.

Allison DB, Cui X, Page GP and Sabripour M (2006). Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet*. **7**:55-65.

Altuvia S, Weinstein-Fischer D, Zhang A, Postow L and Storz G (1997). A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell*. **90**:43-53.

Altuvia S, Zhang A, Argaman L, Tiwari A and Storz G (1998). The Escherichia coli OxyS regulatory RNA represses fhlA translation by blocking ribosomal binding. *EMBO J*. **17**:6069-75.

Amara RR and Vijaya S (1997). Specific polyadenylation and purification of total messenger RNA from *Escherichia coli*. *Nucleic Acids Res.* **25**: 3465–70.

Anders S and Huber W (2010). Differential expression analysis for sequence count data. *Genome Biol*. **11**: R106. Epub.

Andersen J and Delihas N (1990). micF RNA binds to the 5' end of ompF mRNA and to a protein from Escherichia coli. *Biochemistry*. **29**:9249-9256.

Andersen J, Forst SA, Zhao K, Inouye M and Delihas N (1989). The function of micF RNA. micF RNA is a major factor in the thermal regulation of OmpF protein in *Escherichia coli*. *J Biol Chem*. **264**:17961-17970.

André G, Even S, Putzer H, Burguière P, Croux C, Danchin A, Martin-Verstraete I, Soutourina O (2008). S-box and T-box riboswitches and antisense RNA control a sulphur metabolic operon of *Clostridium acetobutylicum*. *Nucleic acid Res*. **36**:5955-5969.

Anjum MF, Marooney C, Fookes M, Baker S, Dougan G, Ivens A and Woodward MJ (2005). Identification of core and variable components of the *Salmonella enterica* subspecies I genome by microarray. *Infect Immun*. **73**:7894-7905.

Argaman L and Altuvia S (2000). fhlA repression by OxyS RNA: kissing complex formation at two sites results in a stable antisense-target RNA complex. *J Mol Biol*. **300**:1101-1112.

Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H and Altuvia S (2001). Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol*. **26**:941-950.

Arthur DC, Ghetu AF, Gubbins MJ, Edwards RA, Frost LS and Glover JN (2003). FinO is an RNA chaperone that facilitates sense-antisense RNA interactions. *EMBO J*. **22**:6346-6355.

Axmann IM, Kensche P, Vogel J, Kohl S, Herzel H and Hess WR (2005). Identification of cyanobacterial non-coding RNAs by comparative genome analysis. *Genome Biol*. **6**:R73.

Bailey TL, Williams N, Misleh C and Li WW (2006). MEME: discovering and analysing DNA protein sequence notifs. *Nucleic Acids Res*. **34**:W369-W373.

Bajaj V, Hwang C and Lee CA (1995). *hilA* is a novel ompR/toxR family member that activates the expression of *Salmonella* Typhimurium invasion genes. *Mol Microbiol*. **18**:715-27.

Bakowski MA, Braun V and Brumell JH (2008). *Salmonella*-containing vacuoles: directing traffic and nesting to grow. *Traffic*. **9**:2022-31.

Bandyra KJ, Said N, Pfeiffer V, Górna MW, Vogel J and Luisi BF (2012). The seed region of a small RNA drives the controlled destruction of the target mRNA by the endoribonuclease RNase E. *Mol Cell*. **47**:943-953.

Baron S, Peake RC, James DA, Susman M, Kennedy CA, Singleton MJD and Schuenke S (1996). Medical Microbiology, 4th edition. The University of Texas Medical Branch at Galveston.

Bäumler AJ and Heffron F (1995). Identification and sequence analysis of lpfABCDE, a putative fimbrial operon of *Salmonella* Typhimurium. *J Bacteriol*. **177**:2087-2097.

Bäumler AJ, Tsolis RM, Bowe FA, Kusters JG, Hoffmann S and Heffron F (1996). The *pef* fimbrial operon of *Salmonella* Typhimurium mediates adhesion to murine small intestine and is necessary for fluid accumulation in the infant mouse. *Infect Immun*. **64**:61–68.

Beuzón CR and Holden DW (2001). Use of mixed infections with Salmonella strains to study virulence genes and their interactions in vivo. *Microbes Infect.* **3**:1345-1352.

Bispham J, Tripathi BN, Watson PR and Wallis TS (2001). Salmonella pathogenicity island 2 influences both systemic salmonellosis and Salmonella-induced enteritis in calves. *Infect Immun*. **69**:367-377.

Bogomolnaya LM, Santiviago CA, Yang HJ, Baumler AJ and Andrews-Polymenis HL (2008). 'Form variation' of the O12 antigen is critical for persistence of *Salmonella* Typhimurium in the murine intestine. *Mol Microbiol*. **70**:1105-1119.

Borukhov S, Sagitov V and Goldfarb A (1993). Transcript cleavage factors from *E. coli*. *Cell*. **72**:459-466.

Bouet JY, Rech J, Egloff S, Biek DP and Lane D (2005). Probing plasmid partition with centromere-based incompatibility. *Mol Microbiol.* **55**:511–525.

Boyd EF and Hartl DL (1998). Chromosomal regions specific to pathogenic isolates of Escherichia coli have a phylogenetically clustered distribution. *J Bacteriol*. **180**: 1159-1165.

Boyd EF, Wang FS, Whittam TS and Selander RK (1996). Molecular genetics relationships of the Salmonellae. *Appl Environ Microbiol*. **62**:804-808.

Bradford JR, Hey Y, Yates T, Li Y, Pepper SD and Miller CJ (2010). A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics*. **11**:282.

Bradley DE (1984). Characteristics and function of thick and thin conjugation pili determined by transfer-derepressed plasmids of incompatibility I1, I2, I5, B, K and Z. *J Gen Microbiol*. **130**:1489-502.

Brantl S (2002). Antisense RNAs in plasmids: control of replication and maintenance. *Plasmid.* **48**:165-73.

Brantl S (2007). Regulatory mechanisms employed by *cis*-encoded antisense RNAs. *Curr Opin Microbiol.* **10**:102-109.

Bustamante VH, Martínez LC, Santana FJ, Knodler LA, Steele-Mortimer O and Puente JL (2008). HilD-mediated transcriptional cross-talk between SPI-1 and SPI-2. *Proc Natl Acad Sci* U S A. **105**:14591-14596.

Caimano MJ, Sivasankaran SK, Allard A, Hurley D, Hokamp K, Grassmann AA, Hinton JC and Nally JE (2014). A model system for studying the transcriptomic and physiological changes associated with mammalian host-adaptation by Leptospira interrogans serovar Copenhageni. *PLoS Pathog*. **10**:e1004004.

Caldwell AL and Gulig PA (1991). The Salmonella Typhimurium virulence plasmid encodes a positive regulator of a plasmid-encoded virulence gene. *J Bacteriol*. **173**:7176-7185.

