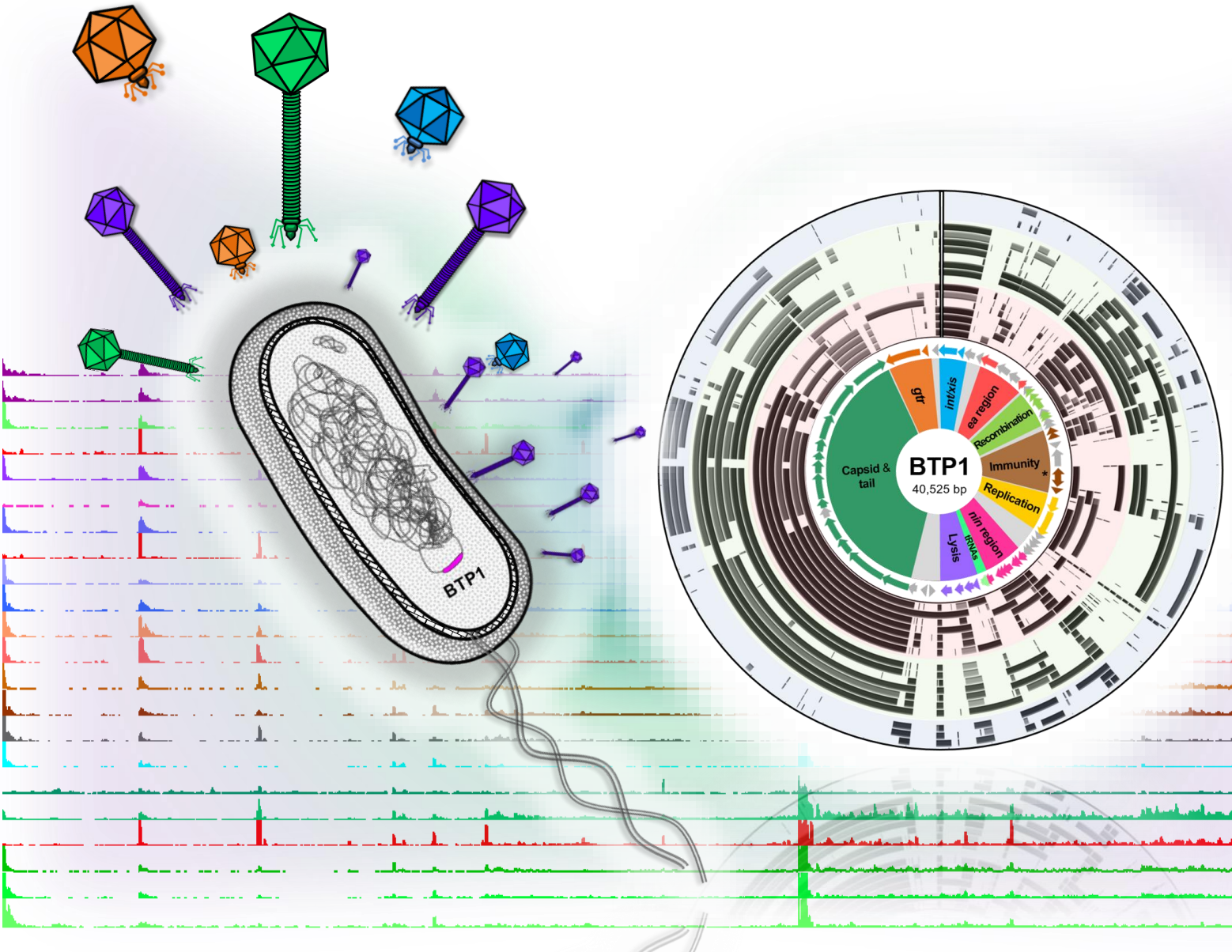
UNIVERSITY OF LIVERPOOL

# Exploring the prophage biology of *Salmonella enterica* serovar Typhimurium ST313

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

**Siân V. Owen**

September 2017



# Abstract

Title of thesis: **Exploring the prophage biology of *Salmonella enterica* serovar Typhimurium ST313**

Author: **Siân V. Owen**

In the past 30 years, *Salmonella* bloodstream infections have become a significant health problem in sub-Saharan Africa and are responsible for the deaths of ~390,000 people each year. The disease is largely caused by a recently described sequence type of *Salmonella* Typhimurium: ST313. Comparative genomic analysis showed that the ST313 lineage is closely-related to the ancestral gastroenteritis-associated Typhimurium sequence type ST19, but carries a distinct prophage repertoire. I hypothesised that prophages contribute to the biology of this clinically-relevant ST.

In this thesis I show that the African ST313 representative strain D23580 contains 5 full length prophages. Prophage BTP1 and BTP5 are undescribed, novel prophages specific to African ST313 strains, whilst Gifsy-2, ST64B and Gifsy-1 are well-characterised prophages found in other strains of *S.* Typhimurium. Of the five prophages, only BTP1 and BTP5 showed evidence for functional phage production, and mutations responsible for the inactivation of Gifsy-2, ST64B and Gifsy-1 were identified. The BTP1 prophage spontaneously induced at a prolific rate, estimated to result in the phage-mediated lysis of approximately 0.2% of the lysogenic cell population. A GFP reporter system was developed to visualise the spontaneous induction of the BTP1 prophage at the single-cell level. Though the BTP5 phage could not be studied using traditional plaque assay methodology, there was evidence that the BTP5 prophage was capable of forming viable BTP5 phage that could lysogenise naïve hosts.

I analysed the genomes of recently discovered ST313 isolates from the UK and show that ST313 in the UK represents a distinct population of antibiotic susceptible strains associated with gastrointestinal infection. Additionally, analysis of the UK-ST313 genomes indicated that the BTP1 and BTP5 prophages were acquired independently by the two African ST313 lineages, showing convergent evolution to acquire and conserve the BTP1 and BTP5 prophages.

Finally I present evidence that the prophages, in particular BTP5, effect the global gene expression of ST313, and the *ST313-td* gene of BTP1 mediates lysogenic conversion by functioning as a superinfection immunity factor against infection by *Salmonella* phage P22.

The implications of these findings for understanding the pathogen in terms of ecological niche, host range and invasiveness in humans is discussed.

# Acknowledgements

Firstly, I would like to express my gratitude to Jay Hinton, my primary supervisor, who first launched me on my PhD journey by welcoming me into his lab. Despite being the sole PhD student amongst an assortment of highly accomplished postdocs for the majority of my time in the Liverpool lab, from day one Jay has treated my ideas with as much credibility, and given me as much responsibility, as anyone in the lab. Jay has tolerated, and even encouraged, my less-scientific PhD exploits (I suspect not every supervisor would be so enthused by how much time I have dedicated to creative endeavours like graphic design and paper bacteriophages). The collaborative and supportive environment Jay has fostered within the group has meant that I finish my PhD a more empowered, confident person and scientist, than I was four years ago. I would additionally like to thank my secondary supervisor, Nick Feasey for his support, encouragement, and invaluable clinical insight, and Melita Gordon, who acted as my secondary supervisor during the first year of my PhD.

I am extremely fortunate to have completed my PhD surrounded and mentored by so many talented and generous postdocs and colleagues in the Hinton lab (Figure i) namely Disa Hammarlöf, Rocío Canals, Carsten Kröger, Shabi Srikumar, Aoife Colgan, Kristian Händler, Sathesh Sivasankaran, Lauren Gordon, Ben Kumwenda, Lizeth Lacharme Lora, Will Rowe, Blanca Perez Sepulveda, Wai Yee Fong and Caisey Pulford. I am particularly indebted to Rocío, for her endless encouragement, advice, and friendship, and for listening so patiently to all my weird ideas over the years; to Disa, for so much enthusiasm and reassurance in those important first few months; to Carsten for all the support and his enduring good humour; and to Nico, for schooling me in old-school genetics, for the laughs and the inspiring discussions. I am grateful to Blanca for so much encouragement and sympathy during the writing of this thesis, and I reserve a special kind of ‘gratitude’ for Lizeth, for not being afraid to push the boundaries of what constitutes a prank, and for all the fun.

Outside of the Hinton lab, I have been fortunate to be surrounded by wonderful colleagues in and around Lab H and the IIB over the years, in particular Paul Loughnane, Marta Veses Garcia, Bruno Manso, Jamie Casswell, Ewan Young, Mohammed Mohaisen, Stuart McEwan, Jo Moran, Evelien Adriaenssens, Charlotte Chong, Breno Salgado, Anushul Gupta, Sean Goodman, Jamie Alison and Carla Solórzano among others. I thank the other academic staff in Microbiology including Alan McCarthy, Mal Horsburgh, Heather Allison, Kate Baker and Al Darby for creating a stimulating and friendly academic environment. In particular, I thank Heather for persistently encouraging my attendance at phage conferences, for adopting me into her group at said phage conferences, and for lending me her copy of “A Genetic Switch” during my first week as a PhD student.

I acknowledge and thank Abram Aertsen and Angela Makumi from KU Leuven for a fruitful collaboration on the BTP1 prophage, and Phil Ashton, Tim Dallman and others at Public Health England for a fun collaboration on the UK-ST313 strains. I also wish to acknowledge the assistance of Karsten Hokamp and Paul Wigley at various stages of the project.

Finally, I take this opportunity to apologise to my friends and family outside of Liverpool for being a neglectful friend, daughter and sister these past four years; Dad, Mum, Tom, Rhys, eSteph, HM, Patsy, Dan, Fred and Rose. Thank you for your endless encouragement.

This thesis is dedicated to you all.

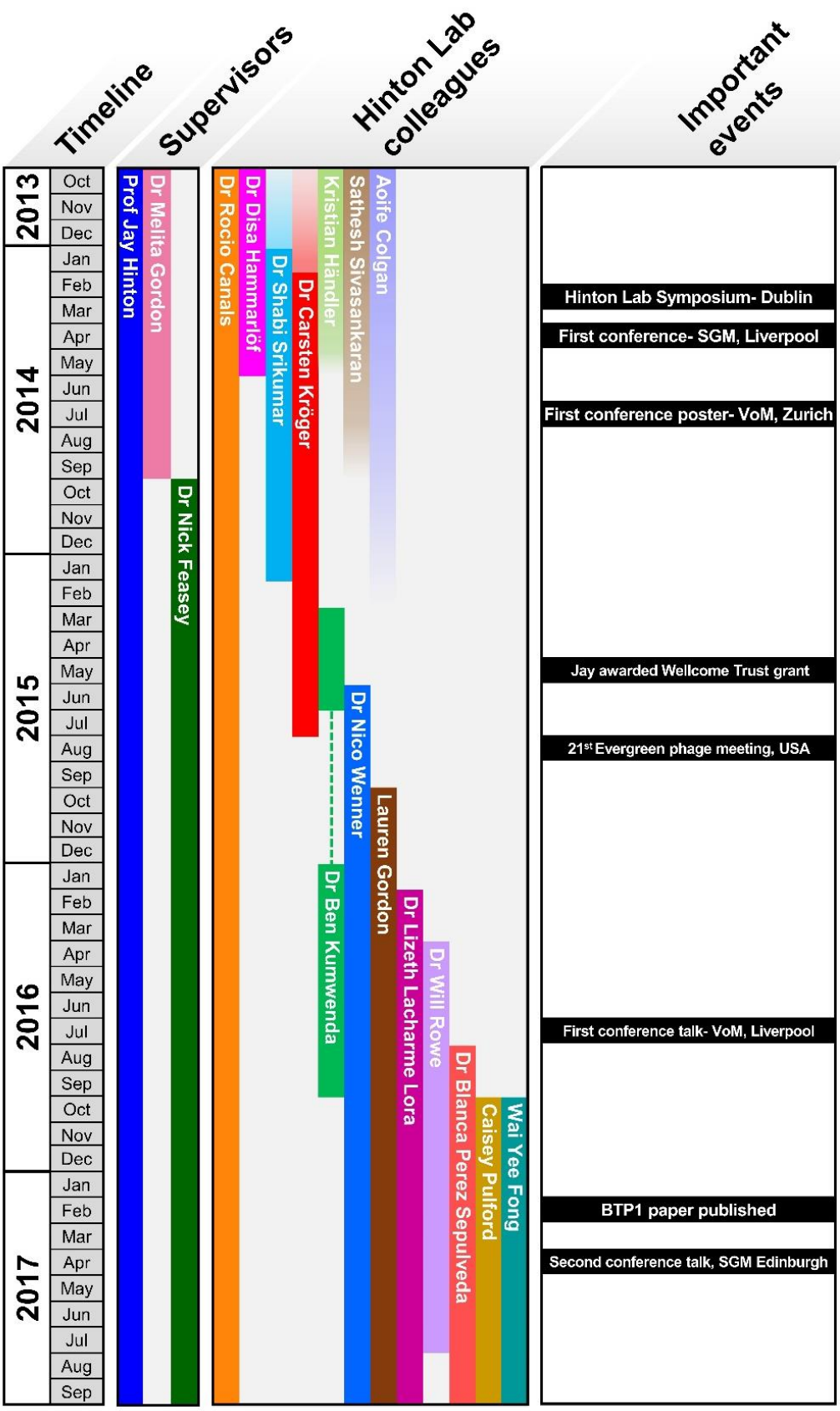



**Figure i Graphical representation of my PhD in the Hinton Lab**

# Publications and Conference Presentations

- **Publications in peer reviewed journals resulting from work described in this thesis:**

* Publications were I am first or joint-first author are indicated with an asterisk

Srikumar, S., Kröger, C., Hébrard, M., Colgan, A., **Owen, S.V**., Sivasankaran, S.K., Cameron, A.D., Hokamp, K., Hinton, J.C., 2015. RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of *Salmonella* Typhimurium. PLoS Pathog. 11, e1005262.

**Owen, S.V.***, Wenner, N., Canals, R., Makumi, A., Hammarlöf, D.L., Gordon, M.A., Aertsen, A., Feasey, N.A., Hinton, J.C.D., 2017. Characterization of the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. Front. Microbiol. 8, 235. doi:10.3389/fmicb.2017.00235

Ashton, P.M.*, **Owen, S.V.*,** Kaindama, L., Rowe, W.P.M., Lane, C., Larkin, L., Nair, S., Jenkins, C., Pinna, E. de, Feasey, N., Hinton, J.C.D., Dallman, T., 2017. Public health surveillance in the UK revolutionises our understanding of the invasive *Salmonella* Typhimurium epidemic in Africa. Genome Med. 9:92. doi:10.1186/s13073-017-0480-7

- **Preprints (currently under review at international journals) resulting from work described in this thesis:**

Hammarlöf, D.L.*, Kröger, C.*, **Owen, S.V.***, Canals, R., Lacharme-Lora, L., Wenner, N., Wells, T.J., Henderson, I.J., Wigley, P., Hokamp, K., Feasey, N.A., Gordon, M.A., Hinton, J.C.D., 2017. The role of a single non-coding nucleotide in the evolution of an epidemic African clade of *Salmonella*. bioRxiv 175265. doi:10.1101/175265

- **Conference presentations resulting from work described in this thesis:**

September 2014: EMBO Viruses of Microbes 2014, Zurich, Switzerland

> Poster presentation entitled: **RNA-seq from 16 infection-relevant conditions resolves the gene expression profiles and possible contribution to virulence of temperate bacteriophages within the genome of an invasive African *Salmonella* Typhimurium strain**

August 2015: 21$^{ST}$ Biennial Evergreen Bacteriophage Meeting, Evergreen State College, Washington State, USA

> Poster presentation entitled: **BTP1: A highly active prophage carried by African *Salmonella* Typhimurium ST313**

April 2016: Microbiology Society Annual Conference 2016, Liverpool, UK

Poster presentation entitled: **The distribution of prophages in UK and African populations of *Salmonella* Typhimurium ST313**

July 2016: EMBO Viruses of Microbes 2016, Liverpool UK

15 minute offered oral presentation entitled: **Prolific spontaneous induction of novel prophage BTP1 in epidemic African *Salmonella* Typhimurium**

September 2016: 5th ASM Conference on *Salmonella*, Potsdam, Germany

Poster presentation entitled: **Novel prophage BTP1 is associated with epidemic African *Salmonella***

**Typhimurium ST313 but not present in diverse UK lineages**

April 2017: Microbiology Society annual conference 2017, Edinburgh, UK

15 minute offered oral presentation entitled: **Contrasting evolutionary fortunes for prophages in an epidemic *Salmonella* lineage**

April 2017: Centennial Celebration of Bacteriophage Research, Institute Pasteur, Paris, France

Poster presentation entitled: **Contrasting evolutionary fortunes for prophages in an epidemic *Salmonella* lineage**

# Contents

# List of figures

# List of tables

# List of abbreviations

| | |
|---|---|
| 3' UTR | Three prime untranslated region |
| 4/74 | *Salmonella enterica* serovar Typhimurium strain 4/74 |
| 5’ RACE | Five prime rapid amplification of cDNA ends |
| 5' UTR | Five prime untranslated region |
| aa | Amino acids |
| ACT | Artemis comparison tool |
| Amp | Ampicillin |
| AMR | Antimicrobial resistance |
| asRNA | Antisense RNA |
| *attB* | Bacterial attachment site |
| *attP* | Phage attachment site |
| BEAST | Bayesian Evolutionary Analysis Sampling Trees |
| BLAST | Basic Local Alignment Search Tool |
| bp | Basepairs |
| BRIG | Blast ring image generator |
| cDNA | Complementary DNA |
| CDS | Coding sequence |
| CFU | Colony forming units |
| Cm | Chloramphenicol |
| CPM | Counts per million |
| DCs | Dendritic cells |
| DE | Differentially expressed |
| $dH_2O$ | De-ionised water |
| Dig | Digoxigenin |
| DMEM | Dulbellcco’s Minimal Essential Medium |
| DMSO | Dimethyl sulfoxide |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleotide |
| dRNA-seq | Differential RNA-sequencing |
| dsDNA | Double stranded DNA |
| EDTA | Ethylenediaminetetraacetic acid |
| ESP | Early stationary phase |
| FLP | Flippase recombinase |
| FRT | Flippase recognition target |
| gDNA | Genomic DNA |
| GFP | Green fluorescent protein |
| Gm | Gentamicin |
| Gtr | Glycosyltransferase |
| HBD | Head binding domain |
| HGT | Horizontal gene transfer |
| HPD | Highest posterior density |
| IGB | Integrated Genome Browser |
| inSPI2 | SPI2-inducing conditions |
| iNTS | Invasive non-typhoidal Salmonella |
| Km | Kanamycin |
| kb | Kilobases |
| LB | Lysogeny broth |
| LEP | Late exponential phase |
| LPS | Lipopolysaccharide |
| LSP | Late stationary phase |

| | |
|---|---|
| MDR | Multi-drug resistant |
| MEP | Mid exponential phase |
| MLST | Multi locus sequence type |
| MOI | Molarity of infection |
| MRCA | Most recent common ancestor |
| mRNA | Messenger RNA |
| ncRNA | Non-coding RNA |
| Non-SPI2 | Non-SPI2 inducing conditions |
| $NoO_2$ | Anaerobic growth |
| nt | Nucleotides |
| NTS | Non-typhoidal *Salmonella* |
| $OD_{600}$ | Optical density at 600 nm |
| ORF | Open reading frame |
| OxyTc | Oxytetracycline |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PEG | Polyethylene glycol |
| PFGE | Pulse field gel electrophoresis |
| PFU | Plaque forming unit |
| PHE | Public Health England |
| qPCR | Quantitative real-time PCR |
| RBD | Receptor binding domain |
| RDAR | Red, dry and rough (colony morphology) |
| RNA | Ribonucleic acid |
| RNA-seq | RNA sequencing |
| RNAP | RNA polymerase |
| RNase | Ribonuclease |
| RNA-seq | High-throughput cDNA sequencing |
| SNP | Single nucleotide polymorphism |
| SPI | *Salmonella* pathogenicity island |
| sRNA | Small non-coding RNA |
| sSA | Sub-Saharan Africa |
| ssDNA | Single stranded DNA |
| ST19 | Sequence type 19 |
| ST313 | Sequence type 313 |
| STnc*XXXX* | *Salmonella* Typhimurium non-coding |
| TAE | Tris Acetate Ethylenediaminetetraacetic acid |
| Tc | Tetracycline |
| TEM | Transmission electron microscopy |
| TPM | Transcripts per million |
| tRNA | Transfer RNA |
| TSS | Transcription start site |
| TTSS | Type three secretion system |
| UV | Ultraviolet |
| WT | Wild-type |

# Chapter 1

## Introduction

## 1.1 The genus *Salmonella*; a short history

*Salmonella* is a genus of Gram-negative, rod-shaped bacteria belonging to the Enterobacteriaceae family, and more broadly to the class Gammaproteobacteria (Garrity et al., 2007). Like other members of the Enterobacteriaceae, *Salmonella* typically inhabit the intestinal tracts of vertebrates, however the ecology of *Salmonella* in the natural environment is not well-studied. In contrast, due to their ability to cause disease, particularly in higher mammals, the history of the study of the genus *Salmonella* is intertwined with that of mankind. Clinical descriptions suggestive of *Salmonella* disease can be found in Greek, Roman and ancient Chinese scriptures (Cunha, 2004) suggesting that *Salmonella* disease has plagued humanity throughout history. Indeed it has been suggested that Alexander the Great died from Typhoid fever (Cunha, 2004). More recently, DNA-based studies of archaeological material have implicated *Salmonella* in a variety of historical events, such as the fall of the Aztec civilisation in the 1500s (Vågene et al., 2017), and the ancient cause of death of a young woman in Trondheim, Norway in approximately 1200 (Zhou et al., 2017).