Camacho EM, Serna A, Madrid C, Marqués S, Fernández R, de la Cruz F, Juárez A and Casadesús J (2005). Regulation of finP transcription by DNA adenine methylation in the virulence plasmid of *Salmonella enterica*. *J Bacteriol*. **187**:5691-5699.

Cascales E, Buchanan SK, Duché D, Kleanthous C, Lloubès R, Postle K, Riley M, Slatin S and Cavard D (2007). Colicin biology. *Microbiol Mol Biol Rev*. **71**:158-229.

Chao Y, Papenfort K, Reinhardt R, Sharma CM and Vogel J (2012). An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J*. **31**:4005-4019.

Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang J, van Diemen PM, Buckley AM, Bowen AJ, Pullinger GD, Turner DJ, Langridge GC, Turner AK, Parkhill J, Charles IG, Maskell DJ and Stevens MP (2013). Comprehensive assignment of roles for *Salmonella* Typhimurium genes in intestinal colonization of food-producing animals. *PLoS Genet*. **9**:e1003456.

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

Chiu CM and Thomas CM (2004). Evidence for past integration of IncP-1 plasmids into bacterial chromosome. *FEMS Microbiol Lett.* **241**: 163-169.

Cho BK, Zengler K, Qiu Y, Park YS, Knight EM, Barrett CL, Gao Y and Palsson BØ (2009). The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol*. **27**:1043-1049.

Cock PJ, Fields CJ, Goto N, Heuer ML and Rice PM (2010). The Sanger FASTQC file format sequences with quality scores and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. **38**:1767-1771.

Corcoran CP, Podkaminski D, Papenfort K, Urban JH, Hinton JC and Vogel J (2012). Superfolder GFP reporters validate diverse new mRNA targets of the classic porin regulator, MicF RNA*. Mol Microbiol*. **84**:428-445.

Cortes T, Schubert OT, Rose G, Arnvig KB, Comas I, Aebersold R and Young DB (2013) Genome-wide Mapping of Transcriptional Start Sites Defines an Extensive Leaderless Transcriptome in *Mycobacterium tuberculosis*. *Cell Rep*. **5**:1121-1131.

Couturier M, Bex F, Bergquist PL and Maas WK (1988). Identification and classification of bacterial plasmids. *Microbiol Rev* **52**: 375-395.

Critchlow SE, O'Dea MH, Howells AJ, Couturier M, Gellert M and Maxwell A (1997). The interaction of the F plasmid killer protein, CcdB, with DNA gyrase: induction of DNA cleavage and blocking of transcription. *J Mol Biol*. **273**:826-839.

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

Croucher NJ and Thomson NR (2010). Studying bacterial transcriptome using RNA-Seq. *Curr Opin Microbiol.* **13**: 619-24.

Crump JA, Luby SP and Mintz ED (2004). The global burden of typhoid fever. *Bull World Health Organ*. **82**:346-53.

Dam M and Gerdes K (1994). Partitioning of plasmid R1.Ten direct repeats flanking the parA promoter constitute a centromere-like partition site parC, that expresses incompatibility. *J Mol Biol.* **236**:1289–98.

Datsenko KA and Wanner BL (200). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 usinf PCR products. *Proc Natl Acad Sci* U S A. 97:6640-6645.

Datta N and Hedges RW (1971). Compatibility groups among FI-R factors. *Nature* **234**:222-223.

Davis MA and Austin SJ (1988). Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. *EMBOJ.* **7**:1881–1888.

Delihas N and Forst S (2001). MicF: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J Mol Biol*. **313**:1-12.

Dennis PP and Omer A (2005). Small non-coding RNAs in Archaea. *Curr Opin Microbiol.* **8**:685-694.

Desnoyers G, Bouchard MP and Massé E (2013). New insights into small RNA-dependent translation regulation in prokaryotes. *Trends Genet*. **29**:92-98.

Dillies MA, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N, Keime C, Marot G, Castel D, Estelle J, Guernec G, Jagla B, Jouneau L, Laloë D, Le Gall C, Schaëffer B, Le Crom S, Guedj M, Jaffrézic F and French StatOmique Consortium (2013). A comprehensive

evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*. **14**:671-683.

Dornenburg JE, Devita AM, Palumbo MJ and Wade JT (2010). Widespread antisense transcription in *Escherichia coli*. *MBio.* **1**:e00024-10.

Doyle M, Fookes M, Ivens A, Mangan MW, Wain J and Dorman CJ (2007). An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science*. **315**:251-2.

Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K and Sharma CM (2013) High-resolution transcriptome maps reveal strand-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet*. **9**:e1003495.

Dühring U, Axmann IM, Hess WR and Wilde A (2006). An internal antisense RNA regulates expression of the photosynthesis gene isiA. *Proc Natl Acad Sci* U S A. **103**:7054-8.

Ebersbach G and Gerdes K (2005). Plasmid segregation mechanisms. *Annu Rev Genet*. **39**:453-479.

Escolar L, Pérez-Martín J and de Lorenzo V (1999). Opening the iron box: transcriptional metalloregulation by the Fur protein. *J Bacteriol*. **181**:6223-6229.

European Food Safety Authority and European Centre for Disease Prevention and Control (2014). The European Union Summary Report on Trends and Scources of Zoonoses, Zoonotic Agents and Food-borne Outbreaks in 2012. *EFSA Journal*. *12*:3547.

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

Fang FC, Libby SJ, Buchmeier NA, Loewen PC, Switala J, Harwood J and Guiney DG (1992) The alternative sigma factor katF (rpoS) regulates *Salmonella* virulence. *Proc Natl Acad Sci* U S A **89**:11978-11982.

Fang G, Passalacqua KD, Hocking J, Llopis PM, Gerstein M, Bergman NH and Jacobs-Wagner C (2013). Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. *BMC Genomics*. **14**:450.

Fernández RO and Antón DN (1987). Bacteriostatic action of streptomycin on ribosomally resistant mutants (rpsL) of *Salmonella* Typhimurium. *Antimicrob Agents Chemother*. **31**:1627-1631.

Fineran PC, Blower TR, Foulds IJ, Humphreys DP, Lilley KS, and Salmond GP (2009). The phage abortive infection system, ToxIN, functions as a protein- RNA toxin-antitoxin pair. *Proc Natl Acad Sci* U S A. **106**:894-899.

Foley SL, Zhao S, Walker RD (2007). Comparison of molecular typing methods for the differentiation of Salmonella foodborne pathogens. *Foodborne Pathog Dis*. **4**:253-76.

Fonseca NA, Rung J, Brazma A and Marioni JC (2012). Tools for mapping high-throughput sequencing data. *Bioinformatics*. **28**:3169-3177.

Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR, Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Bäumler AJ, Nuccio SP, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi A, Roberts M, Frankel

G, Parkhill J, Dougan G and Thomson NR (2010). *Salmonella* bongori provides insights into the evolution of the Salmonellae. *PLoS Pathog*. **7**:e1002191.

Francia MV, Varsaki A, Garcillán-Barcia MP, Latorre A, Drainas C and de la Cruz F (2004). A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiol Rev.* **28**:79-100.