The first clinical descriptions of infections likely to be caused by *Salmonella* were made from the 1600s onwards and were known as either 'typhus' or 'typhoid'. Thomas Willis first accurately differentiated the two diseases in 1643 by identifying characteristic typhoid ulcers in the ilium (Cunha, 2004). However, confusion between typhus, caused by infection by members of the genus *Rickettsia,* and typhoid fever, caused by specific serovars of *Salmonella*, prevailed up until the late 1800s. The diagnostic challenge associated with invasive *Salmonella* disease persists to this day due to the non-specific nature of the clinical syndrome associated with blood stream *Salmonella* infection. *Salmonella* bacteria themselves were first observed in the 1880s, by Karl Eberth, who reported 'rod-shaped' organisms in the spleens of typhoid patients. *Salmonella* were first cultured by Georg Gaffky from patients in Germany in 1884 (Ellermeier and Slauch, 2006) and by Theobald Smith at the Bureau of Animal Industry in the USA. It was Smith, who proposed the name *Salmonella* in 1900, in honour of his mentor, American veterinarian Daniel E. Salmon. Theobald Smith was endeavouring to identify the cause of swine fever or hog cholera when he isolated what become known as *Salmonella* serovar Choleraesuis from a diseased pig (although the swine fever was later found to be caused by a virus). *Salmonella* was officially adopted as the genus name in 1934, when taxonomy was revised such that the genus *Bacterium* was abandoned and the genus *Bacillus* condensed to include

only spore-forming bacteria (The *Salmonella* Subcommittee of the Nomenclature Committee of the International Society for Microbiology, 1934).

## 1.2 Contemporary *Salmonella* phylogeny, taxonomy and typing

The genus *Salmonella* currently contains two recognised species, *S. enterica* and *S. bongori*. *Salmonella enterica* is subdivided into six subspecies (Tindall et al., 2005) (Figure 1.1):

- subsp. I *enterica*
- subsp. II *salamae*
- subsp. IIIa *arizonae*
- subsp. IIIb *diarizonae*
- subsp. IV *houtenae*
- subsp. VI *indica*

The peculiarity in the numbering of the sub-species is the result of *S. bongori* previously being classed as *S. enterica* subsp. V, and recognition that *arizonae* and *diarizonae* should form separate subspecies. The species and subspecies can be distinguished from one another based on biochemical reactions (biotyping) (Malorny et al., 2011).

*Salmonella* is closely-related to other genera in the Enterobacteriaceae such as *Escherichia, Shigella* and *Citrobacter* with which there is extensive genetic synteny and phenotypic similarities (Battistuzzi et al., 2004; Desai et al., 2013). *Salmonella* has been estimated to have diverged from a last common ancestor with *Escherichia* and *Shigella* around 100 million years ago (Doolittle et al., 1996). A defining difference between *Salmonella* and *Escherichia* is the presence of pathogenicity islands (known as SPIs for *Salmonella* pathogenicity islands) in the *Salmonella* genome. The SPIs encode secretion systems and translocatable effector genes which are critical for the intracellular lifestyle of *Salmonella. S. bongori,* which is predominantly associated with the infection of cold-blooded animals, contains only a limited repertoire of SPIs, and shares a more recent common ancestor with *Escherichia* than *S. enterica*. Consequently, *S. bongori* can be thought of as an evolutionary intermediate between *Salmonella* and *E. coli*, and the acquisition of various SPIs and modulation of metabolic function are thought to reflect the distinct niche adaption of *S. bongori and S. enterica* (Fookes et al., 2011).

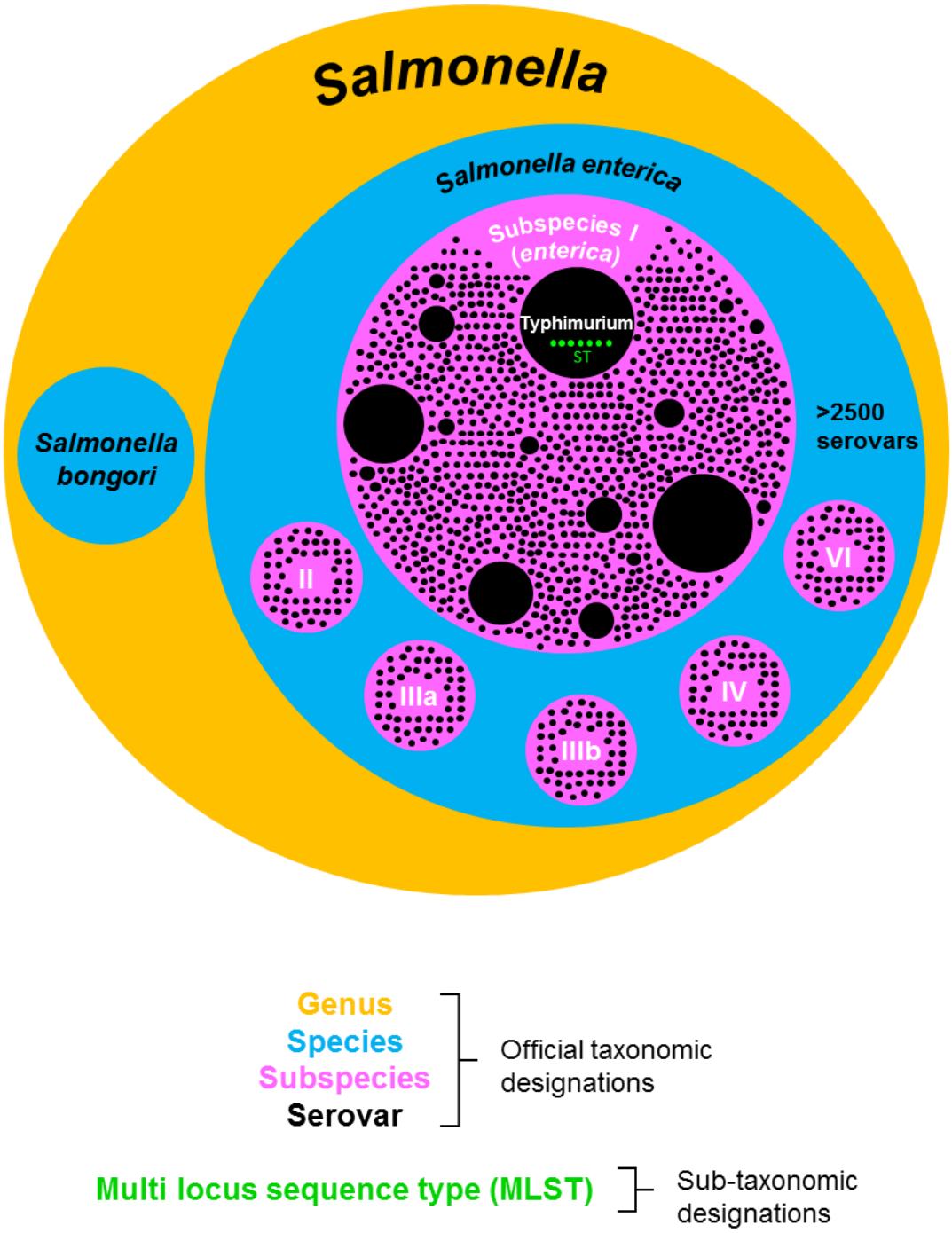



**Figure 1.1 Infographical representation of the taxonomic structure of genus *Salmonella*.** The relative size of black circles representing serovars illustrates that the vast majority of known serovars are rarely isolated, whilst a minority are commonly isolated. For example, in the USA in 2013, just 20 serovars were responsible for 70% of all laboratory-confirmed *Salmonella* infections (Centers for Disease Control and Prevention (CDC), 2016). The majority of sequence types associated with disease in mammals belong to subspecies I, enterica.

Below the level of subspecies, *Salmonella* are delineated or typed based on a method known as serotyping. The method of serotyping was developed by Fritz Kauffmann in the 1930s based on the work of Phillip Bruce White, and the ensuing taxonomic scheme was designated 'The Kauffmann-White Schema' (The *Salmonella* Subcommittee of the Nomenclature Committee of the International Society for Microbiology, 1934). The scheme represented a typing method for *Salmonella* based on the agglutination of serum in response to the diverse antigenic properties of the *Salmonella* cell surface (known as serology) (Kauffmann., 1940). Though serum agglutination had been used since 1896 to characterise *Salmonella* isolates (Widal, 1896), Kauffmann's approach represented the first standardisation of the method, and meant that *Salmonella* typing could be reproduced independently in different laboratories.

In the intermediate years before the introduction of the Kauffmann-White Scheme, numerous reports had been published describing typhoid-like bacilli in different animals, or in humans that showed characteristics distinct to the typical typhus bacillus (collectively known as paratyphoid bacilli), and the naming system calf-, mouse-, swine-, etc., typhus, had been commonly used (Tenbroeck, 1920). Consequently, a vast array of names, often incorporating the original host animal, are present in the early *Salmonella* literature but now considered to be synonyms. For example, *Bacillus typhimurium, Bacterium aertrycke, S. pestis caviae and S. psittacosis* are all considered synonymic for *Salmonella* Typhimurium (Weisbroth, 1979). The modern-day iteration of the Kauffman-White scheme is known as the White-Kauffman-Le Minor, in recognition of the contribution of Léon Le Minor, who described the majority of the known *Salmonella* serovars. Though the scheme has been updated several times in recent history, to allow for the description of an ever increasing number of serovars, the basis of the scheme remains unchanged since its inception by Franz Kauffman in the 1930s (The *Salmonella* Subcommittee of the Nomenclature Committee of the International Society for Microbiology, 1934; Grimont and Weill, 2007).

The serotyping scheme relies on the use of a set of standardised rabbit sera that has been raised against well-characterised *Salmonella* strains. Depending on the agglutination reaction to each test serum, bacteria are given an antigenic formula. The *Salmonella* antigens that contribute to serotyping are located on either lipopolysaccharide molecules (LPS), known as 'O' antigens, or the flagella- 'H' antigens (Grimont and Weill, 2007). The 'H' antigens are encoded by the *fliC* and *fljB*

genes which represent the motile phase 1 and phase 2 of *Salmonella* respectively. The formula is written in the format 'O' antigen(s): 'H' antigen phase 1 : H antigen phase 2 (Grimont and Weill, 2007). The latest volume of the White-Kauffmann-Le Minor scheme describes over 2,500 serovars of *Salmonella*. Historically, new serovars have been named after the cities in which they are first described, though as the number of serovars has increased, convention has been revised such that new serovars are given numerical antigenic formulae only (Grimont and Weill, 2007).

Approximately 60% of all the recognised *Salmonella* serovars belong to subspecies I (*enterica*) (Grimont and Weill, 2007), which is also the most important subspecies in relation to human health. However, the distribution of infections by the ~1,500 serovars of subsp. *enterica* is highly geometric and the majority of infections tend to be caused by only a handful of the most common serovars. For example, in the USA in 2013, just 20 serovars were responsible for 70% of all laboratory-confirmed *Salmonella* infections (Centers for Disease Control and Prevention (CDC), 2016). The vast majority of the >2,500 *Salmonella* serovars are extremely rare in the clinical setting.

### 1.2.1 Use of nomenclature in this thesis

To avoid confusion between species, subspecies and serovar names, it has become conventional to abbreviate the full taxonomic names of *Salmonella* simply to genus name and serovar (Grimont and Weill, 2007). For example, "*Salmonella enterica* subsp. *enterica* serovar Typhimurium" is conventionally abbreviated to "*Salmonella* Typhimurium", or "*S.* Typhimurium". For brevity, this convention will be adopted throughout this thesis.

Serotyping was initially designed as a pragmatic typing tool to aid distinction between *Salmonella* serovars in diagnostic and public health microbiology, rather than to inform phylogeny. A combination of the diversity of the genus *Salmonella*, the general phenotypic uniformity within serovars and technological limitations have meant that until recently, serovar-driven taxonomy has prevailed (Brenner *et al.,* 2000). Disease surveillance and diagnostic guidelines have consequently also been built on serology, which can be unhelpful when serovars are associated with multiple clinical syndromes (Herikstad *et al.,* 2002). Recent re-evaluation of *Salmonella* phylogeny using genetic methods has shown surprisingly high consistency between serology and phylogeny, with a few notable discrepancies such as the serovars *S.* Paratyphi B, *S.* Java, *S.* Paratyphi C and *S.* Choleraesuis, where serotyping does not accurately represent evolutionary relationships (Achtman et al., 2012).

*Salmonella* strains can be further differentiated into a number of 'phage types' based on their susceptibility to killing by a specific set of bacteriophages (viruses that infect bacteria) (Rabsch, 2007). Phage-typing correlates well with serology, because like the antibodies in serum, phages also recognise receptors on the bacterial surface. Phage typing has been used to type *Salmonella* at the sub-serovar level since the 1940s (Felix and Callow, 1943; Lilleengen, 1948) and the Anderson phage typing scheme has been used to classify *S.* Typhimurium strains in labs worldwide since 1966 (Rabsch et al., 2011). Phage typing has been utilised particularly effectively for epidemiological surveillance, where it is expected that strains isolated from cases associated with a point source outbreak will have the same phage type (Baggesen et al., 2010).

Though serology and phage typing have played important roles in *Salmonella* disease epidemiology, a common drawback of these approaches is that they rely on morphological properties of *Salmonella* which may not reflect real evolutionary relationships. Pulse field gel electrophoresis (PFGE) is an early way of typing *Salmonella* based on properties of the genome (Goering, 2010). PFGE is a method of high resolution separation of DNA molecules in a gel matrix by the application of an electric field of alternating direction. Genomic DNA is digested with the XbaI restriction enzyme and separated such that the pattern of DNA molecules of varying lengths (known as the DNA fingerprint) reflects the number and location of XbaI-restriction sites in the genome. PFGE has been applied to the global surveillance of food borne disease by pathogens such as *Salmonella*, *Shigella* and *E. coli*, as pioneered by the PulseNet organisation. However, more discriminatory genome-based methodologies have now largely replaced PFGE-typing for *Salmonella* (Swaminathan et al., 2001).

Multi-Locus Sequence Typing (MLST) is a method that has been employed to supplement, or replace, traditional *Salmonella* typing methods such as serology and phage typing or PFGE (Achtman et al., 2012). Different MLST schemes have been established for many bacterial phyla, but in *Salmonella* the original scheme uses the DNA sequences (or alleles) of seven highly conserved so-called 'housekeeping' loci. The genes used in the *Salmonella* MLST scheme are: *thrA, purE, sucA, hisD*, *aroC*, *hemD* and *dnaN* (Maiden et al., 2013). The sequence of each loci is determined for a given isolate. Traditionally this is accomplished by targeted PCR amplification and Sanger sequencing, but increasingly MLST typing is carried out *in silico* using whole genome sequence data. Each known allelic variant is assigned a number, and each

new combination of the 7 alleles is given a numerical 'type'. As STs are resolved at the level of single nucleotides, MLST is highly discriminatory and is more likely than serology to recognise evolutionary groupings. For serovars such as *S.* Typhimurium, which contain a high degree of allelic variation, MLST is used to discriminate isolates at the sub-serovar level (Kingsley et al., 2013; Okoro et al., 2012). For example, the most common sequence type of *S.* Typhimurium worldwide causing disease in humans is (ST) 19, though more than 100 sequences types of *S.* Typhimurium have currently been defined (EnteroBase: http://enterobase.warwick.ac.uk [Accessed 16 June 2017]) (Figure 1.2).

As the cost of whole genome sequencing has plummeted (Muir et al., 2016), typing *Salmonella* using whole-genome based approaches has become increasingly common. Since 2014, the UK public health agency, Public Health England, has been using whole genome sequencing as a routine typing tool for public health surveillance of *Salmonella*, and adopting *in silico* MLST as a replacement for traditional serotyping (Ashton et al., 2016). More discriminatory schemes have subsequently been applied to collections of whole genome sequences, such as core-genome and whole-genome MLST (cgMLST, wgMLST) which use the sequence of 3,002 and 21,065 loci respectively (EnteroBase: http://enterobase.warwick.ac.uk [Accessed 1 January 2016]). Lastly, classifying *Salmonella* based on statistical clustering of whole genome phylogenies has become increasingly common, and has recently been used to classify global *S.* Typhi populations (Wong et al., 2016).

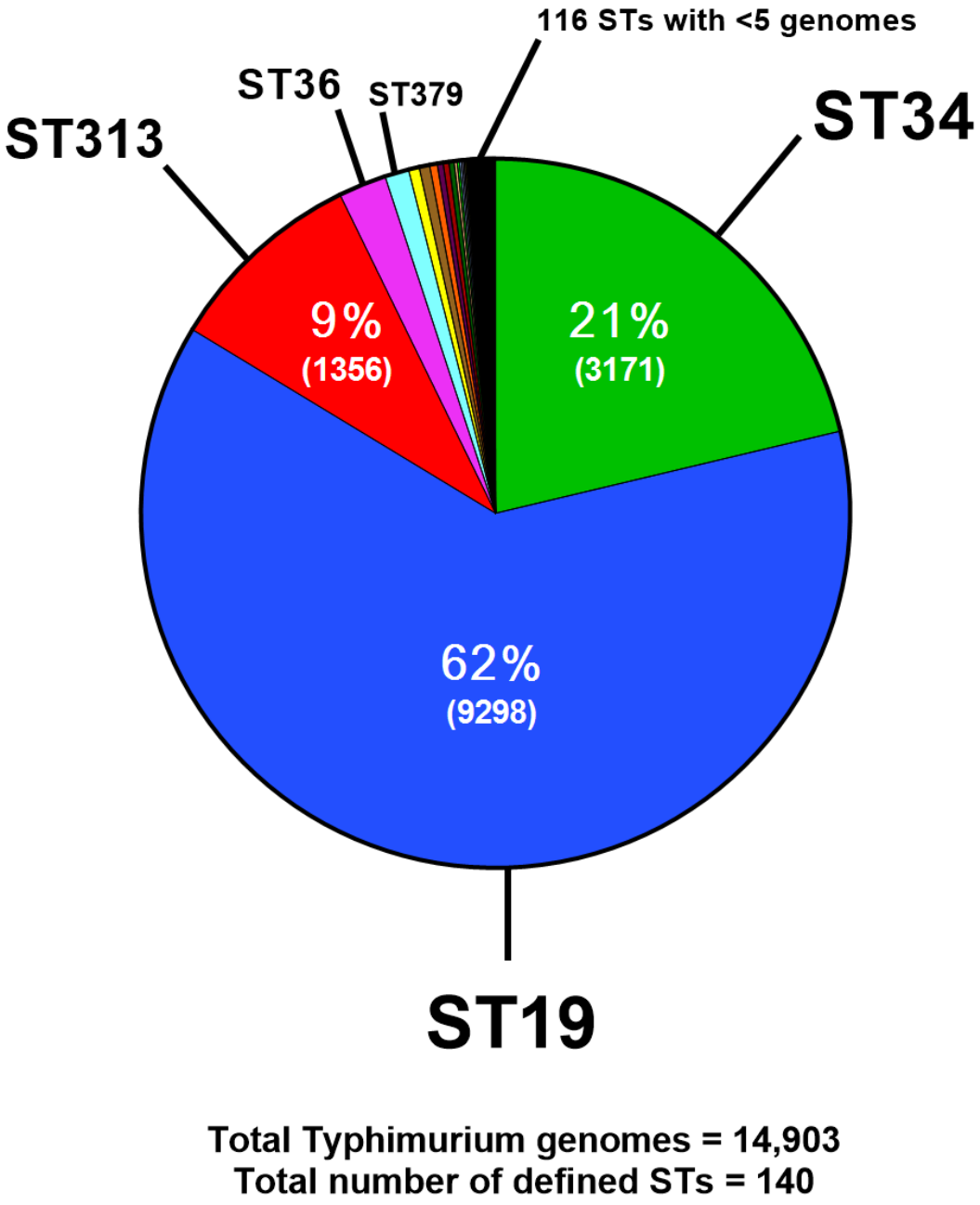



**Figure 1.2 All sequenced *S*. Typhimurium genomes in the Enterobase genome database shown by multi-locus sequence type (MLST).** The pie chart shows the proportion of all sequenced *S.* Typhimurium genomes belonging to each MLST sequence type. The percent and proportion are shown for the largest 3 sequence types. Data obtained from EnteroBase: **http://enterobase.warwick.ac.uk** [Accessed 20 June 2017]

## 1.3 Clinical syndromes and host specificities associated with *Salmonella*

*Salmonella* causes a huge range of disease in humans, including gastroenteritis, sepsis, meningitis, abscess, osteomyelitis and mycotic aneurysms. Such is the breadth of disease and the diversity of serovars, clinicians have found it useful to consider *Salmonella* in terms of the two most prominent clinical syndromes associated with *Salmonella*, these being gastroenteritis and invasive disease/bloodstream infection. This has led them to associate specific serotypes with particular disease types. This crude distinction is not without utility, but neither is it without problems. It follows the assumption that isolates belonging to the same serovar are homogenous in aspects of disease syndrome and host range. This convention existed because certain serovars were associated with a specific host range, and with particular infection severities (Jones et al., 2008). Typically there is an inverse relationship between the breadth of host range and the severity of infection of *Salmonella* serovars (Figure 1.3). For example, *Salmonella* serovars with broad host ranges tend to be associated with self-limiting gastrointestinal disease, which is less severe than the extra-intestinal, often fatal, infections caused by serovars restricted to a single host. It has been suggested that this is the result of an evolutionary trade-off between niche specialisation and niche versatility (Bäumler and Fang, 2013)

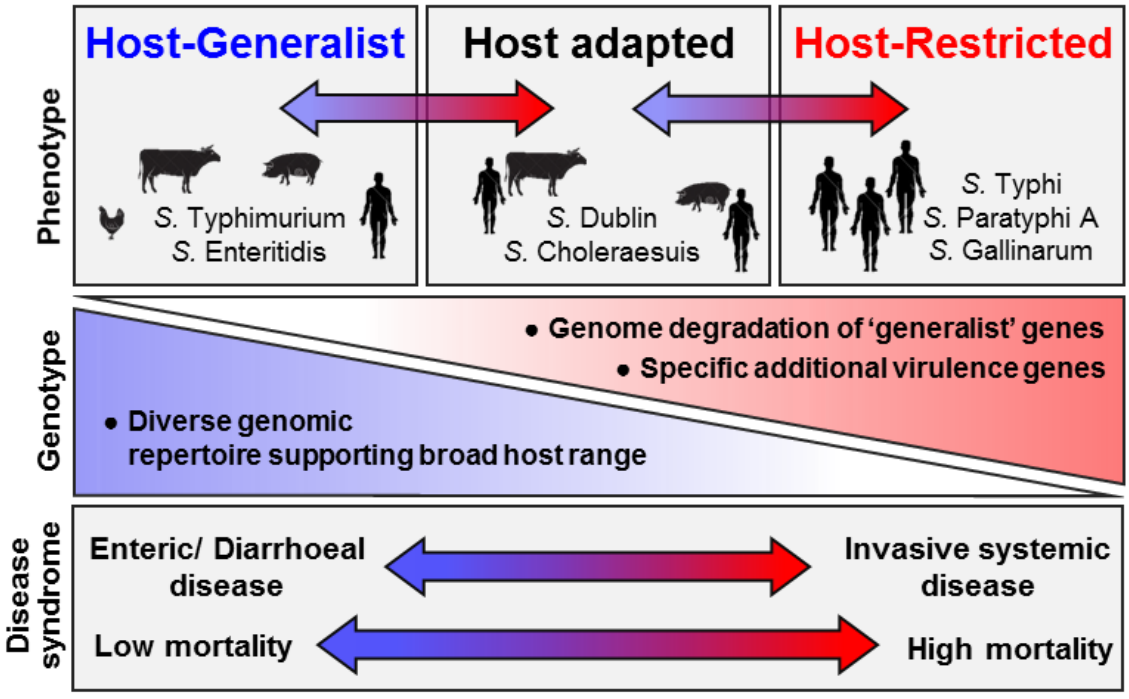



**Figure 1.3 The *Salmonella* host adaption and disease syndrome paradigm.**

The most prominent example of this is the case of Typhoid fever, a clinical syndrome restricted to humans caused by serovars *S.* Typhi, *S.* Paratyphi A, and to a lesser extent, *S.* Paratyphi B and C. Consequently these human-restricted pathogens associated with invasive (systemic) disease are considered to be 'typhoidal' serovars. The clinical syndrome caused by these serovars is known as enteric fever or Typhoid fever (occasionally the term Paratyphoid fever is used when the causative agent is serovar *S.* Paratyphi A, B or C, though the diseases are thought to be clinically indistinguishable) (Crump and Mintz, 2010). Serovars that are considered host-restricted in other animal hosts include *S.* Gallinarum, *S.* Typhisuis and *S.* Abortusequi which cause invasive disease in chicken, pig and horse respectively (Uzzau et al., 2000).