Frost L, Lee S, Yanchar N and Paranchych W (1989). FinP and fisO mutations in FinP anti-sense RNA suggest a model for FinOP action in the repression of bacterial conjugation by the Flac plasmid JCFL0. *Mol Gen Genet*. **218**:152-160.

Furuya N, Nisioka T and Komano T (1991). Nucleotide sequence and functions of the oriT operon in IncI1 plasmid R64. *J Bacteriol*. **173**:2231-2237.

Garcia-del Portillo F, Foster JW and Finlay BB (1993). Role of acid tolerance response genes in *Salmonella* Typhimurium virulence. *Infect Immun*. **61**:4489-92.

García-Quintanilla M and Casadesús J (2011). Virulence plasmid interchange between strains ATCC 14028, LT2, and SL1344 of *Salmonella enterica* serovar Typhimurium. *Plasmid*. **65**:169-175.

García-Quintanilla M, Prieto AI, Barnes L, Ramos-Morales F and Casadesús J (2006). Bile-induced curing of the virulence plasmid in *Salmonella enterica* serovar Typhimurium. *J Bacteriol*. **188**:7963-7965.

García-Quintanilla M, Ramos-Morales F and Casadesús J (2008). Conjugal transfer of the *Salmonella enterica* virulence plasmid in the mouse intestine. *J Bacteriol*. **190**:1922-1927.

Garmendia J, Beuzón CR, Ruiz-Albert J and Holden DW (2003). The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the *Salmonella* Typhimurium SPI-2 type III secretion system. *Microbiology*. **149**:2385-96.

Geissmann TA and Touati D (2004). Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J*. **23**:396-405.

Gemski P Jr and Stocker (1967) Transduction by Bacteriophage P22 in Non smooth Mutants of *Salmonella* Typhimurium. *J Bacteriol*. **93**:1588-1597.

Georg J and Hess WR (2011). C*is*-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev*. **75**:286-300.

Georg J, Voss B, Scholz I, Mitschke J, Wilde A and Hess WR (2009). Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Mol Syst Biol*. **5**:305.

Gerdes K, Christensen SK and Løbner-Olesen A (2005). Prokaryotic toxinantitoxin stress response loci. *Nat Rev Microbiol* **3**:371-382.

Gerdes K, Møller-Jensen J and Jensen RB (2000). Plasmid and chromosome partitioning: surprises from phylogeny. *Mol Microbiol.* **37**:455–66.

Gerdes K, Poulsen LK, Thisted T, Nielsen AK, Martinussen J and Andreasen PH (1990). The *hok* killing gene family in Gram-negative bacteria. *New. Biol.* **2**:946-56.

Glöckner S, Streckel W, Struy H, Fruth A, Morenz J and Tschape H (1997). Further search for virulence factors encoded by *Salmonella* serovar-specific plasmids. *Zentralbl Bakteriol* **286**:69–82.

Gong H, Vu GP, Bai Y, Chan E, Wu R, Yang E, Liu F and Lu S (2011). A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog*. **7**:e1002120.

Gonzalo-Asensio J, Ortega AD, Rico-Pérez G, Pucciarelli MG and García-Del Portillo F (2013). A novel antisense RNA from the *Salmonella* virulence plasmid pSLT expressed by non-growing bacteria inside eukaryotic cells. *PLoS One*. **8**:e77939.

Gracia-Quintanilla M and Casdesus J (2011). Virulence plasmid interchanges between strains ATCC 14028, LT2, and SL1344 of *Salmonella* enterica serovar Typhimurium. *Plasmid*. **65**:169-75.

Grimont PAD and Weill FX (2007). Antigenic formulae of the *Salmonella* serovars, Ninth Edition, world health organization collaborating centre for reference and research on *Salmonella*. Institute Pasteur, Paris.

Gruber TM and Gross CA (2003). Multiple sigma subunits and partitioning of bacterial transcription space. Annu Rev Microbiol. **57**:441-466.

Guan S, Bastin DA and Verma NK (1999). Functional analysis of the O antigen glucosylation gene cluster of Shigella flexneri bacteriophage SfX. *Microbiology*. **145**:1263-1273.

Guiney DG and Fierer J (2011). The role of the *spv* gene in *Salmonella* pathogenesis. *Front. Microbiol*. **2**:e129.

Guiney DG, Fang FC, Krause M and Libby S (1994) Plasmid-mediated virulence genes in non-typhoid *Salmonella* serovars. *FEMS Microbiol Lett*. **124**:1–9.

Guiney DG, Libby S, Fang FC, Krause M and Fierer J (1995). Growth phase regulation of plasmid virulence genes in *Salmonella*. *Trends Microbiol.* **3**:275-279.

Gulig PA (1990). Virulence plasmids of *Salmonella Typhimurium* and other *salmonellae*. *Microb.Pathog.* **8**:3-11.

Gulig PA and Curtiss R (1987). Plasmid-associated virulence of *Salmonella* Typhimurium. *Infect. Immun*. **55**:2891-901.

Gulig PA and Curtiss R (1987). Plasmid-associated virulence of *Salmonella* Typhimurium. *Infect.Immun.* **55**:2891-2901.

Gulig PA and Doyle TJ (1993). The *Salmonella* Typhimurium virulence plasmid increases the growth rate of salmonellae in mice. *Infect.Immun.* **61**:504-511.

Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F and Rhen M (1993). Molecular analysis of *spv* virulence genes of the *Salmonella* virulence plasmids. *Mol Microbiol*. **7**:825-830.

Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F and Rhen M (1993). Molecular analysis of spv virulence genes of the *Salmonella* virulence plasmids. *Mol Microbiol.* **7**:825-830.

Gulig PA, Doyle TJ, Hughes JA, and Matsui H (1998). Analysis of host cells associated with the Spv-mediated increased intracellular growth rate of *Salmonella* Typhimurium in mice. *Infect Immun.* **66**:2471-2485.

Haneda T, Ishii Y, Shimizu H, Ohshima K, Iida N, Danbara H and Okada N (2011). *Salmonella* type III effector SpvC, a phosphothreonine lyase, contributes to reduction in inflammatory response during intestinal phase of infection. *Cell Microbiol*. **14**:485-499.

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

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

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

Hayes F (2003). Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science*. **301**:1496-1499.

Hayes F and Barillà D (2006). The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat Rev Microbiol*. **4**:133-143.

Health Protection Surveillance Centre, Ireland (2012). Annual report of salmonellosis in Ireland.

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

Heithoff DM, Sinsheimer RL, Low DA and Mahan MJ (1999). An essential role for DNA adenine methylation in bacterial virulence. *Science*. **284**:967-970.

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

Hernández JA, Muro-Pastor AM, Flores E, Bes MT, Peleato ML and Fillat MF (2006). Identification of a furA *cis* antisense RNA in the cyanobacterium *Anabaena* sp. PCC7120. *J Mol Biol*. **355**:325-334.

Herring CD, Raffaelle M, Allen TE, Kanin EI, Landick R, Ansari AZ and Palsson BØ (2005). Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J Bacteriol.* **187**:6166-6174.

Hershberg R, Altuvia S and Margalit H (2003). A survey of small RNA-encoding genes in *Escherichia coli*. *Nucleic Acids Res*. **31**:1813-1820.

Hinton JC (1997). The *Escherichia coli* genome sequence: the end of an era or the start of the FUN? *Mol Microbiol.* **26**:417-422.

Ho DK, Tissari J, Järvinen HM, Blom AM, Meri S and Jarva H (2011). Functional recruitment of human complement inhibitor C4B-binding protein to outer membrane protein Rck of *Salmonella*. *PLoS One*. **6**:e27546.

Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF and Hackermüller J (2009). Fast mapping of short sequences with mis-matches, insertions and deletions using index structures. *PLoS Comput Biol.* **5**: e1000502.

Humphreys S, Stevenson A, Bacon A, Weinhardt AB and Roberts M (1999). The alternative sigma factor, sigmaE, is critically important for the virulence of *Salmonella* Typhimurium. *Infect Immun*. **67**:1560-1568.

Hurme R, Berndt KD, Normark SJ and Rhen M (1997). A proteinaceous gene regulatory thermometer in *Salmonella*. *Cell*. **90**:55-64.

Jackson RW, Vinatzer B, Arnold DL, Dorus S and Murillo J (2011). The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elements.* **1**:55-65.

Jarvik T, Smillie C, Groisman EA and Ochman H (2010). Short-term signatures of evolutionary change in the *Salmonella enterica* serovar Typhimurium 14028 genome. *J Bacteriol*. **192**:560-567.

Jensen, R. B., and K. Gerdes (1995). Programmed cell death in bacteria: proteic plasmid stabilization systems. *Mol Microbiol.* **17**:205-210.

Jerome LJ, van Biesen T and Frost LS (1999). Degradation of FinP antisense RNA from F-like plasmids: the RNA-binding protein, FinO, protects FinP from ribonuclease E. *J Mol Biol*. **285**:1457-1473.

Kauffmann F (1941). Die Bakteriologie der *Salmonella*-gruppe. Munksgaard. Copenhagen, Sweden.

Kauffmann F (1966). The bacteriology of *Enterobacteriaceae*. Munksgaard. Copenhagen, Sweden.

Kim JN and Kwon YM (2013). Identification of target transcripts regulated by small RNA RyhB homologs in *Salmonella*: RyhB-2 regulates motility phenotype. *Microbiol Res*. **168**:621-629.

Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE, Gordon MA, Harris D, Clarke L, Whitehead S, Sangal V, Marsh K, Achtman M, Molyneux ME, Cormican M, Parkhill J, MacLennan CA, Heyderman RS and Dougan G (2009). Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res*. **19**:2279-2287.

Klein RJ, Misulovin Z and Eddy SR (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc Natl Acad Sci* U S A. **99**:7542-7547.

Kolb FA, Westhof E, Ehresmann B, Ehresmann C, Wagner EG and Romby P (2001). Four-way juntions in antisense RNA-mRNA complexes involved in plasmid replication control: a common theme? *J Mol Biol*. **309**:605-614.

Komano T, Kim SR and Nisioka T (1987). Distribution of shufflon among IncI plasmids. *J. Bacteriol*. **169**:5317-5319.

Koonin EV (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J Mol Biol.* **229**:1165–1174.

Kowarz L, Coynault C, Robbe-Saule V and Norel F (1994). The *Salmonella* Typhimurium katF (rpoS) gene: cloning, nucleotide sequence, and regulation of spvR and spvABCD virulence plasmid genes. *J Bacteriol*. **176**:6852-6860.

Kowarz L, Robbe-Saule V and Norel F (1996). Identification of *cis*-acting DNA sequences involved in the transcription of the virulence regulatory gene *spvR* in *Salmonella* Typhimurium. *Mol Gen Genet*. **251**:225-235.

Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M and Turner DJ (2009). Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods.* **6**:291-295.

Krause M, Fang FC, el-Gedaily A, Libby S and Guiney DG (1995). Mutational analysis of SpvR binding to DNA in the regulation of the Salmonella plasmid virulence operon. *Plasmid*. **34**:37-47.

Kröger C, Colgan A, Srikumar S, Händler K, Sivasankaran SK, Hammarlöf DL, Canals R, Grissom JE, Conway T, Hokamp K and Hinton JC (2013). An infection-relevant transcriptomic compendium for *Salmonella* serovar Typhimurium. *Cell Host & Microb.* **14**:683-695.

Kröger C, Dillon SC, Cameron AD, Papenfort K, Sivasankaran SK, Hokamp K, Chao Y, Sittka A, Hébrard M, Händler K, Colgan A, Leekitcharoenphon P, Langridge GC, Lohan AJ, Loftus B, Lucchini S, Ussery DW, Dorman CJ, Thomson NR, Vogel J and Hinton JC (2012). The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci* U S A. **109**:E1277-E1286.

Kues U and Stahl U (1989). Replication of Plasmids in Gram-Negative Bacteria. *Microbiol Rev*. **53**:491-516.

Langmead B and Salzberg AL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**:357-359.

Langmead B, Trapnell C, Pop M and Salzberg SL (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J and Turner AK (2009). Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res*. **19**:2308-2316.

Lara-Tejero M and Galán JE (2009). *Salmonella enterica* serovar Typhimurium pathogenicity island 1-encoded type III secretion system translocases mediate intimate attachment to nonphagocytic cells. *Infect Immun*. **77**:2635-2642.

Lasa I, Toledo-Arana A, Dobin A, Villanueva M, de los Mozos IR, Vergara-Irigaray M, Segura V, Fagegaltier D, Penadés JR, Valle J, Solano C and Gingeras TR (2011). Genome-wide antisense transcription derives mRNA processing in bacteria. *Proc Natl Acad Sci* U S A. **108**:20172-20177.

Lease RA, Cusick ME and Belfort M (1998). Riboregulation in Escherichia coli: DsrA RNA acts by RNA:RNA interactions at multiple loci. *Proc Natl Acad Sci* U S A. **95**:12456-12461.

Lederberg J (1952). Cell genetics and hereditary symbiosis. *Physiol Rev*. 32:403-430.

Lee EJ and Groisman EA (2010). An antisense RNA that governs the expression kinetics of a multifunctional virulence gene. *Mol Microbiol*. **76**:1020-1033.

Lennox ES (1955). Transduction of linked genetic characters of the host by bacteriophage P1. *Virology*. **1**:190-206.

Li B, Ruotti V, Stewart RM, Thomson JA and Dewey CN (2010). RNA-Seq gene expression estimation with reads mapping uncertainty. *Bioinformatics*. **26**:493-500

Li H and Durbin R (2009). Fast and Accurate Read Alignment with Burrows-Wheeler Transform. *Bioinformatics.* **25**:1754–1760.