This distinction is useful in the context of Typhoid fever. What is less useful is the convention of considering the remainder of serovars as being "non-typhoidal" and by association not associated with invasive disease. Instead, the majority of serovars are associated with a spectrum of host restriction and generalism and clinical disease. Certain serovars of *Salmonella* are described as 'host-adapted', and show specialisation, though not total restriction, to a specific host. Examples of 'host-adapted' serovars are *S.* Choleraesuis and *S.* Dublin, which primarily infect pigs and cows respectively, but can also cause invasive disease in humans (Uzzau et al., 2000).

At the opposite end of the spectrum, broad-host range "generalist" *Salmonella* serovars tend to cause mild infection in a variety of hosts, manifesting as gastroenteritis in humans. Strictly in regard to human health, all *Salmonella* that are not primarily associated with invasive disease in humans (the 'typhoidal' *Salmonella* already discussed) are loosely described as 'non-typhoidal *Salmonella*' (NTS). These "non-typhoidal" *Salmonella* serovars typically have a broad host-range, and the majority of human cases are foodborne, often originating from zoonotic reservoirs (Mead et al., 1999). NTS are responsible for a significant public health burden worldwide and *S.* Typhimurium and *S.* Enteritidis are the predominant serovars observed in clinical cases in most countries. For example, in England & Wales, 48.7% of the isolates referred to the *Salmonella* Reference Service in 2012 were *S.* Typhimurium or *S.* Enteritidis (Public Health England data- *Salmonella* frequency by serotype 2000 to 2010). The global burden of non-typhoidal *Salmonella* caused gastroenteritis has been estimated to be 150,000 deaths per year (Majowicz et al., 2010).

Serotyping has historically been a useful tool for dividing the diverse genus *Salmonella* into clinically-relevant groupings. However, this simplified view of typhoidal *vs* non-typhoidal *Salmonella* disease becomes complicated by host factors which also play an important role in *Salmonella* disease severity and host susceptibility. For example, in immunodeficient hosts, non-typhoidal serovars such as *S.* Typhimurium are prominent causes of systemic infection, termed invasive non-typhoidal *Salmonella* disease (iNTS) (Gordon, 2011). Specific human genetic variants (primary immunodeficiencies) are known to increase susceptibility to *Salmonella* infection (Gilchrist et al., 2015), but immunodeficiency due to an underlying disease status (secondary immunodeficiency) such as that associated with human immunodeficiency virus (HIV) infection has become increasingly associated with iNTS disease. In fact, iNTS was one of the first HIV/AIDS defining illnesses (Feasey et al., 2012; Gruenewald et al., 1994). Though iNTS disease does not represent a large health burden in the developed world where hygiene standards are high and prevalence of immunodeficiency is relatively low, in sub Saharan Africa (sSA), iNTS is responsible for an estimated 390,000 deaths *per* year (Ao et al., 2015) (Figure 1.4).


**Figure 1.4 Map of the world showing the estimated annual frequency of mortality due to iNTS disease by continent.** Based on estimates by Ao et al., 2014. The numerical estimates of annual deaths per continent are: N. Africa (MidEast) 723, Africa 388,555, Asia (Oceania) 2784, SE Asia 94,452, Europe 152,638, Americas 42,162.

## 1.4 iNTS disease in sub Saharan Africa & *S.* Typhimurium ST313

Reports of bacteraemia in children caused by non-typhoidal *Salmonella* in sub Saharan Africa (sSA) began to appear in the published literature in the 1960s (Ao et al., 2015; Uche et al., 2017). With the subsequent AIDS epidemic, NTS began to be frequently isolated from bloodstream infections in adults as well as children, and NTS was described as a common HIV-associated pathogen in sSA (Gilks et al., 1990). iNTS disease has remained a significant health burden, and is estimated to be responsible for around 39% of all community-acquired blood stream infections in sSA (Uche et al., 2017). Serovars *S.* Typhimurium and *S.* Enteritidis are most frequently isolated from iNTS patients (MacLennan and Levine, 2013; Uche et al., 2017), though a number of other serovars have been associated with iNTS disease in specific locations including *S.* Isangi (South Africa), *S.* Concord (Ethiopia), *S.* Infantis (Ethiopia), *S.* Dublin (Mali), *S.* Stanleyville (Mali), *S.* Heidelberg (Malawi) and *S.* Bovismorbificans (Malawi) (Bronowski et al., 2013; Gordon, 2011; Tennant et al., 2010; Uche et al., 2017; Wain et al., 2013). Recent meta-analysis suggests that *S.* Typhimurium has been responsible for 48% of all reported iNTS disease cases (Uche et al., 2017) and for this reason it has been the focus of research into the pathogenicity of iNTS disease associated *Salmonella*.

Whole genome sequencing of *S.* Typhimurium isolates collected from patients with iNTS disease in Malawi initially identified a novel sequence type (ST), ST313, of *S.* Typhimurium in 2009 (Kingsley et al., 2009) (Figure 1.5A). Subsequent whole genome-based study of isolates from across the African sub-continent confirmed the existence of two distinct phylogenetic lineages of ST313, and spatio-temporal phylogenetic reconstruction estimated that lineage 1 emerged *circa* 1960 in south west Africa, whereas lineage 2 emerged *circa* 1977 in Malawi (Okoro et al., 2012) (Figure 1.5B). Both lineages are associated with antimicrobial resistance (AMR) mediated by differing Tn*21*-like integrons on the virulence plasmid pSLT (Kingsley et al., 2009), and it was proposed that clonal replacement of lineage 1 by lineage 2 had occurred, driven by the acquisition of chloramphenicol resistance by lineage 2 (Okoro et al., 2012). The finished genome of lineage 2 strain D23580 (Accession Number: FN424405) isolated from a child in Malawi, published by Kingsley et al., 2009 is now generally used as the reference strain for ST313.

Since its description in 2009, numerous publications have reported ST313 *S.* Typhimurium to be a leading cause of iNTS disease in countries across sub-Saharan Africa (Summarised in Table 1.1). Most recently, ST313 have been isolated in Brazil (Almeida et al., 2017). The strong association between ST313 and iNTS disease compared to other sequence types of *S.* Typhimurium, and other serovars of *Salmonella,* has led to the hypothesis that ST313 *S.* Typhimurium have adapted to cause systemic disease in the sSA human niche.

Phylogenetic analysis indicates that ST313 is closely-related to ST19- the 'archetypal' *S.* Typhimurium sequence type responsible for self-limiting gastroenteritis worldwide, and containing well studied strains LT2 and SL1344 (Kingsley et al., 2009).

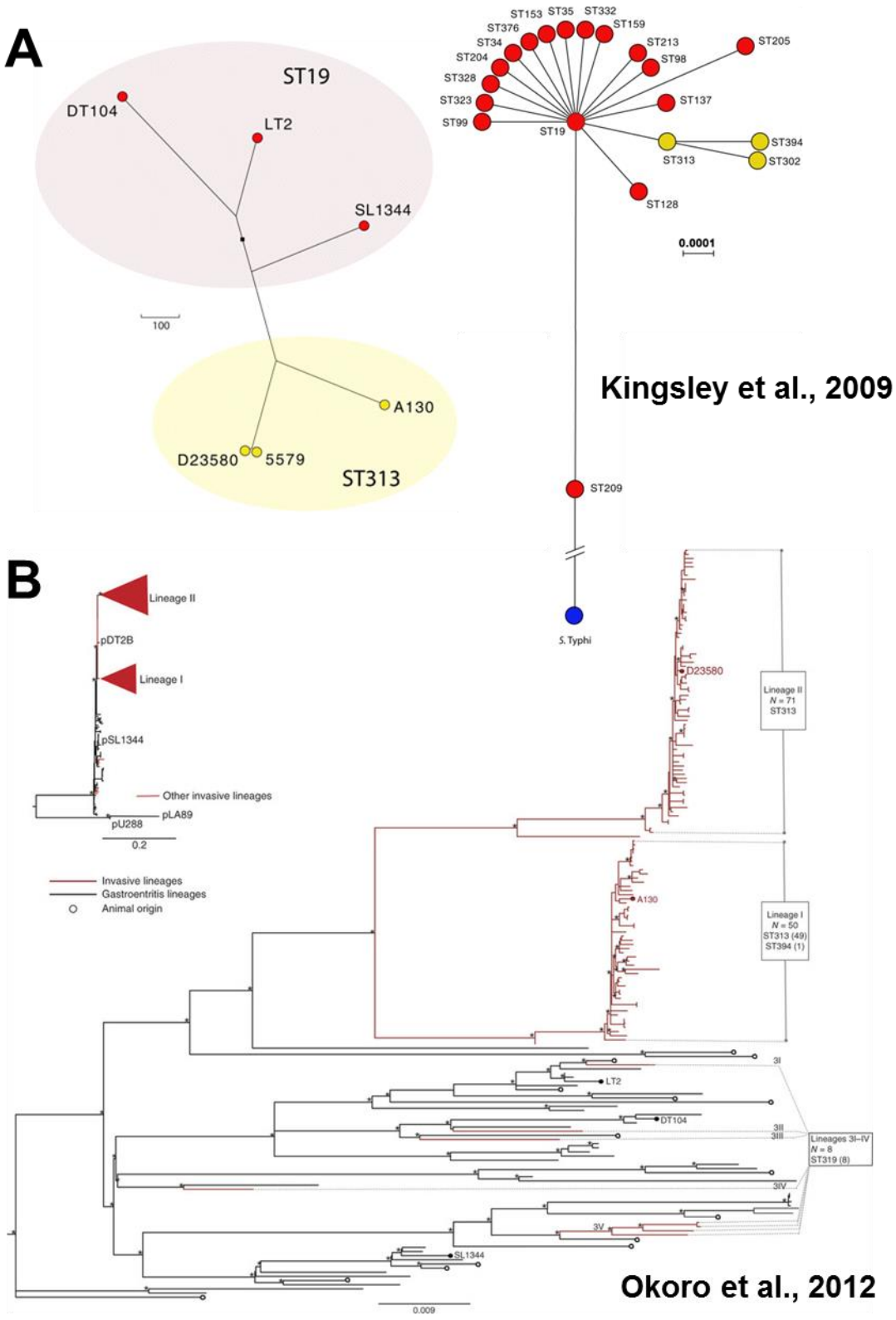



**Figure 1.5 Two landmark papers describing the iNTS disease-associated *S.* Typhimurium ST313.** A. Figures reproduced from Kingsley et al., 2009. (left) Radial phylogeny showing the phylogenetic relationship between 6 strains of ST313 and ST19. (right) Radial phylogeny showing the phylogenetic relationship between sequence types (ST) of *S.* Typhimurium. B. Figure reproduced from Okoro et al., 2012 showing a phylogenetic tree of ST313 and other *S.* Typhimurium isolates. The tree shows that the population structure of ST313 consists of two discrete lineages**.**

## Table 1.1 The *S.* Typhimurium ST313 disease epidemiology literature

| Authors | Year | Associated Country | Summary |
|---|---|---|---|
| **Publications describing ST313 disease epidemiology** | | | |
| **Almeida et al.** | 2017 | Brazil | Antimicrobial-sensitive ST313 isolates from food, faeces and blood in Brazil. |
| **Kariuki & Onsare** | 2015 | Kenya | Emergence of ceftriaxone resistant ST313 in Kenya |
| **Keddy et al.** | 2015 | South Africa | ST313 in *S.* Africa is associated with neonates and $HIV^+$ adults |
| **Moon et al.** | 2015 | Mozambique | ST313 is a prevalent cause of iNTS in Mozambique |
| **Kariuki et al.** | 2015 | Kenya | Identification of an ESBL-producing IncHI2 plasmid in ceftriaxone-resistant ST313 isolates from Kenya |
| **Feasey et al.** | 2014 | Malawi | Identification of an ESBL-producing IncHI2 plasmid in ceftriaxone and ciprofloxacin resistant ST313 isolates from Malawi |
| **Ley et al.** | 2014 | DRC | ST313 is a prevalent cause of iNTS in the DRC |
| **Leekitcharoenphon et al.** | 2013 | Nigeria, DRC | ST313 is a prevalent cause of iNTS in the DRC and Nigeria |
| **Paglietti et al.** | 2013 | Zimbabwe | ST313 is a prevalent cause of iNTS in the DRC |
| **Okoro, et al.** | 2012 | Malawi, Kenya, Mozambique, Uganda, DRC, Nigeria and Mali | ST313 is a cause of iNTS disease across Africa and is composed of 2 discrete phylogenetic lineages |
| **Okoro et al.** | 2012 | Malawi | Recurrence of iNTS disease by ST313 in Malawi is mainly recrudescence not reinfection |
| **Boyle et al.** | 2011 | Kenya, Malawi | ESBL-producing ST313 from Kenya and Malawi carry the $bla_{OXA-1}$ gene |
| **Kingsley et al.** | 2009 | Kenya, Malawi | iNTS disease in Kenya and Malawi is associated with the novel sequence type (ST)313 |
| **Publications reviewing ST313 disease epidemiology** | | | |
| **Wain et al.** | 2013 | - | Review of iNTS disease epidemiology and epidemiological questions remaining |
| **Feasey et al.** | 2012 | - | Review of the epidemiology and clinical features of iNTS disease |
| **de Jong et al.** | 2012 | - | Immunology-focused review of iNTS disease |

Initial genome-based studies proposed that the link between *S.* Typhimurium ST313 and iNTS disease in sSA was that, compared with the generalist *S.* Typhimurium ST19, ST313 has adapted to an extra-intestinal/invasive lifestyle via genome degradation (Kingsley et al., 2009; Okoro et al., 2015). This deduction was prompted by the finding of an accumulation of pseudogenes in pathways associated with gastrointestinal colonization, as observed in host-restricted *Salmonella* serovars such as *S.* Typhi, and in other notable human pathogens such as *Yersinia pestis, Shigella* spp, *Mycobacterium leprae* and *Bordetella pertussis* (Cole et al., 2001; McClelland et al., 2004; McNally et al., 2016; Parkhill et al., 2001; The et al., 2016; Yang et al., 2005). Of the 44 novel pseudogenes and deletions found in ST313-representative strain D23580, compared to ST19 strain LT2, 26 are also degraded in *S.* Typhi (Kingsley et al., 2009). For example, a 1694 bp deletion containing the *allP* and *allB* genes was identified in D23580. These genes are involved in allantoin metabolism, a pathway that has also been lost in the host-restricted *Salmonella* serovars *S.* Typhi, *S.* Paratyphi A and *S.* Gallinarum. Allantoin is a breakdown product of purines in lower mammals, but not produced in humans and avian species, implying that loss of the allantoin metabolism pathway could indicate evolution towards a human/avian restricted niche (Kingsley et al., 2009).

As well as genome degradation, Kingsley et al. reported the presence of two novel prophages, denoted BTP1 and BTP5 (for Blantyre Prophage) as well as the addition of an array of antibiotic resistance genes carried on a Tn*21*-like integron on the pSLT virulence plasmid (Kingsley et al., 2009). Though the structure of the MDR integrons differed in ST313 lineage 1 and 2, both lineages were reported to carry the BTP1 and BTP5 prophages. Subsequent analysis by Okoro et al. showed that the prophage repertoire of ST313 isolates is highly conserved in isolates from both lineages and originating from across the African continent (Kingsley et al., 2009). Interestingly, other publications characterising iNTS associated *S.* Bovismorbificans and *S.* Enteritidis and respective gastroenteritis-associated comparator strains showed a similar pattern, with the major genetic differences being the plasmid and phage elements (Bronowski et al., 2013; Feasey et al., 2016).

The BTP1 prophage carried by ST313 contains the *gtrC* gene, which mediates acetylation of the O-antigen (Kintz et al., 2015). Consequently, the O-antigen morphology of ST313 strains differs from ST19 strains which do not contain the BTP1 prophage. O-antigen modification systems are commonly carried by temperate phages to confer superinfection immunity to the lysogen, and these systems have

also been shown to play a role in bacterial virulence (Davies et al., 2013). Despite genomic degradation that mirrors that of host-restricted *S.* Typhi (McClelland et al., 2004), numerous reports have shown that ST313 are capable of causing infection in a number of animal models and implicating a possible avian reservoir; *S.* Typhimurium ST313 causes invasive disease in chickens (Parsons et al., 2013), and several studies have reported virulent infection in mice (Herrero-Fresno et al., 2014; Okoro et al., 2015; Singletary et al., 2016; J. Yang et al., 2015). Herrero-Fresno et al. (2014) and Yang et al. (2015) reported that ST313 were more invasive than ST19 in the BALB/c mouse and C57/BL6 mouse respectively, and Singletary et al (2015) reported that ST313 disseminated to the mesenteric lymph nodes more rapidly than ST19 in the streptomycin-treated mouse. However, Okoro et al. (2015) reported that ST313 cause less inflammation than ST19 in the streptomycin-treated mouse model, whereas Singletary et al (2015) reported no difference in inflammation caused by ST313 and ST19 in the same model. ST313 have additionally been reported to produce less inflammation than ST19 in the bovine ligated ileal loop infection model (Okoro et al., 2015), whereas no differences were observed in the rhesus macaque ligated ileal loop infection model (Singletary et al., 2016). Only one genetic mechanism has been identified that could explain the reported differences in animal infection models. The *sseI* gene is a pseudogene in ST313 lineage 2 (Kingsley et al., 2009). A recent study showed that the loss of the functional *Salmonella* pathogenicity island 2 (SPI-2) effector SseI protein facilitates hyperdissemination from the gut to systemic sites via migratory dendritic cells (DCs) in the streptomycin pre-treated mouse model (Carden et al., 2017). The lack of agreement between various animal infection models utilised to study ST313, together with the lack of genetic mechanisms, has fuelled debate as to whether ST313 are truly more invasive, or whether the association between ST313 and iNTS disease is predominantly due to host factors (Lokken et al., 2016).

### 1.4.1 A note on the use of the term “invasive”

The term ‘invasive’ was originally applied to iNTS as a clinical term i.e. capable of ‘invading’ the blood stream to cause systemic disease. However the term has been widely confused with the cellular biology definition of ‘invasive’ i.e. when a bacterium is capable of actively invading eukaryotic non-phagocytic cells such as epithelial cells and fibroblasts. The distinction between these two definitions is important, because a pathogen which is more clinically invasive may not necessarily exhibit greater cellular invasiveness.

As well as animal infection models, numerous cellular models have been used to study ST313 pathogenesis. The infection behaviour (including invasion, proliferation and cytotoxicity) of ST313 has been studied in various murine and human cells, including human epithelial cell lines (HeLa), human macrophage cell lines (THP-1), murine macrophage cell lines (J774) and human and mouse primary (bone-marrow derived) macrophages (Almeida et al., 2017; Carden et al., 2015; Herrero-Fresno et al., 2014; Ramachandran et al., 2015) (Table 1.2). As with animal infection models however, no consistent infection phenotype has been reported to distinguish ST313 from ST19, though some patterns have emerged. Generally, ST313 replicate better in murine macrophages than ST19 strains, and illicit less cytotoxicity (Carden et al., 2015; Herrero-Fresno et al., 2014; Ramachandran et al., 2015). There is conflicting data on the ability of ST313 to invade epithelial cells compared to ST19 (Almeida et al., 2017; Carden et al., 2015; Herrero-Fresno et al., 2014). The lack of consensus between studies may reflect the use of various different bacterial strains to represent ST313 and ST19 (Table 1.2). ST313 have also been reported to require a higher concentration of complement for serum-mediated bacterial killing than ST19 or *S.* Paratyphi B (Goh and MacLennan, 2013) and to be more resistant to high fluid shear forces than ST19, which may indicate adaption toward survival in the blood stream (Yang et al., 2016).