Li R, Li Y, Kristiansen K and Wang J (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics.* **24**:713–714.

Libby SJ, Adams LG, Ficht TA, Allen C, Whitford HA, Buchmeier NA, Bossie S and Guiney DG (1997). The *spv* genes on the *Salmonella* Dublin virulence plasmid are required for severe enteritis and systemic infection in the natural host. *Infect Immun.* **65**:1786-1792.

Lindberg J and Lundeberg J (2010). The plasticity of the mammalian transcriptome. *Genomics*. **95**:1-6.

Liu JM and Camilli A (2010). A broadening world of bacterial small RNAs. *Curr Opin Microbiol*. **13**:18-23.

Liu MY, Gui G, Wei B, Preston JF 3rd, Oakford L, Yüksel U, Giedroc DP and Romeo T (1997). The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in Escherichia coli. *J Biol Chem*. **272**:17502-17510.

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

MacLean D, Jones JD and Studholme DJ (2009). Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol.* **7**:287-96.

Madsen L, Aarestrup FM and Olsen JE (2000). Characterisation of streptomycin resistance determinants in Danish isolates of *Salmonella* Typhimurium. *Vet Microbiol*. **75**:73-82.

Maetschke SR, Towsey MW and Hogan JM (2006). Bacterial promoter modelling and prediction for *E. coli* and *B. Subtilis* with beagle. In Proc. Workshop on Intelligent Systems for Bioinformatics (WISB), Hobart, Australia. CRPIT, 73. Boden M and Bailey TL, Eds., ACS. 9-13.

Magnuson R (2007). Hypothetical functions of toxin-antitoxin systems. *J Bacteriol*. **189**:6089-6092.

Majdalani N, Cunning C, Sledjeski D, Elliott T and Gottesman S (1998). DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci* U S A. **95**:12462-12467.

Majowicz SE, Musto J, Scallan E, Angulo FJ, Kirk M, O'Brien SJ, Jones TF, Fazil A and Hoekstra RM (2010). International Collaboration on Enteric Disease 'Burden of Illness' Studies. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis*. **50**:882-889.

Maloy S, Stewart VJ and Taylor RK (1996). Genetic Analysis of Pathogenic Bacteria: A Laboratory Manual. Cold Spring Harbor Press.

Mamanova V, Andrews RM, James KD, Sheridan EM, Ellis PD, Langford CF, Ost TW, Collins JE and Turner DJ (2010). FRT-Seq. Amplification free, strand-specific transcriptome sequencing. *Nat Methods.* **7**:130-132.

Marioni JC, Mason CE, Mane SM, Stephens M and Gilad Y (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res*. **18**:1509-1517.

Mariscotti JF and García-del Portillo F (2009). Genome expression analyses revealing the modulation of the Salmonella Rcs regulon by the attenuator IgaA. *J Bacteriol*. **191**:1855-1867.

Marr MT and Roberts JW (2000). Function of transcription cleavage factors GreA and GreB at a regulatory pause site*. Mol Cell.* **6**:1275-1285.

Marshall DG, Sheehan BJ and Dorman CJ (1999). A role for the leucine-responsive regulatory protein and integration host factor in the regulation of the *Salmonella* plasmid virulence (*spv*) locus in Salmonella Typhimurium. *Mol Microbiol*. **34**:134-145.

Mavris M, Manning PA and Morona R (1997). Mechanism of bacteriophage SfII-mediated serotype conversion in *Shigella flexneri*. *Mol Microbiol*. **26**:939-950.

McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S, Sabo A, Meyer R, Bieri T, Ozersky P, McLellan M, Harkins CR, Wang C, Nguyen C, Berghoff A, Elliott G, Kohlberg S, Strong C, Du F, Carter J, Kremizki C, Layman D, Leonard S, Sun H, Fulton L, Nash W, Miner T, Minx P, Delehaunty K, Fronick C, Magrini V, Nhan M, Warren W, Florea L, Spieth J and Wilson RK (2004). Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nat Genet*. **36**:1268-1274.

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

McGettigan PA (2013). Transcriptomics in the RNA-Seq era. *Curr Opin Chem Biol*. **17**:4-11.

Mendoza-Vargas A, Olvera L, Olvera M, Grande R, Vega-Alvarado L, Taboada B, Jimenez-Jacinto V, Salgado H, Juárez K, Contreras-Moreira B, Huerta AM, Collado-Vides J, Morett E (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One*. **4**:e7526.

Minor LL and Popoff MY (1987). Designation of *Salmonella enterica* spp. nov., nom. Rev., as the type and only species of the genus *Salmonella*: request for an opinion. *Int J Syst Bacteriol*. **37**:465-468.

Mira A, Martín-Cuadrado AB, D'Auria G and Rodríguez-Valera F (2010). The bacterial pan-genome: a new paradigm in microbiology. *Int Microbiol.* **13**:45-57.

Mitschke J, Vioque A, Haas F, Hess WR and Muro-Pastor AM (2011). Dynamics of transcriptional start site selection during nitrogen stress-induced cell differentiation in *Anabaena* sp. PCC7120. *Proc Natl Acad Sci* U S A. **108**:20130-20135.

Møller-Jensen J, Borch J, Dam M, Jensen RB, Roepstorff P and Gerdes K (2003). Bacterial Mitosis: ParM of plasmid R1 moves DNA by an actin-like insertional polymerization mechanism. *Mo. Cell* **12**:1477–1487.

Moore SD and Sauer RT (2007). The tmRNA system for translational surveillance and ribosome rescue. *Annu Rev Biochem*. **76**:101-124.

Moreno Switt AI, den Bakker HC, Cummings CA, Rodriguez-Rivera LD, Govoni G, Raneiri ML, Degoricija L, Brown S, Hoelzer K, Peters JE, Bolchacova E, Furtado MR and Wiedmann M (2012). Identification of characterization of novel Salmonella mobile elements involved in the dissemination of genes linked to virulence and transmission. *PLoS One*. **7**:e41247.

Mortazavi A, Williams BA, McCue K, Schaeffer L and Wold B (2008). Mapping and quantifying mammalian transcriptome by RNA-Seq. *Nat Methods.* **5**: 621-628.

Murray PR, Baron EJ, Jorgensen JH, Landry ML, and Pfaller MA. (Eds.) (2007). *Manual of Clinical Microbiology.* 9th edn. Washington: ASM Press.

Mutz KO, Heilkenbrinker A, Lönne M, Walter JG and Stahl F (2013). Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol*. **24**:22-30.

Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M and Snyder M (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*. **320**:1344-1349.

Nicholson B and Low D (2000). DNA methylation-dependent regulation of *Pef* expression in *Salmonella* Typhimurium. *Mol Microbiol*. **35**:728–742.

Nicol JW, Helt GA, Blanchard SG Jr, Raja A and Loraine AE (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics.* **25**: 2730-2731.

Nielsen AK and Gerdes K (1995). Mechanism of post-segregation killing by *hok*-homologue pnd of plasmid R483: two translational contral elements in the *phd* mRNA. *J Mol Biol.* **249**:270-282.

Norman A, Hansen LH and Sørensen SJ (2009). Conjugative plasmids: vessels of the communal gene pool. *Philos Trans R Soc Lond B Biol Sci*. **364**:2275-2289.

Novick RP, Iordanescu S, Projan SJ, Kornblum J and Edelman I (1989). p181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell.* **59**:395-404.

O'Neil D, Glowatz H and Schlumpberger M (2013). Ribosomal RNA depletion for efficient use of RNA-Seq capacity. *Curr Protoc Mol Biol*. **103**:4.19.1-4.19.8.

O'Byrne CP and Dorman CJ (1994). The spv virulence operon of *Salmonella* Typhimurium LT2 is regulated negatively by the cyclic AMP (cAMP)-cAMP receptor protein system. *J Bacteriol*. **176**:905-912.

O'Byrne CP and Dorman CJ (1994). Transcription of the *Salmonella* Typhimurium spv virulence locus is regulated negatively by the nucleoid-associated protein H-NS. *FEMS Microbiol Lett*. **121**:99-105.

Ono S, Goldberg MD, Olsson T, Esposito D, Hinton JC and Ladbury JE (2005). H-NS is a part of a thermally controlled mechanism for bacterial gene regulation. *Biochem J*. **391**:203-213.

Opdyke JA, Fozo EM, Hemm MR and Storz G (2010). RNase III participates in GadY-dependent cleavage of the *gadX-gadW* mRNA. *J Mol Biol*. **406**:29-43.

Osmundson J, Dewell S and Darst SA (2013). RNA-Seq reveals differential gene expression in Staphylococcus aureus with single-nucleotide resolution. *PLoS One*. **8**:e76572.

Ozsolak F and Milos PM (2011). RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet.* **12**:87-98.

Padalon-Brauch G, Hershberg R, Elgrably-Weiss M, Baruch K, Rosenshine I, Margalit H and Altuvia S (2008). Small RNAs encoded within genetic islands of *Salmonella* Typhimurium show host-induced expression and role in virulence. *Nucleic Acids Res.* **36**:1913–1927.

Paesold G, Guiney DG, Eckmann L and Kagnoff MF (2002). Genes in the *Salmonella* pathogenicity island 2 and the Salmonella virulence plasmid are essential for *Salmonella*-induced apoptosis in intestinal epithelial cells. *Cell Microbiol*. **4**:771-781.

Paget MS and Helmann JD (2003). The sigma70 family of sigma factors. *Genome Biol*. **4**:203. Epub.

Pandey DP and Gerdes K (2005). Toxin-antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic Acids Res*. **33**:966-976.

Pansegrau W, Lanka E, Barth PT, Figurski DH, Guiney DG, Haas D, Helinski DR, Schwab H, Stanisich VA and Thomas CM (1994). Complete nucleotide sequence of Birmingham IncPα plasmids. Compliation and comparative analysis. *J Mol Biol*. **239**:623-663.

Pansegrau W, Schröder W and Lanka E (1993). Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc Natl Acad Sci* U S A. **90**:2925-2929.

Pansegrau W, Ziegelin G and Lanka E (1990). Covalent association of the traI gene product of plasmid RP4 with the 5'-terminal nucleotide at the relaxation nick site. *J Biol Chem*. **265**:10637-10644.

Papenfort K, Pfeiffer V, Mika F, Lucchini S, Hinton JC and Vogel J (2006). SigmaE-dependent small RNAs of *Salmonella* respond to membrane stress by accelerating global omp mRNA decay. *Mol Microbiol*. **62**:1674-1688.

Papenfort K, Podkaminski D, Hinton JC and Vogel J (2012). The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci* U S A. **109**:E757-E767.

Park SY, Moon MW, Subhadra B and Lee JK (2010). Functional characterization of the glxR deletion mutant of *Corynebacterium glutamicum* ATCC 13032: involvement of GlxR in acetate metabolism and carbon catabolite repression. *FEMS Microbiol Lett.* **304**:107-115.

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

Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME and Bergman NH (2009). Structure and complexity of bacterial transcriptome. *J. Bacteriol.* **191**: 3203-3211.

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD, Yu L, Assefa SA, He M, Croucher NJ, Pickard DJ, Maskell DJ, Parkhill J, Choudhary J, Thomson NR, Dougan G (2009) A strand-specific RNA-Seq analysis of the transcriptome of the typhoid bacillus Salmonella typhi. *PLoS Genet*. **5**:e1000569.

Perocchi F, Xu Z, Clauder-Münster S, Steinmetz LM (2007) Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic acid Res*. **35**:e128. Epub.

Pizarro-Cerdá J, Desjardins M, Moreno E, Akira S and Gorvel JP (1999). Modulation of endocytosis in nuclear factor IL-6(-/-) macrophages is responsible for a high susceptibility to intracellular bacterial infection. *J Immunol*. **162**:3519-3526.

Porwollik S, Boyd EF, Choy C, Cheng P, Florea L, Proctor E, and McClelland M (2004). Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. *Journal of Bacteriology.* **186**: 5883-5898.

Postma PW (1977). Galactose transport in Salmonella Typhimurium. *J Bacteriol*. **129**:630-639.

Puniya BL, Kulshreshtha D, Verma SP, Kumar S and Ramachandran S (2014). Integrated gene co-expression network analysis in the growth phase of Mycobacterium tuberculosis reveals new potential drug targets. *Mol Biosyst*. **9**:2798-2815.

Quackenbush J (2003). Genomics. Microarrays--guilt by association. *Science*. **302**:240-1.

Raghavan R, Sloan DB and Ochman H (2012). Antisense transcription is pervasive but rarely conserved in enteric bacteria. *MBio*. **3**:e00156-12.

Rasmussen S, Nielsen HB and Jarmer H (2009). The transcriptionally active regions in the genome of *Bacillus subtilis*. *Mol Microbiol*. **73**:1043-1057.

Reddy PS and Chaterji D (1994). Evidence for a pyrimidine-nucleotide-specific initiation site (the I site) on Escherichia coli TNA polymerase. Proximity relationship with the inhibitor binding domain. *Eur J Biochem.* **225**:737-745.

Reeves MW, Evins GM, Heiba AA, Plikaytis BD and Farmer JJ (1989). Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. Nov. *J Clin Microbiol*. **27**:313-320.

Rhen M and Sukupolvi S (1988). The role of the *traT* gene of the *Salmonella Typhimurium* virulence plasmid for serum resistance and growth within liver macrophages. *Microb.Pathog.* **5**: 275-285.

Richardson EJ, Limaye B, Inamdar H, Datta A, Manjari KS, Pullinger GD, Thomson NR, Joshi RR, Watson M and Stevens MP (2011). Genome sequences of *Salmonella enterica* serovar Typhimurium, Cholerasuis, Dublin, and Gallinarum strains of well-defined virulence in food-producing animals. *J. Bacteriol*. **193**:3162-3163.