Phenotypes not directly related to infection have also been described (Table 1.2). Yang et al. (2015) reported that ST313 was more acid resistant than ST19 strains. Ramachandran et al. (2016) showed ST313 form less surface-associated biofilm and exhibit less long-term survival under desiccating conditions than ST19. Furthermore, ST313 cannot form the biofilm-associated RDAR (red dry and rough) morphology at 25 °C, and are defective in stationary-phase catalase activity, due to point mutations in the *bcsG* and *katE* genes respectively (Singletary et al., 2016). It has been suggested that, taken together, these phenotypes may support the hypothesis that ST313 has lost an adaption to an environmental niche that is extant in ST19. Finally, ST313 has been shown to have altered motility compared to ST19. Yang et al. (2015) reported that ST313 strains showed higher swimming motility than ST19, whereas Ramachandran et al. (2015) reported reduced motility in ST313 strains compared to ST19, and Carden et al. (2015) showed reduced flagellin expression in ST313 compared to ST19 (Carden et al., 2015; Ramachandran et al., 2015; J. Yang et al., 2015).

Despite considerable research into the effects of core genome degradation, along with *in vivo* and *in vitro* phenotypes that differentiate ST313 from other clades of *S.* Typhimurium, a mechanism that would confer an enhanced ability of *S.* Typhimurium ST313 to cause systemic disease has not yet been described. It is therefore clear that our understanding of *S.* Typhimurium ST313 pathogenesis, and its association with iNTS disease is incomplete.

## Table 1.2 The *S.* Typhimurium ST313 phenotypic literature

| Authors | Year | Key Strains used | Phenotypes described |
|---|---|---|---|
| **Publications describing ST313 phenotypes** | | | |
| **Almeida et al.** | 2017 | Brazilian ST313 & ST19 isolates, SL1344 | Brazilian ST313 isolates are heterologous in invasion of epithelial cells (HeLa) and replication in mouse macrophages (J774) |
| **Owen et al.** | 2017 | D23580 | **Chapter 3** |
| **Carden et al.** | 2017 | D23580, SL1344 | The pseudogenisation of *sseI* in lineage 2 ST313 facilitates hyperdissemination from the gut to systemic sites via migratory DCs in the streptomycin pre-treated mouse model (C57BL/6). |
| **Yang et al.** | 2016 | D23580, χ3339[1] | ST313 are more resistant to high fluid shear conditions (likely to be encountered in the bloodstream) than ST19 |
| **Ramachandran et al.** | 2016 | D23580, SL1344 + Malian ST313 & ST19 strains | ST313 form less biofilm than ST19 and are less resistant to desiccation |
| **Singletary et al.** | 2016 | D23580, 14028 | Point mutations in *katE* and *bcsG* in ST313 result in loss of stationary phase catalase production and RDAR morphology formation, linked to environmental persistence. No differences between ST313 and ST19 in rhesus macaque ligated ileal loop infection model but ST313 more rapidly disseminates to the mesenteric lymph nodes in the streptomycin pre-treated mouse. |
| **Yang et al.** | 2015 | D23580, A130, SL1344 | ST313 are more acid resistant, more motile, and more invasive in BALB/c mice than ST19. |
| **Carden et al.** | 2015 | D23580, A130, SL1344, DT104 | ST313 invade epithelial cells (HeLa) less efficiently and are less cytotoxic, and illicit less inflammasome activity in primary mouse macrophages. Expression of *sopE2* and *fliC* is lower in ST313 than in ST19. |
| **Kintz et al.** | 2015 | D23580 | ST313 possess a modified O-antigen due to acetylation by the BTP1 prophage-encoded *gtrC* gene. |
| **Okoro et al.** | 2015 | D23580, A130, 5597 (Kenya), 5579 (Kenya), SL1344 | ST313 have a distinct metabolic profile compared to ST19. ST313 illicit less inflammation in streptomycin pre-treated mouse and bovine ligated ileal loop infection model. |
| **Ramachandran et al.** | 2015 | D23580, SL1344 + Malian ST313 & ST19 isolates | ST313 are less motile than ST19, are more virulent in cultured macrophages (THP-1), and primary mouse and human macrophage models compared to ST19. ST313 are less cytotoxic compared to ST19. |
| **Herrero-Fresno et al.** | 2014 | 02-03/002 (DRC), 4/74, 14028 | ST313 possess the *st313-td* gene, which increases replication in murine macrophages (J774) and increases invasiveness in mice (C57/BL6) relative to ST19, but does not affect invasion of epithelial cells (HeLa). |
| **Parsons et al.** | 2013 | D23580, Q456, 4/74, F98 | ST313 causes invasive infection of chickens that is more rapid than ST19. ST313 exhibits lower gastrointestinal colonization. |
| **Goh et al.** | 2013 | D23580, LT2, CVD1901 (Paratyphi A) | ST313 require a higher concentration of complement for serum-mediated bacterial killing compared to ST19 and Paratyphi A. |
| **Kingsley et al.** | 2009 | D23580 | ST313 have undergone genome degradation analogous to that found in host-restricted *Salmonella* such as *S.* Typhi or *S.* Gallinarum. |
| **Publications reviewing ST313 phenotypes** | | | |
| **Lokken et al.** | 2016 | - | Review of reported phenotypes of ST313 and host factors associated with iNTS disease. |

[1]Animal passaged derivative of SL1344

## 1.5 An introduction to bacteriophages

Bacteriophages or ‘phages’ are viruses that infect bacteria. Phages were first discovered independently by the British pathologist Frederick William Twort in 1915, and the French-Canadian microbiologist Félix d’Hérelle in 1917 (Twort, 1915; d’Hérelle, 1917). Both scientists reported a filter-passing agent capable of lysing bacteria, that could be transmitted from infected to uninfected cultures of the same bacteria, though it was d’Hérelle who coined the term bacteriophage, literally meaning “eater of bacteria”. d’Hérelle subsequently began isolating bacteriophages against gastroinstinal pathogens including *E. coli* and *Shigella* species, and was the first to isolate phages against *S.* Typhi and *S.* Typhimurium (d’Hérelle, 1922).

Not long after the initial descriptions of bacteriophages by Twort and d’Hérelle, other microbiologists reported the appearance of bacteriophage in cultures spontaneously, without infection, and without evidence of culture lysis. These bacteriophages could subsequently infect and lyse isogenic strains of the same bacterial species, but could not infect the strain from which they had originated (Lisbonne and Carrére, 1922; Bordet, 1923). The cultures which produced these bacteriophages were termed lysogenic- literally meaning ‘generating lysis’ (Lwoff, 1953). More than 25 years before the discovery of DNA as a heritable genetic material, Jules Bordet wrote on the subject of a bacterial lysogen that “The faculty to produce bacteriophage is inscribed in the heredity of the bacterium; it is ‘inserted’ in the normal physiology of the bacterium” (Bordet, 1925; Lwoff, 1953). Bordet had foreseen what would later inspire a breakthrough in molecular biology; the origin and all characteristics of a phage are determined by the nucleic acids that make up its genome, which in the case of temperate phages (those originating from a lysogen) can become integrated into the genome of the host (Lwoff, 1953). This latent phage state within the heritable material of the host was termed a pro-bacteriophage, or more commonly, prophage (Lwoff and Gutmann, 1950). All phages can be classed as either temperate, or lytic, depending on whether or not they possess the faculty to enter the latent prophage state.

Despite their early discovery, phages were not confirmed to be true viruses (as opposed to another factor, such as a lytic enzyme) until the application of electron microscopy, and were first visualised by the German doctor Helmut Ruska in 1940 (Ruska, 1940). Subsequently, phages became important tools in the development of techniques in molecular biology, and the pioneering experiments of Alfred Hershey and Martha Chase in 1952 using phage T2 contributed to the discovery of DNA as a heritable genetic material (Hershey and Chase, 1952). Study of temperate phages, in

particular phage lambda (λ), underpinned the discovery of important molecular biology concepts such as the regulation of gene expression and genetic recombination, as well as contributing to the field of genetic engineering (Casjens and Hendrix, 2015; Murray, 2006). Due to the depth in which it was studied, much of our current understanding of the biology of temperate phages is based on the study of phage lambda.

Based on their morphology and type of nucleic acid carried, phages can be classified into broad taxonomic groups (Ackermann, 2006) (Figure 1.6). The heritable genetic material carried by phages can take 4 forms: single stranded DNA (ssDNA), double stranded DNA (dsDNA), single stranded RNA (ssRNA) and double stranded RNA (dsRNA). However, dsDNA phages are by far the most well studied and most abundant group. The order of tailed dsDNA phages, *Caudovirales*, constitutes 96% of all known phages (Fokine and Rossmann, 2014) and the vast majority of known phages associated with the genus *Salmonella* belong to one of the three families of the *Caudovirales*; the *Myoviridae, Siphoviridae* and *Podoviridae* (Kropinski et al., 2007). These families are differentiated based on further morphological features. The *Myoviridae* are characterised by a rigid contractile tail. The *Siphoviridae* have long, non-contractile tails and the *Podoviridae* have short, non-contractile tails. All three *Caudovirales* families contain both temperate and lytic phages. The lifecycle of a temperate phage (and a lysogenic cell), is shown in Figure 1.7. Temperate phage undergo a more complicated life cycle than that of lytic phage, which are only capable of infection leading to cell lysis. The process of lysogeny, particularly how phage enter into, and come out of, a latent state has been a topic of intense study for more than half a century, though much of the process remains a mystery or understood only in well-studied phage systems such as lambda.

The beginning of both of the lytic and lysogenic phage life cycle is infection. In tailed phages, this process is usually mediated by interaction of the phage tail with specific receptors on the bacterial surface, a process known as adsorption. Once the phage has adsorbed, infection continues with the translocation of the phage nucleic acids from the phage capsid, into the bacterial cell. At this point, the lytic and lysogenic life cycles diverge. Temperate phages are capable of both lytic and lysogenic infection, and the complex molecular process that defines which path is taken is known as the lytic-lysogenic decision (Ptashne, 2004). Lysogeny is thought to be favoured in poor growth conditions, because phages require cells that are metabolically active to grow lytically. The lytic-lysogenic decision in phage lambda is tuned to the metabolism of

the cell by the fate of single protein CII, one of the first proteins to be synthesised by the incoming phage (Ptashne, 2004).

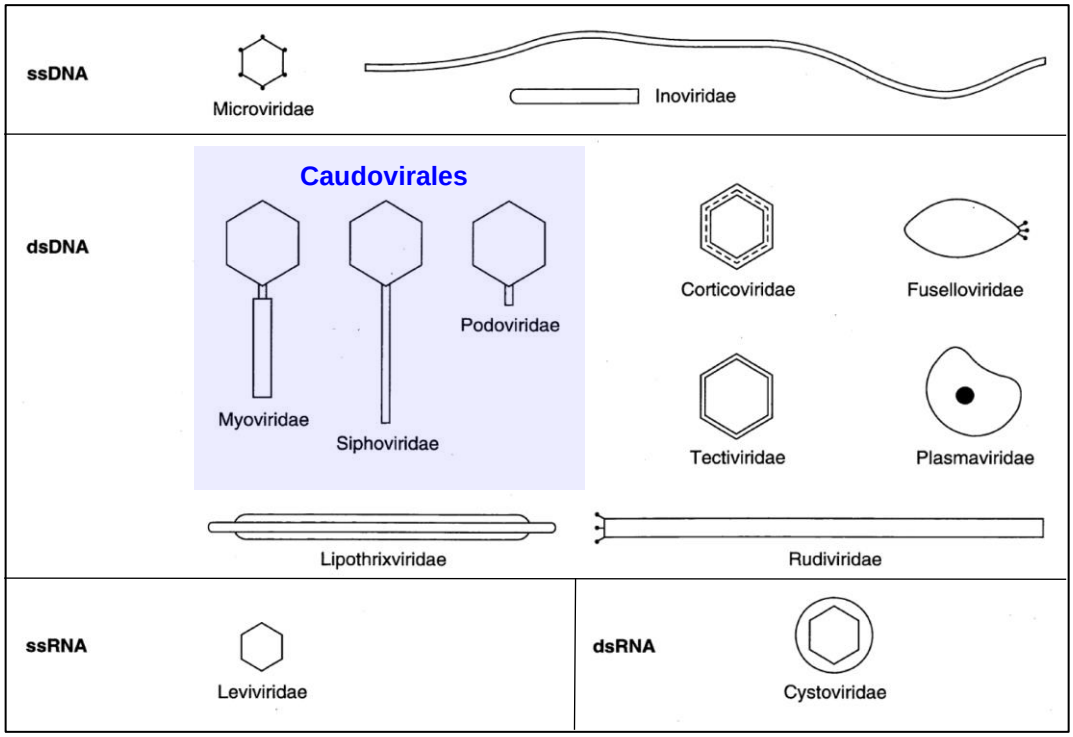



**Figure 1.6 Schematic representation of the major phage groups.** Reproduced and adapted from Ackermann, 2006**.** The structural characteristics of all major phage groups are shown. The *Caudovirales* order of tailed phage are highlighted in blue, as the three phage families in this order: *Myoviridae, Siphoviridae* and *Podoviridae* are the most frequently associated with the *Salmonella* genus, and therefore most relenvant to this thesis. The *Myoviridae* are characterised by a rigid contractile tail. The *Siphoviridae* have long, non-contractile tails and the *Podoviridae* have short, non-contractile tails.

The CII activator protein is rapidly degraded by host proteases in metabolically-active cells, which allows the lytic development of the phage. Inactive cells have low protease levels and therefore high levels of CII protein accumulates, driving the synthesis of the phage CI repressor and integrase proteins, to block the lytic replication of the phage and allow stable integration into the host chromosome.

During lysogeny, the CI repressor is constitutively expressed, repressing the production of all genes required for phage replication (Ptashne, 2004). The prophage replicates concomitantly with the host genome unless an induction signal is detected. The induction of lambdoid prophages is tuned to the SOS response of the host cell. The bacterial SOS response is an ancient gene expression cascade which is activated by DNA-damage, and is conserved across the Bacterial kingdom (Baharoglu and Mazel, 2014; Erill et al., 2007; Kreuzer, 2013; Michel, 2005). Upon interaction with single-stranded DNA, the bacterial RecA protein becomes 'activated' (Sassanfar and Roberts, 1990). Though the exact molecular basis of RecA protein 'activation' is not fully understood, in the case of *Salmonella* phage P22, presence of activated-RecA is known to stimulate the autoproteolysis of the P22 repressor protein leading to de-repression of phage replication genes (Campoy et al., 2006; Ptashne, 2004). Consequently, the phage lytic replication pathway proceeds. Specifically, the prophage excises from the host chromosome, replicates its DNA, synthesises and packages itself inside a protein capsid and tail, and induces cell lysis by the synthesis of proteins which hydrolyse the bacterial cell wall. The molecular processes and transcription events that control phage infection, and prophage induction are summarised in Figure 1.8.

Though the process of prophage induction has likely evolved to allow prophages to efficiently escape from doomed host cells (frequently likened in the literature to rats leaving a sinking ship), prophages are unable to prevent the natural accumulation of mutations in their genome during generations in lysogeny. Consequently, wild-type bacterial strains frequently harbour prophages that have been inactivated by mutations. These prophages are known as defective, or cryptic prophages. For example, a strain of *E. coli* K-12 contained 9 defective prophage elements (Wang et al., 2010) and in *Salmonella,* prophage-derived elements have been estimated to represent up to 20% of the genome (Casjens, 2003). Frequently, genes encoding bacterial virulence factors are encoded on active or defective prophage elements (Brüssow et al., 2004).

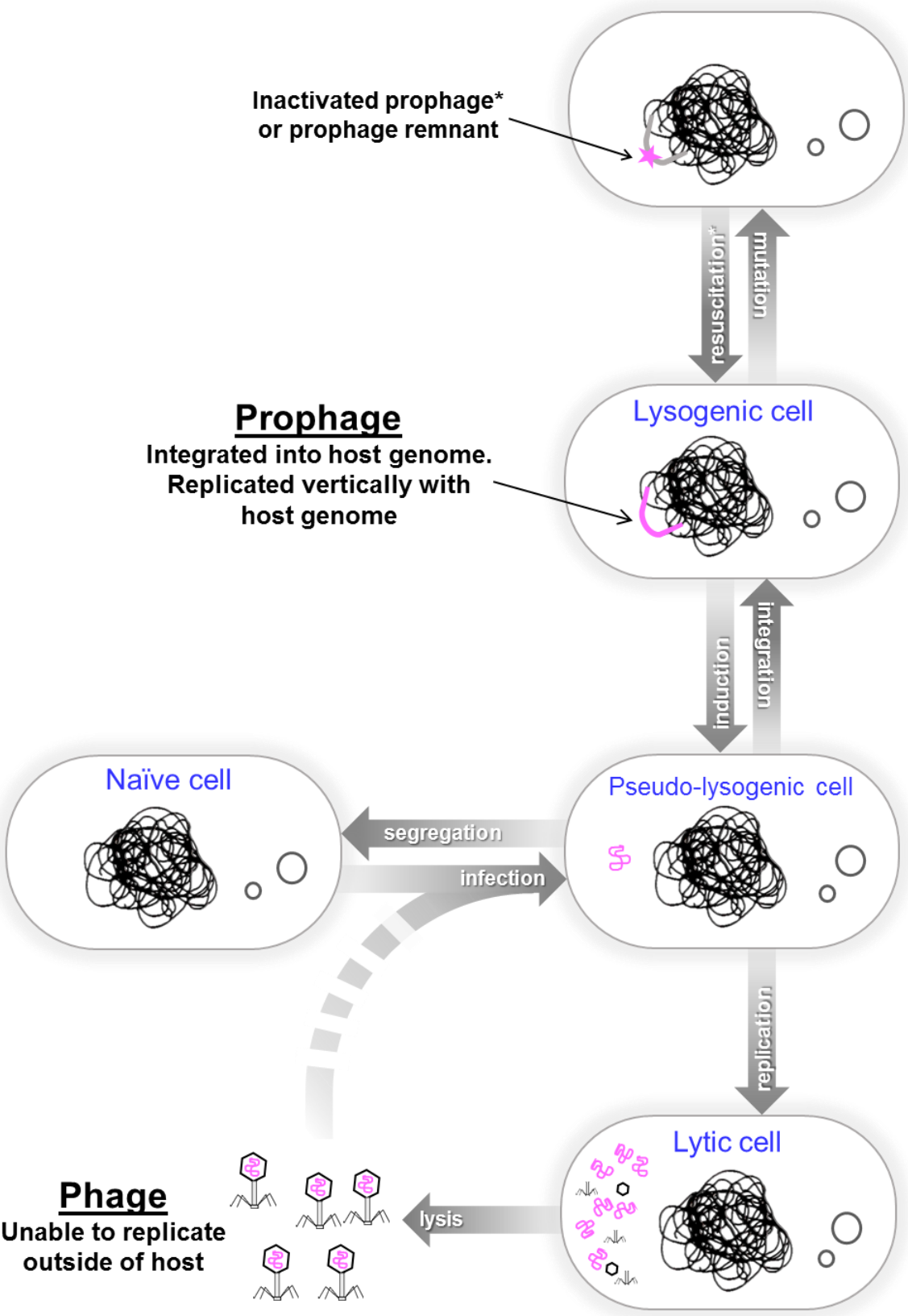



**Figure 1.7 The lifecycle of a temperate phage.** The processes which lead to each cell-state are represented by arrows. The topmost state represents a cell containing a remnant prophage. All *Salmonella* genomes contain remnant phage genes. Phage DNA is shown represented in pink

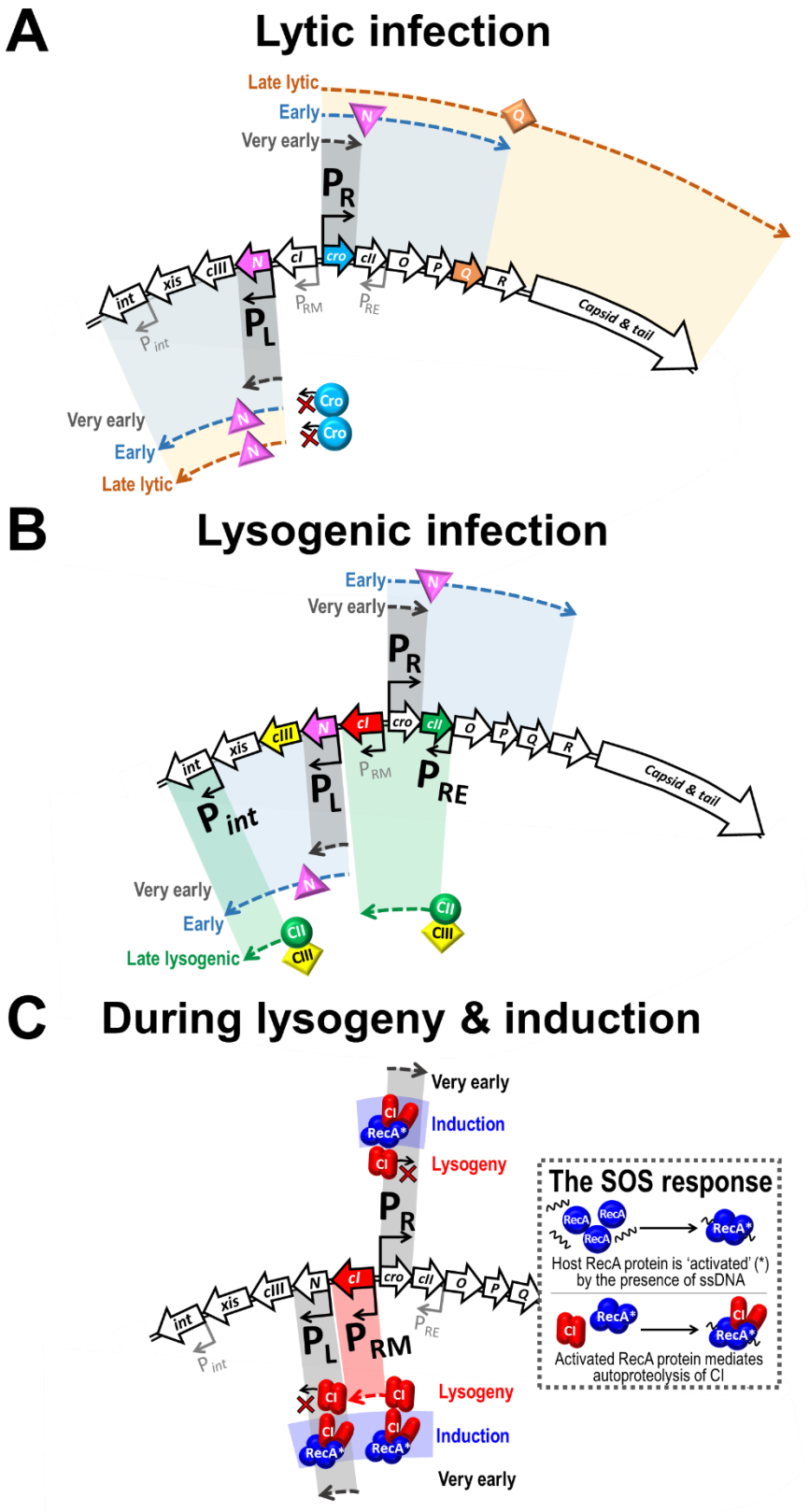



**Figure 1.8 Schematics showing the paradigm of the phage lambda molecular genetic switch.** Important molecular events governing each outcome is represented. Dashed lines indicate transcripts. Important promoters for each stage are shown in large bold text. A. During lytic infection N and Q protein antitermination allows late gene transcription. B. During lysogenic infection, CII protein activates CI repressor synthesis. C. During lysogeny, CI repressor protein is constitutively produced, preventing phage replication. Activation of the SOS response leads to the RecA-dependent inactivation of the repressor protein.