Robbe-Saule V, Schaeffer F, Kowarz L and Norel F (1997). Relationships between H-NS, sigma S, SpvR and growth phase in the control of *spvR*, the regulatory gene of the *Salmonella* plasmid virulence operon. *Mol Gen Genet*. **256**:333-347.

Roh SW, Abell GC, Kim KH, Nam YD and Bae JW (2010). Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Treands Biotechnol*. **28**:291-299.

Rolfe MD, Rice CJ, Lucchini S, Pin C, Thompson A, Cameron AD, Alston M, Stringer MF, Betts RP, Baranyi J, Peck MW and Hinton JC (2012). Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J Bacteriol*. **194**:686-701.

Rotger R and Casadesus J (1999). The virulence plasmid of *Salmonella*. *Int. Microbiol*. **2**:177-184.

Roudier C, Fierer J and Guiney DG (1992). Characterisation of translation termination mutations in the spv operon of the *Salmonella* virulence plasmid pSDL2. *J Bacteriol*. **174**:6418-6423.

Ruffalo M, Koyutürk M, Ray S and LaFramboise T (2012). Accurate estimation of short reads mapping quality for next-generation genome sequencing. *Bioinformatics*. **28**:i349-i355.

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

Sakai H and Komano T (1996). DNA replication of IncQ broad-host-range plasmids in Gram-negative bacteria. *Biosci Biotechnol Biochem*. **60**:377-82.

Salmon DE and Smith T (1886). The bacterium of swine plague. *American Monthly Microbiology Journal*. **7**: 204.

Sambrook J and Russell D (2000). Molecular cloning. Cold Spring Harbor Laboratory Press. New York. U S A.

Santiviago CA, Reynolds MM, Porwollik S, Choi SH, Long F, Andrews-Polymenis HL and McClelland M (2009). Analysis of pools of targeted *Salmonella* deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog*. **5**:e1000477.

Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL and Griffin PM (2011). Foodborne illness acquired in the United States- -major pathogens. *Emerg Infect Dis*. **17**:7-15.

Selimovic BM, Babic T, Kocic B, Stojanovic P, Ristic L and Dinic M (2007) Bacterial Plasmids. *Acta Medica Medianae.* **46**: 61-65.

Selinger DW, Cheung KJ, Mei R, Johansson EM, Richmond CS, Blattner FR, Lockhart DJ and Church GM (2000). RNA expression analysis using a 30 base pair resolution *Escherichia coli*. *Nat Biotechnol.* **18**:1262-1268.

Sen D, Van der Auwera GA, Rogers LM, Thomas CM, Brown CJ and Top EM (2011). Broad-host-range plasmids from agricultural soils have IncP-1 backbones with diverse accessory genes. *Appl Environ Microbiol*. **77**: 7975-7983.

Seo JH, Hong JS, Kim D, Cho BK, Huang TW, Tsai SF, Palsson BO and Charusanti P (2012). Multiple-omic data analysis of *Klebsiella pneumoniae* MGH 78578 reveals its transcriptional architecture and regulatory features. *BMC Genomics*. **13**:679.

Serganov A and Nudler E (2013). A decade of riboswitches. *Cell*. **152**:17-24.

Sesto N, Wurtzel O, Archambaud C, Sorek R and Cossart P (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat Rev Microbiol*. **11**:75-82.

Sharma CM, Darfeuille F, Plantinga TH 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 Dev*. **21**:2804-2817.

Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermüller J, Reinhardt R, Stadler PF and Vogel J (2010). The primary transcriptome of the human pathogen *Helicobacter pylori*. *Nature*. **11**: 250-255.

Sharma CM, Papenfort K, Pernitzsch SR, Mollenkopf HJ, Hinton JC and Vogel J (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol*. **81**:1144-1165.

Shea JE, Beuzon CR, Gleeson C, Mundy R and Holden DW (1999). Influence of the *Salmonella* Typhimurium pathogenicity island 2 type III secretion system on bacterial growth in the mouse. *Infect Immun*. **67**:213-219.

Shelobolina ES, Sullivan SA, O'Neill KR, Nevin KP and Lovley DR (2004). Isolation, characterization, and U(VI)-reducing potential of a facultatively anaerobic, acid-resistant Bacterium from Low-pH, nitrate- and U(VI)-contaminated subsurface sediment and description of *Salmonella* subterranea sp. *Appl Environ Microbiol*. **70**:2959-2965.

Shultzaberger RK, Chen Z, Lewis KA and Schneider TD (2007). Anatomy of *Escherichia coli* singma70 promoters. *Nucleic Acids Res*. **35**:771-788.

Silvaggi JM, Perkins JB and Losick R (2006). Genes for small, noncoding RNAs under sporulation control in *Bacillus subtilis*. *J Bactriol*. **188**:532-541.

Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC and Vogel J (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet*. **4**:e1000163.

Skinner ME, Uzilov AV, Stein LD, Mungall CJ and Holmes IH (2009). JBrowse: a next-generation genome browser. *Genome Res.* **19**: 1630-1638.

Sleigh JD and Duguid JP (1989). *Salmonella*. In: Colle JG, Duguid JP, Fraser AG, Marmion BP (Eds.) Practical medical microbiology, 13th edn. Churchill Livingstone, Edinburgh, 456.

Smalla K, Osborn M and Wellington EMH (2000). Isolation and characterisation of plasmid from bacteria, p. 207-248. In Thomas CM (Eds.), The horizontal gene pool. Bacterial plasmids and gene spread. harwood Academic Publishers, Amsterdam.

Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP and de la Cruz F (2010). Mobility of plasmids. *Microbiol Mol Biol Rev*. **74**:434-452.

Sorek R and Cossart P (2010). Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet.* **11**: 9-16.

Stazic D, Lindell D and Steglich C (2011). Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. *Nucleic acid Re*s. **39**:4890-4899.

Stecher B, Denzler R, Maier L, Bernet F, Sanders MJ, Pickard DJ, Barthel M, Westendorf AM, Krogfelt KA, Walker AW, Ackermann M, Dobrindt U, Thomson NR and Hardt WD (2012). Gut inflammation can boost horizontal gene transfer between pathogenic and commensal *Enterobacteraceae*. *Proc Natl Acad Sci* U S A. **109**:1269-1274.

Sterzenbach T, Nguyen KT, Nuccio SP, Winter MG, Vakulskas CA, Clegg S, Romeo T and Bäumler AJ (2013). A novel CsrA titration mechanism regulates fimbrial gene expression in *Salmonella* Typhimurium. *EMBO J*. **32**:2872-2883.

Stolt P and Zillig W (1993). Antisense RNA mediated transcriptional processing in an archaebacterium, including a novel kind of RNase activity. *Mol Microbiol*. **7**:875-882.

Storz G, Opdyke JA and Zhang A (2004). Controlling mRNA stability and translation with small, noncoding RNAs. *Curr Opin Microbiol*. **7**:140-144.