## 1.6 Temperate bacteriophages and the evolution of bacterial pathogens

Temperate prophages have long been recognised to carry important virulence factors. Before the actual virulence factors were characterised, changes in the virulence of bacterial strains upon lysogenisation by prophages was recognised, and was termed 'lysogenic conversion'. The earliest reported examples of this phenomenon are lysogenic conversion of *Corynebacterium diphtheriae* by a phage designated β, by Victor Freeman (Freeman, 1951), who showed that previously avirulent *C. diphtheriae* strains could cause disease in guinea pigs once lysogenised. Much later, the protein toxin *tox*, carried by the β-phage was characterised and shown to be the cause of lysogenic conversion (Holmes, 2000). The gene encoding the C1 neurotoxin, carried by the C1 prophage of *Clostridium botulinum* is a second early example of a prophage-encoded gene responsible for the virulence of a notorious pathogen (Barksdale and Arden, 1974). Subsequently, many examples of prophage-encoded virulence factors, capable of mediating lysogenic conversion, have been described (Brüssow et al., 2004) (summarised in Table 1.3). In *Salmonella*, important prophage-encoded virulence factors include the *sodC* gene of Gifsy-2 and the guanine nucleotide exchange factor carried by the ϕsopE prophage (Figueroa-Bossi & Bossi, 1999 ; Mirold *et al.,* 1999).

Recently it has been recognized that prophages can contribute to bacterial virulence and physiology by more complex mechanisms than simply encoding virulence factors. For example, prophages that use attachment sites corresponding to important genetic loci can act as regulatory switches by integration and excision, such as the Com system of *Listeria monocytogenes*, where controlled prophage excision and reintegration is required during infection to promote bacterial escape from macrophage phagosomes (Rabinovich et al., 2012). Spontaneous phage induction can directly contribute to virulence in a number of other bacterial pathogens (Nanda et al., 2015). Prophage induction causes an increase in biofilm formation by releasing extracellular DNA in pathogens including *Streptococcus pneumoniae* and *Shewanella oneidensis* (Carrolo et al., 2010; Gödeke et al., 2011). More recently spontaneous prophage induction was shown to modulate *Pseudomonas aeruginosa* biofilm formation by stimulating biogenesis of membrane vesicles (Turnbull et al., 2016). Finally, it has been reported that phage-mediated lysis can function as a crude method of toxin delivery, facilitating human gingival fibroblast invasion by *Aggregatibacter*

*actinomycetemcomitans* (Stevens et al., 2013), and enabling Stx toxin release in Shiga toxin-producing *E. coli* (Shimizu et al., 2009; Xu et al., 2012).

The contribution of prophages to the evolution of bacterial pathogens is undeniable, but the reason for this association remains a topic of debate (Abedon and LeJeune, 2007). One hypothesis that has widely prevailed in the literature is that the inclusion of virulence factors by prophages increases phage fitness indirectly, by increasing the fitness of the host cell. Whilst in the latent prophage state, the phage genome is replicated vertically with the bacterial chromosome, therefore, the fitness of the prophages is tied to that of the host cell. By increasing the success of the pathogen, the prophage increases the chance of its own survival.

**Table 1.3 Prophage-encoded virulence factors and associated mechanisms of action. Adapted from Brüssow et al. (2004).**

| Protein | Gene | Phage | Bacterial host |
|---|---|---|---|
| **Extracellular toxins** | | | |
| **Diphtheria toxin** | *tox* | β-Phage | *C. diphtheriae* |
| **Neurotoxin** | C1 | Phage C1 | *C. botulinum* |
| **Shiga toxins** | *stx1, stx2* | H-19B | *E. coli* |
| **Enterohaemolysin** | *hly2* | φFC3208 | *E. coli* |
| **Cytotoxin** | *ctx* | φCTX | *P. aeruginosa* |
| **Enterotoxin** | *see, sel* | NA | *Staph. aureus* |
| **Enterotoxin P** | *sep* | φN315 | *Staph. aureus* |
| **Enterotoxin A** | *entA* | φ13 | *Staph. aureus* |
| **Enterotoxin A** | *sea* | φMu50A | *Staph. aureus* |
| **Exfoliative toxin A** | *eta* | φETA | *Staph. aureus* |
| **Toxin type A** | *speA* | T12 | *Strep. pyogenes* |
| **Toxin type C** | *speC* | CS112 | *Strep. pyogenes* |
| **Cholera toxin** | *ctxAB* | CTXφ | *V. cholerae* |
| **Leukocidin** | *pvl* | fPVL | *Staph. aureus* |
| **Superantigens** | *speA1*, *speA3*, *speC*, *speI*, *speH*, *speM*, *speL*, *speK*, *ssa* | 8232.1 | *Strep. pyogenes* |
| **Cytolethal distending toxin** | *cdt* | Unnamed | *E. coli* |
| **Proteins altering antigenicity** | | | |
| **Membrane proteins** | Mu-like | Pnm1 | *N. meningitidis* |
| **Glucosylation** | *rfb* | $\varepsilon^{34}$ | *S. enterica* |
| **Glucosylation** | *gtr* | P22 | *S. enterica* |
| **O-antigen acetylase** | *oac* | Sf6 | *Sh. flexneri* |
| **Glucosyl transferase** | *gtrII* | SfII, SfV, SfX | *Sh. flexneri* |
| **Effector proteins involved in invasion** | | | |
| **Type III effector** | *sopE* | SopEΦ | *S. enterica* |
| **Type III effector** | *sseI (gtgB)* | Gifsy-2 | *S. enterica* |
| **Type III effector** | *sspH1* | Gifsy-3 | *S. enterica* |
| **Enzymes** | | | |
| **Superoxide dismutase** | *sodC* | Sp4, 10 | *E. coli* O157 |
| **Superoxide dismutase** | *sodC-I* | Gifsy-2 | *S. enterica* |
| **Superoxide dismutase** | *sodC-III* | Fels-1 | *S. enterica* |
| **Neuraminidase** | *nanH* | Fels-1 | *S. enterica* |
| **Hyaluronidase** | *hylP* | H4489A | *Strep. pyogenes* |
| **Leukocidin** | *pvl* | φPVL | *Staph. aureus* |
| **Staphylokinase** | *sak* | φ13 | *Staph. aureus* |
| **Phospholipase** | *sla* | 315.4 | *Strep. pyogenes* |
| **DNase/streptodornase** | *sdn, sda* | 315.6, 8232.5 | *Strep. pyogenes* |
| **Serum resistance** | | | |
| **OMP*b*** | *bor* | λ | *E. coli* |
| **OMP** | *eib* | λ-like | *E. coli* |
| **Adhesions for bacterial host attachment** | | | |
| **Vir** | *vir* | MAV1 | *M. arthritidis* |
| **Phage coat proteins** | *pblA*, *pblB* | SM1 | *Strep. mitis* |
| **Others** | | | |
| **Mitogenic factors** | *mf2, mf3, mf4* | 370.1, 370.3, 315.3 | *Strep. pyogenes* |
| **Mitogenic factor** | *toxA* | Unnamed | *P. multocida* |
| **Virulence** | *gtgE* | Gifsy-2 | *S. enterica* |
| **Antivirulence** | *grvA* | Gifsy-2, Fels-1 | *S. enterica* |

## 1.7 Aims and hypotheses

Compelling evidence exists to implicate prophages in the evolution of bacterial pathogens. Due to their intrinsic mobility, prophages can rapidly modulate the disease phenotype of a bacterial population. It was therefore hypothesised that the prophage repertoire of *S.* Typhimurium ST313 has contributed to the epidemic of invasive non-typhoidal *Salmonella* disease in the immunosuppressed human niche that is highly prevalent in sub-Saharan Africa.

The aims of this project were as follows:

- Functionally characterise the phage biology of each of the prophages found in the representative ST313 strain D23580
- Assess the biological contribution of each of the prophages of D23580, by the construction and phenotypic screening of knock-out mutants for each prophage
- Identify specific prophage genes which may modulate the phenotype of ST313 representative strain D23580
- Apply the findings from strain D23580 to strains representing the diversity of the ST313 population, to make broader conclusions about the contribution of prophages to the biology of this clinically-relevant sequence type.

# Chapter 2

## Materials and Methods

## 2.1 Reagents and chemicals

A list of chemicals and reagents used in this thesis, along with purchase suppliers can be found in Table 2.1.

## 2.2 Media and antibiotics

Recipes for all liquid and solid media used in this thesis are given in Table 2.1. All media were made up with water purified by de-ionisation and filtration to a standard of 18.2 MΩ/cm ($dH_2O$). All media were sterilised by autoclaving using a standard sterilisation cycle of 121°C at 15 psi (100 kPa) above atmospheric pressure for 15 minutes. InSPI2 media was not sterilized by autoclaving and instead by vacuum-driven filtration through 0.22 µm Steritop$^{TM}$ filter unit (Millipore). All sugar solutions were sterilised by syringe-driven filtration using Millex-GP 0.22 µm (Hydrophilic Polyethersulfone 33 mm) filter units and added to media post-autoclave, or before vacuum-driven filtration. Once prepared, all solid media was kept at 4°C until use.

Lysogeny broth (Lennox formulation) (LB) contained 10 g tryptone, 5 g yeast extract and 5 g NaCl per litre (pH 7.0). When required, LB was supplemented with the following antibiotics: ampicillin (Ap), 100 µg/ml; chloramphenicol (Cm), 25 µg/ml; kanamycin (Km), 50 µg/ml; gentamicin (Gm), 20 µg/ml; oxytetracycline (OxyTc), 10 µg/ml; tetracycline (Tc), 20 µg/ml. Antibiotic stock solutions were made in $dH_2O$, Ethanol, Methanol or NaOH as specified in Table 2.1. Stock concentration and final concentrations of all antibiotics used are given in Table 2.1. All antibiotic stock solutions were aliquoted and stored at -20°C and thawed on ice before use.

Agar plates were prepared by aliquoting approximately 25 ml of molten LB bacto-agar (1.5%) cooled to 50°C into sterile petri dishes. Lids were left ajar under sterile conditions to allow the plates to cool and dry for 30 minutes. Agar plates were stored in sealed plastic bags at 4°C. Where agar was supplemented with antibiotics, stock solutions were added to the agar at 50°C and mixed thoroughly before pouring the plates.

**Table 2.1 Chemicals and reagents used in this thesis, including the supplier and catalogue numbers.** Reagents are listed in order of use in media, so that media recipes can be referenced. The stock and final concentrations (including solvent) are given where appropriate.

| Use | Chemical/ Reagent | Source | Cat number | Stock conc. used | Final conc. used |
|---|---|---|---|---|---|
| **LB media** | Tryptone | Appleton Woods | MN649 | N/A | 10 g/L |
| | Bacto-Yeast Extract | Appleton Woods | DM832 | N/A | 5 g/L |
| | NaCl | Sigma-Aldrich | S3014 | N/A | 5 g/L |
| | Bacto-agar | Appleton Woods | 214010 | N/A | Various |
| | | | | | |
| **inSPI2 media** | MES (pH 5.8) | Sigma-Aldrich | M8250 | 400 mM | 80 mM |
| | Tricine | Sigma-Aldrich | T5817 | 400 mM | 4 mM |
| | FeCl3 | Sigma-Aldrich | 236489 | 100 mM | 100 µM |
| | K2SO4 | Sigma-Aldrich | P0772 | 376 mM | 376 µM |
| | K2HPO4 | Sigma-Aldrich | 60353 | 100 mM | 0.4 mM |
| | KH2PO4 | Sigma-Aldrich | P8416 | 100 mM | 0.4 mM |
| | Glucose | Sigma-Aldrich | G8270 | 20% | 0.4% |
| | NH4Cl | Sigma-Aldrich | A9434 | 1.5 M | 15 mM |
| | MgSO4 | Sigma-Aldrich | 83266 | 1 M | 1 mM |
| | CaCl2 | Sigma-Aldrich | C3306 | 1 M | 0.01 mM |
| | Na2MoO4 | Sigma-Aldrich | 737860 | 100 µM | 10 µM |
| | Na2SeO3 | Sigma-Aldrich | S5261 | 100 µM | 10 nM |
| | H3BO3 | Sigma-Aldrich | B6768 | 4 µM | 4 nM |
| | CoCl2 | Sigma-Aldrich | C8661 | 3 mM | 300 nM |
| | CuSO4 | Sigma-Aldrich | C1297 | 1 mM | 100 nM |
| | MnCl2 (tetrahydrate) | Sigma-Aldrich | 529680 | 8 mM | 800 nM |
| | ZnSO4 | Sigma-Aldrich | Z4750 | 10 µM | 1 nM |
| | | | | | |
| **Green plate media** | Tryptone | Appleton Woods | MN649 | N/A | 8 g/L |
| | Bacto-yeast extract | Appleton Woods | DM832 | N/A | 1 g/L |
| | NaCl | Sigma-Aldrich | S3014 | N/A | 5 g/L |
| | Bacto-agar | Appleton Woods | 214010 | N/A | 11 g/L |
| | Methyl Blue | Sigma-Aldrich | M6900 | N/A | 0.1 g/L |
| | Alizarin Yellow | Sigma-Aldrich | 206709 | N/A | 0.6 g/L |
| | Glucose | Sigma-Aldrich | G8270 | 20% | 0.7% |
| | | | | | |
| **M9 Salts** | Na2HPO4 | Sigma-Aldrich | S9390 | N/A | 64 g/L |
| | K2HPO4 | Sigma-Aldrich | 60353 | N/A | 15 g/L |
| | NaCl | Sigma-Aldrich | S3014 | N/A | 2.5 g/L |
| | NH4Cl | Sigma-Aldrich | A9434 | N/A | 5 g/L |
| | | | | | |
| **M9 Media** | M9 salts (5X) | N/A | N/A | 5X | 1X |
| | Glucose | Sigma-Aldrich | G8270 | 20% | 0.4% |
| | MgSO4 | Sigma-Aldrich | 83266 | 1 M | 2 mM |

| | | | | | |
|---|---|---|---|---|---|
| | CaCl2 | Sigma-Aldrich | C3306 | 1 M | 0.1 mM |
| | | | | | |
| **Media supplement** | Mitomycin C | Sigma-Aldrich | M4287 | 2 mg/ml (H2O) | 2 µg/ml |
| | Kanamycin mono-phosphate | Melford | K0126 | 50 mg/ml (H2O) | 50 µg/ml |
| | Tetracycline | Sigma-Aldrich | 87128 | 20 mg/ml (Methanol) | 20 µg/ml |
| | Ampicillin sodium salt | Melford | A0104 | 100 mg/ml (H2O) | 100 µg/ml |
| | Chloramphenicol | Sigma-Aldrich | C0378 | 20 mg/ml (Ethanol) | 20 µg/ml |
| | Nalidixic acid | Sigma-Aldrich | N8878 | 50 mg/ml (H2O) | 50 µg/ml |
| | Gentamicin sulphate | Melford | G0124 | 40 mg/ml (H2O) | 20 µg/ml |
| | L-Arabinose | Sigma-Aldrich | A3256 | 20% (H2O) | 0.20% |
| | *m-toluic* acid | Sigma-Aldrich | T36609 | 0.5 M (NaOH) | 1 mM |
| | Congo red | Sigma-Aldrich | C6767 | N/A | 40 µg/ml |
| | $H_2O_2$ (30%) | Merck Millipore | 107209 | N/A | 20 µl/ml |
| | | | | | |
| **Protocol reagent** | Chloroform | Sigma-Aldrich | C2432 | N/A | N/A |
| | Isopropanol | Sigma-Aldrich | 19516 | N/A | N/A |
| | Ethanol | Sigma-Aldrich | 469836 | N/A | N/A |

## 2.3 General bacterial methods

### 2.3.1 Bacterial strains and culture conditions

A list of strains and plasmids used in this study can be found in Table 2.2. Bacteria were routinely cultured in glass universal bottles in a volume of 5 ml LB broth inoculated from single colonies on LB plates for 16 hours at 37°C with shaking at 220 rpm (this procedure is referred to in this thesis as 'overnight culture'). Unless otherwise stated, for experiments, overnight cultures were used to sub-inoculate 1:1000 in 25 ml LB broth in 250 ml sterilize Erlenmeyer flasks and grown at 37°C with 220 rpm agitation. For growth in minimal media, 1 ml of overnight culture was harvested by centrifugation at 12,000 x g at room temperature for 1 minute. Cell pellets were washed three times in minimal media and sub-inoculated 1:500 in 25 ml of minimal media in 250 ml sterilized Erlenmeyer flasks and grown at 37°C with 220 rpm agitation. Optical density (OD) measurements were taken using a Jenway 67-series spectrophotometer at a wavelength of 600 nm. When colony forming unit (CFU) counts were required, 10-fold serial dilutions of cultures (up to a factor of $10^{-8}$) were carried out in sterile LB broth and triplicate 10 µl drops plated on LB agar plates (Miles and Misra method). CFU counts are given in units per ml (CFU/ml). Frozen stocks of all bacterial strains were kept at -80°C for the duration of the work. To prepare frozen stocks, overnight cultures were mixed with sterile (autoclaved) 50% glycerol to a final concentration of 20% in a 1.5 ml volume in a sterile 2 ml cryovial tube (i.e. 900 µl overnight and 600 µl 50% glycerol). Bacterial colonies were maintained on agar plates at 4°C for a maximum of 1 week.

### 2.3.2 Motility assays

0.3% LB agar was made-up and whilst molten, exactly 25 ml of agar were aliquoted into petri dishes to ensure minimal variation in agar depth between plates. The plates were left to dry at room temperature overnight (16 hours). 3 µl of overnight cultures were spotted on to the 0.3% LB agar plates. Care was taken not to pierce the agar surface with the pipette tip. Plates were left at room temperature for 30 minutes before incubation at 37°C for 5 hours. Motility was assessed by measurement of the diameter of migration. All strains were tested independently in triplicate.

### 2.3.3 RDAR morphotype assays

The methodology used to assay RDAR morphology was based a recent publication (Singletary et al., 2016). 2 µl of overnight bacterial cultures was spotted onto two sets of LB plates without NaCl and supplemented with 40 µg/ml Congo red (Table 2.1). Plates were incubated in parallel at 25°C and 37°C for 7 days without inversion. To avoid drying of the agar over the prolonged incubation time, plates were incubated

within a plastic bag. All experiments were conducted in triplicate. After 7 days colonial morphology was photographed using the ImageQuant LAS 4000 imager (GE Healthcare). The presence of a filamentous, patterned texture of the colony at 25°C indicated RDAR morphology, whereas smooth colony morphology was judged to be RDAR negative. RDAR morphology does not form at 37°C therefore this condition acted as a negative control.