Storz G, Vogel J and Wassarman KM (2011). Regulation by small RNAs in bacteria: expanding frontiers. *Mol Cell*. **43**:880-891.

Struhl K (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat Struct Mol Biol*. **14**:103-105.

Stuart JM, Segal E, Koller D and Kim SK (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*. **302**:249-255.

Su LH and Chiu CH (2007). *Salmonella*: clinical importance and evolution of nomenclature. *Chang Gung Medical Journal.* **30**:210-219.

Tarazona S, García-Alcalde F, Dopazo J, Ferrer A and Conesa A (2011). Differential expression in RNA-Seq: a matter of depth. *Genome Res*. **21**:2213-2223.

Texter FL, Radford SE, Laue ED, Perham RN, Miles JS and Guest JR (1988). Site-directed mutagenesis and 1H NMR spectroscopy of an interdomain segment in the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *Biochemistry*. **27**:289-296.

Texter FL, Radford SE, Laue ED, Perham RN, Miles JS and Guest JR (1988). Site-directed mutagenesis and 1H NMR spectroscopy of an interdomain segment in the pyruvate dehydrogenase multienzyme complex of Escherichia coli. *Biochemistry*. **27**:289-296.

Thomas CM (1980) Molecular genetics of broad host range plasmid RK2. *Plasmid.* **5**:10-19.

Thomas CM (2000) Paradigms of plasmid organization. Mlo Microbiol. **37**:485-491.

Thomason MK and Storz G (2010). Bacterial antisense RNAs: how many are there, and what are they doing? *Annu Rev Genet.* **44**:167-188.

Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S, Churcher C, Quail MA, Stevens M, Jones MA, Watson M, Barron A, Layton A, Pickard D, Kingsley RA, Bignell A, Clark L, Harris B, Ormond D, Abdellah Z, Brooks K, Cherevach I, Chillingworth T, Woodward J, Norberczak H, Lord A, Arrowsmith C, Jagels K, Moule S, Mungall K, Sanders M, Whitehead S, Chabalgoity JA, Maskell D, Humphrey T, Roberts M, Barrow PA, Dougan G and Parkhill J (2008). Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res*. **18**:1624-37.

Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Régnault B, Coppée JY, Lecuit M, Johansson J and Cossart P (2009). The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature*. **459**:950-956.

Tomizawa J, Itoh T, Selzer G and Som T (1981). Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci* U S A. **78**:1421-5.

Torreblanca J, Marqués S and Casadesús J (1999). Synthesis of FinP RNA by plasmids F and pSLT is regulated by DNA adenine methylation. *Genetics*. **152**:31-45.

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

Urbanowski ML, Stauffer LT and Stauffer GV (2000). The gcvB gene encodes a small untranslated RNA involved in expression of the dipeptide and oligopeptide transport systems in Escherichia coli. *Mol Microbiol*. **37**:856-868.

Valentin-Hansen P, Eriksen M and Udesen C (2004). The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol*. **51**:1525-1533.

Van Melderen L, and Saavedra De Bast M (2009). Bacterial toxin-antitoxin systems: more than selfish entities? *PLoS Genet*. **5**:e1000437.

Van Melderen L, Bernard P, and Couturier M (1994). Lon-dependent proteolysis of CcdA is the key control for activation of CcdB in plasmid-free segregant bacteria. *Mol Microbiol*. **11**:1151-1157.

Vandenbosch JL, Kurlandsky DR, Urdangaray R and Jones GW. (1989) Evidence of coordinate regulation of virulence in *Salmonella* Typhimurium involving the *rsk* element of the 95-kilobase plasmid. *Infect Immun.* **57**:2566-2568.

Vernikos GS and Parkhill J (2004). Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*. **22**:2196-2203.

Viegas CA, Lilley AK, Bruce K and Bailey MJ (1997). Description of a novel plasmid replicative origin from a genetically distinct family of conjugative plasmids associated with phytosphere microflora. *FEMS Microbiology Letters.* **149**:121-127.

Vogel J (2009). Rough guide to non-coding RNA world of *Salmonella*. *Mol. Microbiol.* **71**:1-11.

Vogel J and Luisi BF (2011). Hfq and its constellation of RNA. *Nat Rev Microbiol*. **9**:578-589.

Vogel J and Papenfort K (2006). Small non-coding RNAs and the bacterial outer membrane. *Curr Opin Microbiol*. **9**:605-611.

Wagner GP, Kin K and Lynch VJ (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci*. **131**:281-285.

Wagner GP, Kin K and Lynch VJ (2013). A model based criterion for gene expression calls using RNA-seq data. *Theory Biosci*. **132**:159-164.

Wang Z, Gerstein M and Snyder M (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* **10**:57-63.

Wassarman KM, Zhang A, Storz G (1999) Small RNAs in Escherichia coli. Trends *Microbiol.* **7**:37-45.

West NP, Sansonetti P, Mounier J, Exley RM, Parsot C, Guadagnini S, Prévost MC, Prochnicka-Chalufour A, Delepierre M, Tanguy M and Tang CM (2005). Optimization of virulence functions through glucosylation of *Shigella* LPS. *Science*. **307**:1313-1317.

Whitchurch CB and Mattick JS (1994). *Escherichia coli* contains a set of genes homologous to those involved in protein secretion, DNA uptake and the assembly of type-4 fimbriae in other bacteria. *Gene*. **150**:9-15.

Wilkins BM, Boulnois GJ and Lanka E (1981). A plasmid DNA promise active in discontinuous bacterial DNA replication. *Nature*. **290**:217-21.

Wilson JA and Gulig PA (1998). Regulation of the spvR gene of the Salmonella Typhimurium virulence plasmid during exponential-phase growth in intracellular salts medium and at stationary phase in L broth. *Microbiology*. **144**:1823-1833.

Wolf JB (2013). Principles of transcriptome analysis and gene expression quantification: an RNA-Seq tutorial. *Mol Ecol Resour*. **13**:559-572.

Wozniak CE, Lee C and Hughes KT (2009). T-POP array identifies EcnR and PefI-SrgD as novel regulators of flagellar gene expression. *J Bacteriol*. **191**:1498–1508.

Wray C and Sojka WJ (1978). Experimental *Salmonella* Typhimurium infection in calves. *Res Vet Sci*. **25**:139-143.

Wurtzel O, Sesto N, Mellin JR, Karunker I, Edelheit S, Bécavin C, Archambaud C, Cossart P and Sorek R (2012). Comparative transcriptomic of pathogenic and non-pathogenic *Listeria* species. *Mol Syst Biol*. **22**:583.

Ygberg SE, Clements MO, Rytkönen A, Thompson A, Holden DW, Hinton JC and Rhen M (2006). Polynucleotide phosphorylase negatively controls spv virulence gene expression in Salmonella enterica. *Infect Immun*. **74**:1243-1254.

Yoo Sm, Na D and Lee SY (2013). Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. *Nat Protoc*. **8**:1694-1707.