**Table 2.2 All bacterial strains used in this thesis.**

| Bacterial strains | Description [a] | References/origin |
|---|---|---|
| ***S. Typhimurium*** | | |
| | **D23580 derivatives** | |
| JH3621 | D23580 WT | (Kingsley et al., 2009) |
| JH3810 | ΔBTP1::FRT | This study |
| JH3798 | ΔBTP5:FRT | This study |
| JH3874 | ΔGifsy-1 | (Owen et al., 2017) |
| JH3875 | ΔGifsy-2 | (Owen et al., 2017) |
| JH3877 | ΔBTP1 | (Owen et al., 2017) |
| JH3878 | ΔBTP5 | (Owen et al., 2017) |
| JH3881 | ΔBTP1 ΔBTP5 | (Owen et al., 2017) |
| JH3940 | ΔBTP1 ΔBTP5 ΔGifsy-1 | (Owen et al., 2017) |
| JH3942 | ΔBTP1 ΔBTP5 ΔGifsy-1 ΔGifsy-2 | (Owen et al., 2017) |
| JH3949 | ΔΦ (ΔBTP1 ΔBTP5 ΔGifsy-1 ΔGifsy-2 ΔST64B) | (Owen et al., 2017) |
| JH3983 | ΔGifsy-1, *aph-galE496*; $Km^R$ | (Owen et al., 2017) |
| JH3984 | ΔGifsy-2, *aph-galE496*; $Km^R$ | (Owen et al., 2017) |
| JH3986 | ΔST64B | (Owen et al., 2017) |
| JH3987 | PdinI-gfoA from 14028s (C→T substitution, position 2,790,162 in D23580 chromosome) | (Owen et al., 2017) |
| D23580 BTP5::$Km^R$ | BTP5::$Km^R$ (insertion of aph in BTP5); $Km^R$ | This study |
| D23580ΔBTP1 BTP5::$Km^R$ | ΔBTP1 BTP5::$Km^R$ (insertion of aph in BTP5); $Km^R$ | This study |
| D23580 Δ*recA* | D23580 with *recA* gene replaced with *nptII* from pKD13 and subsequently flipped out using the Flp recombinase-encoding pCP20 plasmid | (Owen et al., 2017) |
| D23580 *ST313-tdSTOP* | D23580 with two stop codons introduced into the beginning of the *ST313-td* gene sequence (total change of 4 nucleotides) | This study |
| | **14028s derivatives** | |
| MA5958 | 14028s, WT | L. Bossi |
| MA6684 | ΔGifsy-1 ΔGifsy-2 *bio-106*::Tn10 *galE496*; $Tc^R$ | L. Bossi (Campoy et al., 2006) |
| | **LT2 derivatives** | |
| LT2 | LT2, WT | (Zinder and Lederberg, 1952) |
| MA8508 | Gifsy-1[-] Gifsy-2[-] Fels-2[-] Fels-1[Δ(*int-attR*)*104::cat*] | L. Bossi (Lemire et al., 2011) |
| LT2 [BTP1] | LT2 derivative MA8508 lysogenised with phage BTP1 | (Owen et al., 2017) |
| LT2 [P22] | LT2 derivative MA8508 lysogenised with phage P22 (wildtype) | (Owen et al., 2017) |
| LT2 Δrec | LT2 derivative MA8508 with *recA* gene replaced with *nptII* from pKD13 and subsequently flipped out using Flp recombinase encoding pCP20 plasmid | (Owen et al., 2017) |
| | **4/74 derivative** | |
| | 4/74, WT | (Rankin and Taylor, 1966) |
| JH3864 | ΔGifsy-2 | N. Wenner |
| JH3865 | ΔGifsy-1 | N. Wenner |
| JH3866 | ΔGifsy-1 ΔGifsy-2 | N. Wenner |
| JH3880 | ΔΦSopE | N. Wenner |
| ***E. coli*** | | |
| S17-1 λpir | *pro thi hsdR recA* chromosome::RP4-2 Tc::Mu Km::Tn7/λpir; TpR, SmR | (Simon et al., 1983) |

[a] Relevant antibiotic resistances are indicated by $^R$: Ap, ampicillin; Gm, gentamicin; Km, kanamycin, OxyTc, oxytetracyline; Sm, streptomycin; Tc, tetracycline; Tp, trimethoprim. For clarity purposes, the natural resistances of D23580 and its derivative are not indicated.

### 2.3.4 Stationary phase catalase activity assay

Catalase activity was assayed based on previously described methods (Singletary et al., 2016). 20 µl of 20% aqueous $H_2O_2$ was added to 1 ml of bacterial overnight culture in 1 cm diameter glass test tubes and briefly vortexed to homogenise the suspension. Tubes were incubated at room temperature for 5 minutes and subsequently photographed and the height of the bubble column measured. Assays were conducted in triplicate.

### 2.3.5 Antibiotic resistance typing

Phenotypic antimicrobial susceptibility testing was carried out at Public Health England, Collindale for all UK-isolated ST313 strains. The antimicrobial susceptibility testing was done using breakpoint concentrations. Briefly, an agar dilution method involving Iso-sensitest agar or Muller-Hinton agar was used to determine if isolates were sensitive or resistant to a set concentration of individual antimicrobials (Table 2.3).

**Table 2.3 Antibiotic concentrations used for antimicrobial resistance typing of UK-isolated ST313 strains carried out by PHE (Chapter 4).**

| Antibiotic | Concentrations tested |
|---|---|
| Ampicillin | 8 mg/L |
| Chloramphenicol | 8 and 16 mg/L |
| Colistin | 2 mg/L |
| Sulphonamide | 256 mg/L |
| Gentamicin | 2 mg/L |
| Tobramycin | 8 mg/L |
| Amikacin | 8 mg/L |
| Streptomycin | 16 mg/L |
| Tetracycline | 8 mg/L |
| Trimethoprim | 2 mg/l |
| Nalidixic Acid | 16 mg/L |
| Ciprofloxacin | 0.064 and 0.5 mg/L |
| Ceftazidime | 1 and 2 mg/L |
| Cefotaxime | 0.5 and 1 mg/L |
| Cefoxitin | 8 mg/L |
| Cefpirome | 8 mg/L |
| Ertapenem | 0.064 and 0.5 mg/L |
| Temocillin | 128 mg/L. |

### 2.3.6 Microscopy and preparation of samples

An EVOS FL cell imaging system (Thermo Fisher) fitted with a GFP light cube (470/22 nm excitation, 525/50 nm emission) was used to visualise cells under transmitted light and fluorescent light with an EVOS 40X flurorite coverslip-corrected objective (Thermo Fisher, AMEP4699) and an Olympus 100X super-apochromat, coverslip-corrected oil objective (Thermo Fisher, AMEP4733). 2 µl of cell samples were pipetted onto glass microscopy slides and covered with a glass coverslip. Where cell fixing was required, cell samples were pelleted by centrifugation at 11,000 Xg for 2 minutes, and washed in PBS buffer (Melford, P3203). Cells were further pelleted and resuspended in 4% paraformaldehyde PBS for 20 minutes at room temperature. Cells were again pelleted and resuspended in PBS ready for microscopy. Where the 100X objective was used, a drop of immersion oil (Sigma-Aldrich, 56822) was applied to the objective before imaging.

## 2.4 Phage and prophage methods

### 2.4.1 Phages

All phages were propagated on the strain D23580 ΔΦ (JH3949; Table 2.2). This strain contains no prophages, so lysates produced from phage replication on this strain cannot be contaminated by induced native prophages. Phage lysates were clarified by centrifugation at 12,000 x g for 2 minutes and syringe filtration through a 0.22 µm filter. Lysates were routinely stored at 4°C for the duration of the project.

### 2.4.2 Phage enumeration and plaquing

Phage enumeration and plaque isolation were carried out using the double layer agar technique. Where enumeration of spontaneously induced phage in overnight cultures was required, 1 ml of overnight culture was passed through a 0.22 µm syringe filter. For enumeration, phage lysates or culture supernatants were serially diluted by orders of 10 in sterile LB broth. For overnight culture supernatant, dilution up to $10^{-7}$ was sufficient, whereas for phage lysates, higher dilutions were required (typically up to $10^{-10}$). 4 ml of 0.4% LB agar was seeded with 100 µl of and overnight culture of the required indicator strain (approximately $10^{8}$ CFU) and, once dry, phage dilutions were applied to the cell lawn in 10 µl drops. Once dried, plates were incubated overnight at 37°C. Phage concentrations were calculated as plaque forming units (PFU) per ml of lysate or culture supernatant (PFU/ml).

### 2.4.3 Isolation of phage lysogens

Exponential phase cultures of susceptible host strains and phage were mixed and plated using the double layer agar technique as described above. Turbid centres from

BTP1 or P22 containing plaques were identified using primers for the *pid* gene, pid_Fw and pid_Rev (Table 2.4). Turbid plaque centres that were positive for *pid* were further picked and purified on green indicator plates (Table 2.1) (Maloy, 1990). This medium contains glucose as a carbon source, and a pH indicator dye that turns dark green at sites where phage infection causes cell lysis and the concomitant release of organic acids, while the lysogens remain pale green. Candidate lysogens were picked onto fresh green plates and cross streaked against phage to confirm resistance of the lysogens to superinfection by the target phage.

### 2.4.4 Determination of prophage viability

Viable phages were detected in the supernatant of D23580 WT after 16 h culture. To stimulate chemically mediated prophage induction, mitomycin C (Sigma-Aldrich M4287) was added to exponentially growing cultures ($OD_{600}$ = 0.3) at a final concentration of 2 µg/ml. Phages were enumerated using the double layer agar technique as described above using specific indicator strains to detect each phage (Table 3.2). Prophage viability was defined as the ability to form infective phage particles with or without chemical induction.

To determine the viability of the BTP5 prophage, supernatant from a 16 h culture of the D23580 BTP5::*Km*$^R$ strain was passed through a 0.22 µm filter. 100 µl of filtered supernatant was added to 100 µl of D23580 ∆BTP5 and incubated for 1 h before selection on LB Km. Colonies were screened for insertion of the BTP5::*Km*$^R$ prophage using primers BTP5_int_fw, BTP5_int_rv, BTP5_attB_fw and BTP5_attB_rv to detect both the presence of internal BTP5 sequence and disruption of the chromosomal attachment site. The experiment was repeated using supernatant from one of the isolated BTP5::*Km*$^R$ lysogens to confirm the efficiency of lysogenization (number of colonies per ml of supernatant) was approximately the same, indicating that no resuscitation of the BTP5 prophage had occurred.

### 2.4.5 BTP1 burst size and spontaneously-induced population estimation

Burst size determination was achieved using the single step growth curve protocol previously described (Hyman and Abedon, 2009). Briefly, 50 µl of exponentially growing D23580 ΔBTP1 cells ($OD_{600}$ = 0.3) at a density of $10^8$ CFU/ml were mixed with 50 µl BTP1 phage at a density of $10^7$ PFU/ml to achieve a multiplicity of infection (MOI) of approximately 0.1. After 5 minutes adsorption time at 37°C, 1 µl of the cell-phage mixture was diluted 1000X by addition of 999 µl sterile LB, in order to minimise further adsorption. Free (non-adsorbed) phages were enumerated by plaque assay on a susceptible indicator strain as described above. The number of infected cells

was calculated as: [(number of cells x MOI) - number of unabsorbed phage]. Subsequently, the number of free phage was enumerated by sampling and filtration every 10 minutes. The difference between the phage titer pre- and post-burst divided by the number of infected cells was determined to be the burst size. The duration of time between infection and burst was determined as the latent period.

The percentage of the population undergoing spontaneous induction was estimated as [(100 / average number of cells in 16 h culture) / (average number of phage in 16 h culture / burst size)].

### 2.4.6 Adsorption assays

Overnight cultures were used to re-inoculate fresh 25 ml cultures, and cells were grown until $OD_{600}$ 2.0 (≈$2x10^8$ CFU/ml). Phage suspensions were normalised to a concentration of ≈$1x10^8$ PFU/ml. 1 ml of cells were mixed with 1 µl phage suspension to achieve an MOI of approximately 0.01. A low MOI was used to maximise the adsorption efficiency. The cell and phage suspension were incubated with agitation for 5 minutes to allow adsorption to occur. The suspension was centrifuged at 7,000 x g for 5 minutes to pellet cells and adsorbed phage. The titer of free-phage remaining in the supernatant was determined by titration and spotting as described in 2.3.2. A no-cell control (1 ml of sterile LB medium) was used as negative control. The level of reduction of free-phage PFU/ml in the absorbed supernatant vs. the negative control indicates the efficiency of adsorption of the phage. Absorption assays were conducted in triplicate.

### 2.4.7 BTP1 genomic DNA isolation and sequencing

A pure BTP1 phage stock was produced by picking and resuspending a BTP1 plaque in sterile LB and plating on indicator strain D23580 ΔΦ. This process was repeated three times to ensure purity and clonality. A high titer BTP1 lysate (~$10^{11}$) was produced by mixing 10 µl of BTP1 plaque suspension with 5 ml of exponential phase ($OD_{600}$= 0.3) D23580 ΔΦ. BTP1 virions were concentrated by polyethylene glycol (PEG) precipitation. Briefly, 1 ml of PEG-8000 20% NaCl 2.5 M was added to 4 ml BTP1 lysate and incubated on ice for 2 h. Subsequently the mixture was centrifuged at 15,000 x g for 1 hour at 4°C and liquid decanted. The pellet was resuspended in 50 µl molecular grade water and treated with DNase I for 1 hour at 37°C to remove any contaminating bacterial DNA. Subsequently the sample was heated to 90°C for 10 minutes to simultaneously deactivate DNase I and release BTP1 DNA from capsids. BTP1 DNA was purified using the Zymo Research Quick-DNA™ Universal Kit (cat# D4069) as per the “Biological Fluids & Cells” protocol and eluted in 15 µl

molecular grade water. Preparation of sequencing libraries and sequencing were performed by Tamsin Redgwell and Andy Millard at the Warwick Medical School, University of Warwick, UK. Sequencing libraries were prepared using the Illumina Nextera XT DNA sample kit per the manufacturer's protocol (Illumina, USA) with 1 ng of input DNA. Sequencing was performed on an Illumina MiSeq instrument using V2 chemistry (2 x 250 bp). Sequencing reads were quality trimmed using Sickle (Joshi and Fass J. N., 2011) and assembled using Spades (Bankevich et al., 2012). The assembly was manually reordered to begin at the small terminase. Functional annotation was transferred from the D23580 (FN424405) prophage BTP1 sequence using Prokka 1.11 (Seemann, 2014). Read data, assembly and functional annotation were deposited at the DDBJ/EMBL/GenBank database under the sample accession LT714109 (URL: http://www.ebi.ac.uk/ena/data/view/LT714109) and study accession PRJEB18919.

### 2.4.8 Transmission electron microscopy of phage (TEM)

To prepare samples for TEM, 10 µl of the culture supernatant from a D23580 overnight culture was pipetted onto a carbon/Pioloform-coated copper 200-mesh grid, and allowed to adhere for 10 minutes. To image phage infections, the supernatant of a 16 h D23580 culture containing virions was added to mid-exponential phase *S.* Typhimurium 4/74 ($OD_{600}$ = 0.3) cells in LB media, and incubated at 37°C for 20 minutes before TEM sample preparation. The grids were subsequently washed twice in distilled water for 2 minutes, negative stained with 2% uranyl acetate for 1 minute and examined using a FEI 120 kV Tecnai G2 Spirit BioTWIN transmission electron microscope at the EM Unit, University of Liverpool.

## 2.5 Molecular genetic techniques

### 2.5.1 Primers, plasmids and Sanger sequencing

All primers and plasmids used in this study are listed in Table 2.4 and 2.5 respectively. All primers were purchased from MWG Eurofins (Germany) and were HPSF purified. All plasmids used in this study were isolated from *E. coli* strains using the ISOLATE II Plasmid Mini Kit (Bioline) following the manufacturer's instructions except column elution was always carried out in molecular grade $H_2O$. All Sanger sequencing was carried out by GATC Biotech AG (Germany) using the LIGHTRUN or SUPREMERUN protocols.

**Table 2.4 All oligonucleotide primers used in this thesis**

| Oligo name | Sequence (5'→3') | Purposes |
|---|---|---|
| BTP5_int_fw | GAACGATGCGCCAATACCAC | Amplification of a 260bp product in BTP5 lysogenic strains |
| BTP5_int_rv | CATGCGTGCCGGTTAATGAG | |
| BTP5_attB_fw | CGGCTGGATTACAGCGTAAA | Amplification of a 167bp if attachment site empty. No product if occupied |
| BTP5_attB_rv | AGGCCACTCTTTAGAGTGGC | |
| NW_1 | ACGTGAATTCGCTGGTGCCGACGAACCATG (EcoRI) | Amplification of *attBGifsy-2* from MA6684, pNAW16 construction |
| NW_4 | AGCTGGATCCAACGTACAGTTCTAATGCG (BamHI) | |
| NW_26 | ACGTGAATTCCAATGATTAAAGAGTGGGC (EcoRI) | Amplification of *attBGifsy-1* from MA6684, pNAW15 construction |
| NW_29 | ACGTGGATCCAGGATATCTTTAATGGCGC (BamHI) | |
| NW_52 | ACGTGAATTCTAGCAGCGCAGACATCGCC (EcoRI) | Amplification of *attBBTP1* from 4/74, pNAW17 construction |
| NW_53 | ACGTGGATCCGAAGGCTGGCTTTATCTGGC (BamHI) | |
| NW_54 | ACGTGAATTCATCAGGCCATGCGCAAGCG (EcoRI) | Amplification of *attBBTP5* from 4/74, pNAW18 construction |
| NW_55 | ACGTGGATCCCTGGGCTTTGACGATCAC (BamHI) | |
| NW_61 | GCGGGTTATAACGGCGATG | *galE496* sequencing |
| NW_82 | GAAAAGAATTCTCAGTTCCGTCAAGTTC (BamHI) | |
| NW_86 | GTCGGGGCGCTCGCCAGGAAGCTACTCTCGCGTAAAAAAG | Insertion of *KmR* in *STM0777-galE intergenic region* |
| NW_87 | ACTGACATATGAATATCCTCCTTAG | |
| NW_161 | ACTTCGGAATAGGAACTTCAAGATCCCCTAGGGATAACAGGGTAATCACGCTGCCGCAAGCACTCAGGGCGC (I-SceI) | Construction of pKD4-I-SceI |
| NW_162 | GCGCCCTGAGTGCTTGCGGCAGCGTGATTACCCTGTTATCCCTAGGGGATCTTGAAGTTCCTATTCCGAAGT (I-SceI) | |
| NW_179 | ACGTGAATTCTCGATACGCTGCTCAATGTAG (EcoRI) | Amplification of *attBST64B* from LT2, pNAW42 construction |
| NW_180 | ACGTGGATCCAACATTCAGCAGGGTCGGTG (BamHI) | |
| NW_194 | GCCAGAGCTTTGCCAGCCG | Modification of PdinI-gfoA in D23580, *construction of JH3987* |
| NW_195 | GACGGATGGTTAAGTTGCAG | |
| NW_196 | GGTTATGTAGCTTCTCTATGC | PdinI-gfoA sequencing in JH3987 |
| NW_197 | GCAGGATGCTCTGTATACG | |
| NW_214 | ATTCTCATAATTCCCTCTACATTTAACTACTGTATATAAACACTAGGGATAACAGGGTAATC (I-SceI) | Modification of PdinI-gfoA in D23580, construction of JH3987 |
| NW_215 | GCAAAAAGTGCTATTCACCTCTTGAATATTCTTTCTAACAG GTATCATTTCGAACCCCAGAGTCC | |
| BTP5_kan_f | GCGGAATTTCTCCCGGTAGTCGATGCATAATATCGCCCTCAATGAATTTTGTGTAGGCTGGAGCTGCTTC | Insertion of KmR in BTP5 |
| BTP5_kan_r | TGGGACGTTACCTTAGCGGGTTTTATAGTGGCGCTTTGACGACACTAAATCATATGAATATCCTCCTTAG | |
| BTP5_kan_ext_f | CTGCTGGCGTCAGGTTCTCA | Sequencing of KmR insertion in BTP5 |
| BTP5_kan_ext_r | GCGTTTGTCCTGCCTTTCAA | |
| pid_Fw | GCCCAAATCGCCGCTTGC | Amplification of 380bp product in BTP1 lysogenic strains |
| pid_Rev | GATTATTGTTGCGTGCCC | |
| RecA_pKD13_Fw | GTACGAATTCACTATCCGGTTCAATACCAAGTTGCATGACAGGAGTAATAGTGTAGGCTGGAGCTGCTTC | Construction of LT2 ΔrecA |
| RecA_pKD13_Rev | GTTTTGCTGAATGGCGGCTTCGTTTTGCCCGCCCCACCATCACCTGATGACATATGAATATCCTCCTTAG | |
| 03761_gfp_2_L_f | ATTATTGTTCTGTAAACTTAAAGCAATGATCTGGTTGGTATAAATGTTCTAGATTTAAGAAGGAG | Amplification of *gfp+* gene from pZEP08 with flanking ends of intergenic BTP1 site and pKD4 kanamycin |
| gfp_kan_2_L_r | CTAAGGAGGATATTCATATGGCGACCGGCGCTCAGCTGGA | |
| gfp_kan_2_R_f | TCCAGCTGAGCGCCGGTCGCCATATGAATATCCTCCTTAG | |

| | | |
|---|---|---|
| 03761_gfp_2_R_r | GCCCATCTGGACTATCTCAACTAGTCGATTCATGACATGTGTCACGTGTAGGCTGGAGCTGCTTC | Amplification of kanamycin reistance construct from pKD4 with flanking ends of *gfp*+ gene from pZEP08 and intergenic BTP1 site |
| 03761_gfp_ext_f | CTAAGATTGCTATCACACTG | Amplification of region external to the *gfp-KmR* insertion in BTP1 to verify insertion |
| 03761_gfp_ext_r | TGATGCCGAGCACGCCCATCT | |
| late_gfp_2_L_f | TTCCTAATTCAATAGAGCAAATCCCCTCAATAAAGGGGGTAGAGCGTTCTAGATTTAAGAAGGAG | Alternative primer to 03761_gfp_2_L_f for replacement of BTP1 late genes with *gfp-KmR* |
| late_gfp_2_R_r | TAAGATAAATTTACATGGGTGCTTGTCACCCATGTTTTACAATATGTGTAGGCTGGAGCTGCTTC | Alternative primer to 03761_gfp_2_R_r for replacement of BTP1 late genes with *gfp-KmR* |
| aacA1_F (NW_88) | CTAAATACATTCAAATATGTATCCGGTCCAACCAGCGGCACCAG | Construction of in $pP_L$-Gm : PCR |
| aacA1_R (NW_89) | GTAAACTTGGTCTGACAGTTACCAATTAGGTGGCGGTACTTGGG | Amplification of aacC1 of pME4510 by overlap extension PCR cloning into $pP_L$ (pJV300) |
| Late_gfp_ext_f | AACACAGTATCCTGGATTTGTTCTA | Amplification of region external to late genes to confirm insertion of the *gfp-KmR* cassette |
| Late_gfp_ext_r | TGCTTAACAAGGGTAGGTGATGGCC | |
| waaL_del_fw | AGATTCATTAAAGAGACTCTGTCTCATCCCAAACCTATTGTGGAGAAAAGGTGTAGGCTGGAGCTGCTTC | Insertion of KmR in place of waaL |
| waaL_del_rv | CCTGATGATGGAAAACGCGCTGATACCGTAATAAGTATCAGCGCGTTTTTCATATGAATATCCTCCTTAG | |
| waaL_ext_fw | ACAAGCGTATTTGGAAAGATTCATTAAAG | Amplication of region external to waaL deletion |
| waaL_ext_rv | CTGGTTTGATAAGTGATTGAGTCCTGATG | |
| BTP1_ko_oligo_fw | AATGCGAAGGTCGTAGGTTCGACTCCTATTATCGGCACCAGTTAAATCAAGTGTAGGCTGGAGCTGCTTC | Insertion of KmR in BTP1 attachment site in strain 4/74 |
| BTP1_ko_oligo_rv | TAGATGGTGCCGATAATAGGAGTCGAACCTACGACCTTCGCATTACGAATCATATGAATATCCTCCTTAG | |
| BTP1_JP_fw | CGTGAACACACCCTTCTCAG | External to BTP1 attachment site. Used to confirm absense of BTP1 prophage |
| BTP1_JP_rv | ACTCATGGCGCATGGTAAAC | |
| snp_BTP1_left_fw | AGCGCTTTCTGAACGCCCGA | Diagnostic PCR of closest SNP to BTP1 between D23580 & 4/74 in the 40 Kb upstream of the BTP1 prophage |
| snp_BTP1_left_rv | CATCACTGTTATAACGGCTG | |
| snp_BTP1_right_fw | ATTGCTGCCGGTTTTGAATG | Diagnostic PCR of closest SNP to BTP1 between D23580 & 4/74 in the 40 Kb downstream of the BTP1 prophage |
| snp_BTP1_right_rv | AGAATTGGTATCAGCACTCG | |
| BTP5_KO_fw | GCAGGCGACTCATAATCGCTTGGTCGCTGGTTCAAGTCCAGCAGGGGCCAGTGTAGGCTGGAGCTGCTTC | Insertion of KmR in BTP5 attachment site in strain 4/74 |
| BTP5_KO_rv | AATTTGGTGGCCCCTGCTGGACTTGAACCAGCGACCAAGCGATTATGAGTCATATGAATATCCTCCTTAG | |
| BTP5_ext_fw | CGGCTGGATTACAGCGTAAA | External to BTP5 attachment site. Used |
| BTP5_ext_rv | AGGCCACTCTTTAGAGTGGC | |

| | | |
|---|---|---|
| snp_BTP5_left_fw | CCAGGTGATAACCTCGCGCC | to confirm absense of BTP1 prophage |
| snp_BTP5_left_rv | ATAATGCGGGTAAGTTCGCC | Diagnostic PCR of closest SNP to BTP1 between D23580 & 4/74 in the 40 Kb downstream of the BTP5 prophage |
| snp_BTP5_right_fw | GGCCGCCTGACGATTATGGC | Diagnostic PCR of closest SNP to BTP1 between D23580 & 4/74 in the 40 Kb downstream of the BTP5 prophage |
| snp_BTP5_right_rv | GTATCAGTGGCCTGTGATGT | |
| ST313td_L_bam_f | GGCCGGATCCATAATGCGCTCTTGG | Amplification of ST313td gene with stop codon mutation for construction of pEMG::ST313tdSTOP |
| ST313td_L_bam_r | AAATGGTTAATTAAATATAGTCCATATCACCCCGCC | |
| ST313td_R_eco_f | GGTGATATGGACTATATTTAATTAACCATTTGATTT | Amplification of ST313td gene with stop codon mutation for construction of pEMG::ST313tdSTOP |
| ST313td_R_eco_r | GGCCCTTAAGTCGAGAATTCGTGGA | |
| ST313_STOP_seq_f | ATACTCAAGAATTGCCATGC | for sequencing of stop mutations with template amplified from ST313td_ext_f+r |
| ST313_STOP??_f | GGTGATATGGACTATATTTA | Diagnostic PCR for ST313_stop mutation |
| ST313_STOP??_r | AGATTTGTTGTTTGTAGACG | |
| ST313td_pP$_L$GM_f | GTGAGCGGATAACAAGATACTGAGCACAGACGAAATCGTTAGCGCTT | For insertion of ST313td gene into expression plasmid pP$_L$-Gm using overlap extension PCR cloning |
| ST313td_pP$_L$GM_r | GCCTTTCGTTTTATTTGATGCCTCTAGAtgtgaagctactggaagagt | |
| STnc6030_pP$_L$_F (NW_295) | GTGAGCGGATAACAAGATACTGAGCACAGCAATATAGTCAACCTGAGAAC | For insertion of STnc6030 asRNA into expression plasmid pP$_L$-Gm using overlap extension PCR cloning |
| STnc6030_pP$_L$_R (NW_296) | GCCTTTCGTTTTATTTGATGCCTCTAGACTGCGTATCTGAAGGGGATTAAG | |
| STnc6030_T7 (NW_297) | TAATACGACTCACTATAGGGCTGCGTATCTGAAGGGGATTAAG | For synthesis of anti-STnc6030 riboprobe. Use with STnc60_pP$_L$GM_f |

**Table 2.5 All plasmids used in this thesis**

| Plasmid | Description [a] | Reference |
|---|---|---|
| pEMG | Suicide plasmid; $Km^R$ | (Martínez-García and de Lorenzo, 2011) |
| pKD13 | *nptII-cassette* template plasmid ; $Km^R$ | (Datsenko and Wanner, 2000) |
| pKD4 | *aph-cassette* template plasmid ; $Km^R$ | (Datsenko and Wanner, 2000) |
| pKD4-I-SceI | pKD4 derivative with an I-SceI site cloned upstream of *aph*; $Km^R$ | (Owen et al., 2017) |
| pKD46 | λ Red recombination plasmid, arabinose-inducible; $Ap^R$ | (Datsenko and Wanner, 2000) |
| pKD46 *bla*::Tn10 | pKD46 derivative; $OxyTc^R$ | (Passaris et al., 2014) |
| pCP20 | Plasmid carrying the Flp recombinase to remove kanamycin resistance from pKD13 derived resistance cassette insertions; $Ap^R$ | (Cherepanov and Wackernagel, 1995) |
| pCP20-TcR | Derivative of plasmid carrying the Flp recombinase to remove kanamycin resistance from pKD4 derived resistance cassette insertions; $Tc^R$ | (Kintz et al., 2015) |
| pNAW15 | Suicide plasmid pEMG::*attBGifsy-1*; $Km^R$ | (Owen et al., 2017) |
| pNAW16 | Suicide plasmid pEMG::*attBGifsy-2*; $Km^R$ | (Owen et al., 2017) |
| pNAW17 | Suicide plasmid pEMG::*attBBTP1*; $Km^R$ | (Owen et al., 2017) |
| pNAW19 | Suicide plasmid pEMG::$attB^{\Phi SopE}$; $Km^R$ | This study |
| pNAW18 | Suicide plasmid pEMG::*attBBTP5*; $Km^R$ | (Owen et al., 2017) |
| pNAW42 | Suicide plasmid pEMG::*attBST64B*; $Km^R$ | (Owen et al., 2017) |
| pSIM5-tet | λ Red recombination plasmid, temperature-inducible; $Tc^R$ | (Koskiniemi et al., 2011) |
| pSW-2 | Plasmid for *m-toluate-inducible* expression of the I-SceI enzyme; $Gm^R$ | (Martínez-García and de Lorenzo, 2011) |
| pZEP08 | Used as template to amplify the *gfp+* for construction of a prophage induction reporter construct | (Hautefort et al., 2003) |
| pME4510 | Used to amplify *aacC1* gene for modification of pJV-300; $Gm^R$ | (Rist et al., 1998) |
| pJV-300 ($pP_L$) | Modified to encode *aacC1* from pME4510 ($pP_L$-Gm); $Ap^R$ | (Sittka et al., 2007) |
| pEMG::ST313tdSTOP | Suicide plasmid pEMG::*ST313tdSTOP*; $Km^R$ | This study |
| $pP_L$-Gm | Derived from pJV-300 but *bla* gene replaced with *aacC1* gene from pME4510; $Gm^R$ | This study |
| $pP_L$-STnc6030 | Expression plasmid; STnc6030 sRNA under the control of the PLlacO-1 promoter; $Gm^R$ | This study |
| $pP_L$-ST313td | Expression plasmid; *ST313-td* gene under the control of the PLlacO-1 promoter; $Gm^R$ | This study |
| $pP_L$-ST313tdSTOP | Expression plasmid; *ST313-tdSTOP* gene under the control of the PLlacO-1 promoter; $Gm^R$ | This study |

[a] Relevant antibiotic resistances are indicated by $^R$: Ap, ampicillin; Gm, gentamicin; Km, kanamycin, OxyTc, oxytetracyline; Sm, streptomycin; Tc, tetracycline; Tp, trimethoprim.

### 2.5.2 Polymerase chain reaction (PCR) & agarose gel electrophoresis

For cloning and recombineering procedures, PCR reactions were performed with Q5 High-Fidelity DNA polymerase (New England Biolabs M0491S) according to the manufacturer's instructions. Each dNTP was used at a final concentration 0.2 mM, each primer was used at a concentration of 0.5 pmol/µL and 1 unit of enzyme was used per reaction of 50 µl. For PCR reactions where amplicons were not required for cloning, MyTaq$^{TM}$ Red Mix polymerase mastermix (Bioline) was used according to the manufacturer's instructions. In a typical 20 µL PCR reaction the mastermix was diluted 2-fold, primers were added to a final concentration of 1 pmol/µL and 1 unit of enzyme was used.

Initial denaturation time in the reaction was dependent on the template DNA used. For colony PCR or lysate PCR, initial denaturation proceeded for 10 minutes at 95°C and the time was reduced to 2 minutes when extracted and purified DNA was used as template. Initial denaturation was followed by 30 cycles of denaturation for 15 seconds at 95°C, annealing for 30 seconds at a temperature typically 4°C below the lowest melting temperature of the primer pair and extension for approximately 15-30 seconds per kilobase at 72°C. The cycling was followed by a final extension step at 72°C for 7 minutes. PCR products were held at 4°C when the PCR reaction was completed.

Agarose gels were made using molecular grade Agarose (Bioline). A 50X TAE buffer stock solution containing 242 g/L Tris-base, 57.1 ml/L Acetic acid and 100 ml/L 0.5M sodium EDTA was used. Gels were made up of TAE buffer (1X, made by diluting 20 ml of 50X stock in 980 ml of dd$H_2O$), and typically 1.5% agarose. Midori Green Advanced DNA Stain (NIPPON Genetics EUROPE GmbH) was added to agarose gels to a final concentration of 0.004%. Agarose gels were typically run at a constant voltage of 100 V for 30 minutes, or as required for sufficient separation of the DNA products. The molecular size of electrophoresed DNA products was estimated by comparison with Hyperladder I (Bioline). Separated DNA was visualised under ultra violet or blue light. Where it was necessary to extract DNA from agarose gels, the band of interest was excised from the gel using a clean scalpel. The gel slice was weighed to determine volume of buffer required for solubilisation. DNA products were purified from PCR reactions or agarose gel slices using the ISOLATE II PCR and Gel Kit (Bioline), following the manufacturer's instructions except column elution was always carried out in molecular grade $H_2O$.

### 2.5.3 Generalised transduction

Generalized transduction was performed with the highly efficient transducing phage P22 HT 105/1 *int*-201 (Schmieger, 1972). As D23580 is highly resistant to P22 infection (Kintz et al., 2015), high numbers of P22 transducing particles were necessary to obtain transductants in this background. Therefore, 200 µl of stationary phase culture of the recipient strains were mixed with 100 µl of the P22 lysates. After 1 hour of incubation at 37°C, transductants were selected on LB Km agar plates. Transductants were grown to stationary phase in LB containing 10 mM EGTA (Sigma-Aldrich, E3889) and Km. Cultures were streaked on indicator Green Plates and P22-free colonies were selected for further experiments (Maloy, 1990).

### 2.5.4 Preparation of electro-competent cells and electroporation

Bacterial overnight cultures were diluted 1:1000 and grown in 25 ml LB broth with antibiotics where required, until $OD_{600}$ 0.5. Cells were incubated on ice for 10 minutes, and then pelleted by centrifugation at 4°C and 7,000 x g for 10 minutes. Cells were washed three times in 25 ml ice-cold sterile $dH_2O$, and finally resuspended in 200 µl of ice-cold water. Cells were aliquoted into 40 µl volumes for electroporation reactions. Cells aliquots were mixed with target DNA (typically 200 ng of DNA was used per electroporation of linear dsDNA PCR amplicons, and 20 ng of DNA for electroporation of plasmids), and the suspension was incubated on ice for 5 minutes before transfer to a pre-chilled 2 mm single-use sterile electroporation cuvette (Geneflow, E6-0060). Electroporation was achieved by application of 2.5 kV, 25 µF, 200 Ω current to the cuvette using a 1652076 Gene Pulser Laboratory Benchtop Cell Electroporation Unit (Bio-Rad). 1 ml of room temperature LB broth was added to the cuvette to recover the electroporated cells, and the suspension was transferred to an Eppendorf tube and incubated at 37°C for 1 hour (unless the cells contained temperature sensitive plasmids, in which case a temperature of 30°C was routinely used). Afterwards, cells were pelleted by centrifugation at 12,000 x g for 2 minutes, and resuspended in 100 µl sterile LB. The resuspension was spread onto an LB plate containing the appropriate selectable antibiotic and incubated overnight at 37°C. Cells electroplated without DNA were also plated on the selectable antibiotic agar, as a negative control.

### 2.5.5 Preparation and transformation of chemically competent *E. coli* cells

Bacterial overnight cultures were diluted 1:1000 and grown in 25 ml LB broth with antibiotics where required until $OD_{600}$ 0.5. Cells were incubated on ice for 10 minutes, and then pelleted by centrifugation at 4°C and 7,000 x g for 10 minutes. The cell pellet was resuspended in 40 ml ice-cold 100 mM $CaCl_2$ and incubated on ice for 20 minutes. Cells were again pelleted by centrifugation at 4°C and 7,000 x g for 10

minutes and resuspended in 10 ml ice-cold 100 mM $CaCl_2$. Sterile glycerol was added to final concentration of 10%, and the cells were incubated on ice for a further 20 minutes. The chemically competent cells were aliquoted into 300 µl volumes and stored at -80°C until needed.

For chemical transformation, target DNA was mixed with a thawed aliquot of 50 µl chemically competent cells (20 ng of DNA for transformation of plasmids). The suspension was incubated on ice for 20 minutes. The cells were heat shocked for 90 seconds at 42°C in a pre-heated water bath, followed by incubation on ice for a further 2 minutes. 700 µl of LB broth was added and the cells were incubated at 37°C for 1 hour (unless the cells contained temperature sensitive plasmids, in which case a temperature of 30°C was routinely used). Cells were pelleted by centrifugation at 12,000 x g for 2 minutes, and resuspended in 100 µl sterile LB. The resuspension was spread onto an LB plate containing the appropriate selectable antibiotic and incubated overnight at 37°C. Cells transformed without DNA were also plated on the selectable antibiotic agar, as a negative control.

### 2.5.6 Plasmid construction by restriction/ligation

Construction of plasmids by restriction digest and ligation was achieved by using the EcoRI and BamHI restriction enzymes (Thermo Fisher). Plasmid inserts were amplified using primers designed with overhangs corresponding to the EcoRI and BamHI restriction sites. All amplicons were digested with restriction enzymes in parallel to the pEMG plasmid vector. DNA restriction was carried out as double digestion reactions according the manufacturer's instructions, using the universal Tango buffer 1X (Thermo Fisher). Cut DNA was purified using the ISOLATE II PCR and Gel Kit (Bioline). 50 ng of plasmid and 3-fold molar excess of the insert were ligated using T4 DNA ligase (New England Biolabs) at room temperature for 16 hours according to manufacturer's instructions. Ligated DNA was directly transformed into chemically competent *E. coli* S17-1 λ*pir* cells. Cells harbouring the pEMG variant plasmids were selected by plating on LB agar containing Km.

### 2.5.7 Plasmid construction by overlap-extension PCR cloning

Construction of plasmids by overlap-extension PCR cloning was based on methods previously described (Bryksin and Matsumura, 2010). The pJV-300 plasmid (Sittka et al., 2007) was initially modified to encode gentamicin resistance ($pP_L$-Gm) by overlap-extension PCR cloning. Chimeric primers containing pJV300 plasmid sequence at the 5′ ends and insert sequence at the 3′ ends were first used to PCR-amplify gentamicin resistance locus from the pME4510 plasmid. The amplified gene was used as a mega-

primer in a second PCR with the circular pJV300 template to amplify the entire plasmid but replacing the ampicillin resistance locus with the gentamicin resistance locus. For this reaction, 30 ng of the template plasmids were mixed with 150-300 ng of the insert and Q5 buffer, dNTPs, Q5 DNA polymerase (New England Biolabs) and water were added to a final volume of 50 µl. PCR reactions were carried out as follows:98°C, 30 sec; 25X(98°C, 10sec; 55°C, 30 sec, 72°C, 3 min); 72°C 5 min. The original plasmid template was then destroyed in restriction digests with DpnI in Cutsmart buffer (1X) (New England Biolabs) according to the manufacturer's instructions, and the overlap extension PCR products were used to transform *E. coli* DH5α cells. Cells harbouring the new $pP_L$-Gm (pJV300 gentamicin resistant derivative) plasmid were selected by plating on LB agar containing gentamicin.

Subsequently, overlap-extension PCR cloning was used to insert sequence of interest into the $pP_L$-Gm plasmid downstream of the $P_{LlacO-1}$ constitutive promoter. The same procedure was followed, except that the chimeric primers used to amplify the insert targeted a different region of the $pP_L$-Gm plasmid between the $P_{LlacO-1}$ promoter and the *rrnB* transcriptional terminator, as described in Table 2.7. To confirm correct conformation of plasmid after transformation, primers external to the insertion site were used to sequence the inserted fragment (Sanger sequencing).

## 2.6 Construction of mutant *Salmonella*

### 2.6.1 Construction of single gene mutants by λ Red recombineering

The pKD4 plasmid was used as the template for all constructs used in λ Red recombineering procedures. Primers (Table 2.4) were designed to amplify the kanamycin resistance gene flanked by the FRT recombinase sites from the pKD4 plasmid, with flanking ends corresponding to the chromosomal position to be replaced with the kanamycin resistance construct. Approximately 1 ng of pKD4 plasmid and 0.5 pmol/µL of each of the forward and reverse primers were used to amplify the kanamycin resistance constructs. The expected size of the construct was verified on an agarose gel and the construct cut and purified from the agarose gel ready for recombineering. Recombineering procedures were carried out using the λ Red recombination plasmids pSIM5-*tet* in the D23580 background and pKD46 in the 4/74/14028s/LT2 backgrounds (Datsenko and Wanner, 2000; Koskiniemi et al., 2011). Where the pSIM5-*tet* plasmid was used, an additional step of incubation at 42°C for 15 minutes directly after cells had reached $OD_{600}$ 0.4 in order to induce recombinase expression. Where pKD46 was used, the LB broth in which the competent cells were grown was supplemented with 10 mM L-arabinose to induce

recombinase expression. Electro-competent cells of *Salmonella* carrying pSIM5-*tet* or pKD46 were prepared and transformed with the indicated PCR fragment as described in section 2.5.4 (Datta et al., 2006; Yu et al., 2003) and recombinants were selected on LB agar plates supplemented with kanamycin. Correct insertion of the kanamycin cassette was confirmed by PCR using primers external to the deleted portion of DNA. To eliminate the chance of off-target effects from expression of the λ recombinase, the kanamycin marked mutation was transduced into a clean genetic background using P22 transduction as described in section 2.5.3.

In the case of the *galE496* allele a slightly modified procedure was used. The *galE496* allele (premature stop codon TAG on the 326$^{th}$ codon of *galE*) of MA6684 was transferred into D23580 derivative strains by a two-step procedure. An *aph* (*Km$^R$*) cassette was amplified from pKD4 with primers NW_86 and NW_87 and was inserted by λ Red recombination into the intergenic region *STM0777-galE496* of MA6684. Due to the defective O-antigen of the *galE* mutants, the resulting strain was grown in LB supplemented with 0.2% of glucose and 0.02% of galactose for the preparation of a P22 lysate (Butela and Lawrence, 2012). The *Km$^R$-galE496* module was subsequently transduced into D23580 derivative strains and transductants were screened by PCR with the primers NW_61 and NW_82 and checked by sequencing to confirm the presence of *galE496*.

To construct strain D23580 BTP5::*Km$^R$*, the *Km$^R$* cassette was first amplified from pKD4 with primers BTP5_kan_f and BTP5_kan_r, and then inserted by λ Red recombination into D23580 pSIM5-*tet*. The location of the *Km$^R$* cassette is in a non-coding position between genes *STMMW_32011* and *STMMW_32031*. Insertion of the cassette at the correct location was confirmed by PCR using primers BTP5_kan_ext_f and BTP5_kan_ext_r.

The pCP20-TcR plasmid was used to remove the kanamycin resistance locus where needed, and so reduce the effect of polar effects on the genes downstream of the deletion. The pCP20-TcR plasmid encodes the yeast FLP recombinase and is temperature sensitive for replication and induction of FLP expression. FLP recognises the FLP recombinase target sites (FRT sites) which flank the kanamycin resistance gene in the recombineering construct. FLP mediates site-specific recombination between the two FRT sites resulting in excision of the antibiotic resistance gene, leaving behind an 82-85 nt scar sequence (Datsenko and Wanner, 2000). Deletion mutants containing the kanamycin construct were made electrocompetent as previously described (2.5.4). Electroporation was used to transform cells with 500 ng

of pCP20-TcR plasmid. Cells were selected on tetracycline plates, and then passaged twice at 37°C to cure the temperature sensitive pCP20-TcR plasmid. Loss of the antibiotic resistance cassette and the pCP20 plasmid were screened for by checking for the loss of resistance to kanamycin and tetracycline.

The RecA$^-$ stains were obtained from Angela Makumi and Abram Aertsen, KU Leuven, Belgium. To construct a RecA$^-$ LT2 strain (LT2 derivative MA8508 was used, Table 2.2), the *Km*$^R$ cassette was first amplified from pKD13 with primers RecA_pKD13_Fw and RecA_pKD13_Rev, and then inserted by λ Red recombination into MA8508 pKD46. Kanamycin resistance was subsequently removed using the Flp recombinase encoded in the pCP20 plasmid. The same methodology was applied to D23580 to obtain the D23580 RecA$^-$ strain, but using the pKD46 *bla*::Tn*10* and pCP20-TcR plasmids instead as ampicillin resistance in D23580 (Kingsley et al., 2009) prohibits the use of the original pKD46 and pCP20 plasmids.

### 2.6.2 Construction of single nucleotide substitution mutants

The functional promoter of the antirepressor gene (*gfoA*) in Gifsy-1 of *S.* Typhimurium 14028s was inserted in place of the corresponding promoter of D23580 (Lemire et al., 2011). Therefore, a single nucleotide substitution (C→T position 2790162) was introduced into the promoter controlling *dinI*(*STMMW_26401*)-*gfoA (STMMW_26391)* transcription ($P_{dinI\text{-}gfoA}$) of D23580 by a protocol based on the principle described by Blank and Colleagues (Blank et al., 2011). Due to the multidrug resistance phenotype of D23580, a plasmid harbouring an I-SceI-*Km*$^R$ module was first constructed: the I-SceI restriction site was inserted upstream *aph* (*Km*$^R$) in pKD4, using the Site-Directed Mutagenesis method with primers NW_161 and NW_162 (Laible and Boonrod, 2009). The resulting plasmid (pKD4-I-SceI) was used as template to insert an I-SceI-*Km*$^R$ module into $P_{dinI\text{-}gfoA}$ by λ Red recombination, using primers NW_214 and NW_215. This module was transduced into a clean D23580 background and the resulting strain, D23580 I-SceI-*Km*$^R$-*dinI-gfoA,* was transformed with pSIM5-*tet* for a second λ Red recombination. The functional $P_{dinI\text{-}gfoA}$ of *S.* Typhimurium 14028s was PCR-amplified using the primers NW_194 and NW_195 and the resulting fragment (319 bp) was electroporated into D23580 I-SceI-*Km*$^R$-*dinI-gfoA* carrying pSIM5-*tet*. To select the recombinants, new electro-competent cells were directly prepared after the electroporation, and bacteria were transformed with pSW-2. After selection on LB agar containing Gm and *m*-toluate, colonies were screened by PCR with primers NW_196 and NW_197 and the resulting PCR fragments were sequenced to confirm the mutation. Two passages in LB were used to eliminate pSW-2.

### 2.6.3 Construction of mutant *Salmonella* by suicide plasmid-mediated genome editing

This genome editing protocol was adapted for *Salmonella* Typhimurium, according to the method previously described by Martínez-García and collegues (Martínez-García and de Lorenzo, 2011). To construct the different suicide plasmids, the DNA fragments (700-800 bp) flanking chromosome regions of interest were PCR amplified, fused by overlap extension PCR and inserted into the pEMG suicide plasmid, or by restriction ligation using restriction sites incorporated into the insert amplicon. In pEMG, the inserts are flanked by two I-SceI restriction sites. The pEMG derivative suicide plasmids were mobilized from *E. coli* S17-1 λ*pir* to *S.* Typhimurium by conjugation. *S.* Typhimurium transconjugant clones that had integrated the suicide plasmid by a homologous recombination were selected on solid minimal medium M9 supplemented with 0.2 % of glucose, 50 µg/ml of kanamycin and 20 µg/ml chormaphenicol. This medium is counter-selective for the S17-1λpir donor strain, which is auxotrophic for proline and thiamine. The merodiploid state of the resulting *Salmonella* transconjugants was resolved using the pSW-2 plasmid, encoding the I-SceI endonuclease. To improve the transformation yield of *Salmonella* with pSW-2, this plasmid was purified from *S.* Typhimurium 4/74 pSW-2 or *S.* Typhimurium D23580 pSW-2 and was then introduced into the strains of interest by electroporation. Transformants were selected on LB agar medium supplemented with 20 µg/ml gentamicin and 1 mM of *m*-toluate. In this condition I-SceI is expressed, provoking cleavages of the chromosome at the inserted I-SceI sites. The presence of double strand DNA breaks stimulates the RecABCD dependent DNA repair allowing homologous recombination events. The recombination can lead either to the restoration of the wild type status or to the generation of the mutant. Colonies were screened for kanamycin resistance and sensitive clones were tested by specific PCR. Due to the instability of its replication origin, pSW-2 was cured from the resulting strains by two passages in LB in the absence of gentamicin.

In the case of the D23580$^{ST313\text{-}tdSTOP}$ strain in which four nucleotides were replaced at the 5' end of the *ST313-td* coding sequence to introduce 2 x stop codons, a two-step procedure was used. First, the Site-Directed Mutagenesis method (Laible and Boonrod, 2009) with primers ST313td_L_bam_f / ST313td_L_bam_r and ST313td_R_eco_f / ST313td_R_eco_r where used to amplify an EcoRI, BamHI restriction site containing insert for the pEMG suicide plasmid. The plasmid (pEMG::ST313tdSTOP) was then used to introduce the stop codons into the D2580 genome using the suicide plasmid methodology described above.

### 2.6.4 Construction of GFP reporter strains for prophage induction

In order to act as a marker of prophage induction, a promoter-less *gfp* gene was designed to be inserted into a non-coding region of the late genes of prophage BTP1. In order to insert the *gfp*$^+$ gene into the chromosome, a construct was designed based on the *gfp*$^+$ gene of plasmid pZEP08 (Hautefort et al., 2003), and the FRT-site-flanked kanamycin gene of the pKD4 plasmid. Primers (Table 2.4) were designed such that the orientation of the kanamycin gene was opposite to that of the *gfp*$^+$, in order to minimise the likelihood of polar effects from the kanamycin resistance gene transcription. Primers 03761_gfp_2_L_f and gfp_kan_2_L_r were used to amplify the *gfp*$^+$ gene using pZEP08 DNA as template, and primers 03761_gfp_2_R_r and gfp_kan_2_R_f were used to amplify the kanamycin resistance loci. A third overlap extension PCR reaction was used to fuse both fragments together. The construct was extracted from agarose gel and electroporated into λ Red electrocompetent cells. The kanamycin gene was subsequently removed using the pCP20-TcR gene as has been described.

Subsequently, secondary mutations were designed which, rather than insert the construct within the prophage chromosome, the prophage lysis genes were deleted, and replaced with the *gfp*$^+$ construct. In the latter case, the same protocol was followed, except that the flanking ends of the construct corresponded to either side of the lysis genes targeted for deletion. In this case, primers late_gfp_2_L_f and gfp_kan_2_L_r were used to amplify the *gfp*$^+$ portion of the construct and primers late_gfp_2_R_r and gfp_kan_2_R_f were used to amplify the kanamycin resistance portion. The same approach was taken for prophage P22, except the flanking ends of the primers were modified to correspond to the P22 lysis genes (Table 2.4).

Detection of florescent cells which had become induced for prophage replication was achieved using microscopy of induced and uninduced cultures. An EVOS FL cell imaging system (Thermo Fisher) fitted with a GFP light cube was used to detect and photograph florescent cells (2.3.6).

## 2.7 Construction of expression plasmids

Plasmids to overexpress the *ST313-td* gene and *STnc6030* asRNA were constructed using the overlap extension PCR cloning method described earlier in 2.5.7. The primers used to amplify the *ST313-td* insert were ST313td_p$P_L$GM_f and ST313td_p$P_L$GM_r, and the primers used to amplify the *STnc6030* insert were STnc6030_p$P_L$_F (NW_295) and STnc6030_p$P_L$_R (NW_296) (Table 2.4).

## 2.8 Genetic deletion of prophages

### 2.8.1 Prophage deletion via unoccupied attachment site transduction

The method developed for initial deletion of prophages is described in Figure 2.1. A *Km*$^R$ resistance cassette was inserted into the prophage attachment site in strain 4/74 using lambda red method already described (2.6.1). *Salmonella* generalised transducing phage P22 HT105/1 int-201 was used to transduce a Km$^R$ resistance cassette into strain D23580 by selecting on LB kanamycin plates, thereby selecting for loss of the prophage. Transfer of co-located SNPs was screened by PCR. Transductant colonies with D23580 SNP genotypes in the flanking regions of the prophage (relative to 4/74) were selected for using colony PCR using annealing temperatures shown in Table 2.6. The kanamycin resistance cassette was recombined out of the chromosome to leave an 82bp scar using the pCP20-TcR plasmid as described (2.6.1). The mutation was confirmed by Sanger sequencing over the prophage attachment sites and over left and right SNP loci (primers indicated in Table 2.4). Resulting strains were denoted D23580 ΔBTP1::FRT and D23580 ΔBTP5::FRT to distinguish them from the deletion strain generated by the scarless approach (described subsequently). These strains were used for experiments described in Chapter 6.2.

**Table 2.6 Annealing temperatures used to assay SNPs proximal to prophage attachment sites.** Three temperatures in the discriminatory range for each primer pair are specified.

| | Temp 1 | Temp 2 | Temp 3 |
|---|---|---|---|
| BTP1_L | 59.2°C | 61.2°C | 63.2°C |
| BTP1_R | 67.0°C | 68.5°C | 69.6°C |
| BTP5_L | 68.5°C | 69.6°C | 70.1°C |
| BTP5_R | 63.2°C | 65.2°C | 68.5°C |

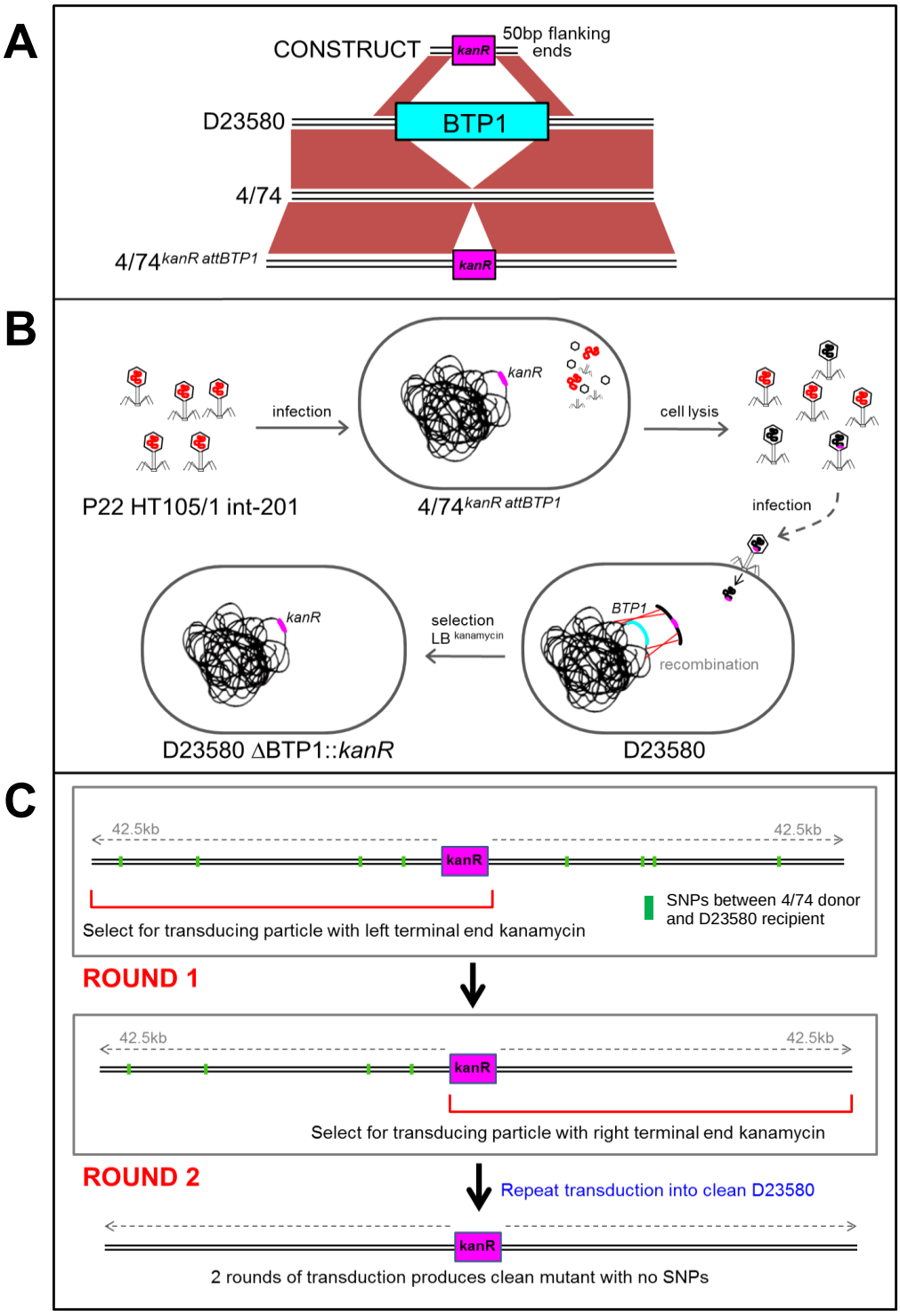



**Figure 2.1 Schematic representation of method developed to create whole-prophage deletions by transduction of unoccupied attachment sites.** A. Kanamycin cassette is first introduced into strain 4/74 in the BTP1 attachment site using lambda red recombineering (Datsenko & Wanner, 2000) B. The kanamycin cassette is then transduced from strain 4/74 into D23580 using phage P22 HT105/1 int-201. As the kanamycin cassette is located within an empty attachment site, transductants had lost BTP1. C. Any SNPs either side of BTP1 that may have been introduced by the transduction from 4/74 were removed by repeated transduction into a clean D23580 background and selecting for polar kanamycin transducing fragments using PCR as described in 2.8.1.

### 2.8.2 Scarless prophage curing

The prophages of S. Typhimurium D23580 and 4/74 were subsequently cured using the scarless genome editing technique described in 2.6.3. The phage attachment site (*attB*) of Gifsy-1$^{D23580}$, Gifsy-2$^{D23580}$, BTP1, BTP5 and ST64B$^{D23580}$ were PCR amplified from *S.* Typhimurium prophage-cured strains or naturally naive strains with the appropriate primers (Table 2.4) and cloned into the suicide plasmid pEMG, as summarized in Table 2.7. *E. coli* S17-1 λ*pir* was used as host strain for the resulting suicide plasmids pNAW15 (*attB*$^{Gifsy-1}$), pNAW16 (*attB*$^{Gifsy-2}$), pNAW17 (*attB*$^{BTP1}$), pNAW18 (*attB*$^{BTP5}$), pNAW42 (*attB*$^{ST64B}$) and pNAW19 (*attB*$^{ΦsopE}$, for 4/74 only). The suicide plasmids were mobilized from S17-1 λ*pir* into the recipient *S*. Typhimurium strains by conjugation and deletion of the prophages using the suicide plasmids was carried out as described in 2.6.3. The scarless phage deletion technique was used to generate strains D23580 ΔBTP1, D23580 ΔGifsy-2, D23580 ΔST64B, D23580 ΔGifsy-1, D23580 ΔBTP5, D23580 ΔBTP1ΔBTP5 and D23580 ΔΦ (Table 2.2). These strains were used in experiments described in Chapter 3.2, 5.3, 6.3 and 6.4.

**Table 2.7 Details of the construction of plasmids used in this thesis**

| Cloning technique | Template plasmid | Insert | Insert Primers Forward / Reverse | Template DNA | Restriction sites used | Resulting plasmid construct |
|---|---|---|---|---|---|---|
| Restriction/ligation | pEMG | *attB*$^{Gifsy-2}$ (BTP2) | NW_1 / NW_4 | MA6684 | EcoRI / BamHI | **pNAW16** (pEMG::*attB*Gifsy-2) |
| Restriction/ligation | pEMG | *attB*$^{Gifsy-1}$ (BTP4) | NW_26 / NW_29 | MA6684 | EcoRI / BamHI | **pNAW15** (pEMG::*attB*Gifsy-1) |
| Restriction/ligation | pEMG | *attB*$^{BTP1}$ | NW_52 / NW_53 | 4/74 | EcoRI / BamHI | **pNAW17** (pEMG::*attB*BTP1) |
| Restriction/ligation | pEMG | *attB*$^{BTP5}$ | NW_54 / NW_55 | 4/74 | EcoRI / BamHI | **pNAW18** (pEMG::*attB*BTP5) |
| Restriction/ligation | pEMG | *attB*$^{ST64B}$ (BTP3) | NW_179 / NW_180 | LT2 | EcoRI / BamHI | **pNAW42** (pEMG::*attB*ST64B) |
| Restriction/ligation | pEMG | *attB*$^{\Phi SopE}$ | NW_64 / NW_65 | D23580 | EcoRI / BamHI | pNAW19 (pEMG::*attB*$^{\Phi SopE}$) |
| Restriction/ligation | pEMG | ST313-tdSTOP | ST313td_L_bam_f / ST313td_L_bam_r | D23580 | EcoRI / | pEMG::ST313tdSTOP |
| Restriction/ligation | pEMG | ST313-tdSTOP | ST313td_R_eco_f / ST313td_R_eco_r | D23580 | BamHI | pEMG::ST313tdSTOP |
| PCR cloning | pP$_L$ (pJV-300) | *aacC1* (Gm$^R$) | NW_88 / NW_89 | pME4510 | N/A | pP$_L$-Gm |
| PCR cloning | pP$_L$-Gm | *STnc6030* | NW_295 | D23580 | N/A | pP$_L$-STnc6030 |
| PCR cloning | pP$_L$-Gm | *STnc6030* | NW_296 | D23580 | N/A | pP$_L$-STnc6030 |
| PCR cloning | pP$_L$-Gm | *ST313-td* | ST313td_pP$_L$GM_f | D23580 | N/A | pP$_L$-ST313td |
| PCR cloning | pP$_L$-Gm | *ST313-td* | ST313td_pP$_L$GM_r | D23580 | N/A | pP$_L$-ST313td |
| PCR Cloning | pP$_L$-Gm | *ST313-tdSTOP* | ST313td_pP$_L$GM_f | D23580 *ST313-td*STOP | N/A | pP$_L$-ST313tdSTOP |
| PCR Cloning | pP$_L$-Gm | *ST313-tdSTOP* | ST313td_pP$_L$GM_r | D23580 *ST313-td*STOP | N/A | pP$_L$-ST313tdSTOP |

**Table 2.8 RNA-seq conditions as described in Kröger et al., 2013.**

| Sample description | Abbreviation | Growth conditions |
|---|---|---|
| Early exponential phase | EEP | Growth in Lennox broth to OD600 0.1 |
| Mid exponential phase | MEP | Growth in Lennox broth to OD600 0.3 |
| Late exponential phase | LEP | Growth in Lennox broth to OD600 1.0 |
| Early stationary phase | ESP | Growth in Lennox broth to OD600 2.0 |
| Late stationary phase | LSP | Growth in Lennox broth to OD600 2.0 + 6 h |
| Low temperature | 25°C | Growth in Lennox broth to OD600 0.3 at 25°C (It's growth at low temperature (ON culture 37dgr, 200 RPM, Lennox, diluted 1:1000 and grown at 25 dgr until OD600 = 0.3) |
| Osmotic shock | NaCl_S | Growth in Lennox broth to OD600 0.3; then addition of NaCl to a final concentration of 0.3 M for 10 min |
| Addition of bile | Bile_S | Growth in Lennox broth to OD600 0.3; then addition of bile to a final concentration of 3% for 10 min |
| Iron limitation | LwFe_S | Growth in Lennox broth to OD600 0.3; then addition of 2,2'-dipyridyl to a final concentration of 0.2 mM for 10 min |
| Anaerobic shock | No_O2_S | Growth in Lennox broth to OD600 0.3 (50 ml), then filled into 50 ml closed centrifuge tube and incubated without agitation for 30 min at 37°C (Falcon tube) |
| Anaerobic growth | No_O2 | Static growth in Lennox broth to OD600 0.3 in a completely filled and closed 50 ml centrifuge tube (Falcon tube) |
| Aerobic shock | O2_S | Static growth in Lennox broth to OD600 0.3 in a completely filled and closed 50 ml centrifuge tube (Falcon tube); then 15 min aerobic growth (baffled flask, 250 rpm) |
| SPI2 non-inducing conditions | noSPI2 | Growth in PCN medium to OD600 0.3 (pH 7.4, 25 mM Pi) |
| SPI2 inducing conditions | inSPI2 | Growth in PCN medium to OD600 0.3 (pH 5.8, 0.4 mM Pi) |
| Oxidative stress | H2O2_S | PCN to OD600 0.3, then addition of H2O2 to final concentration of 1 mM H2O2 for 12 min |
| Nitric oxide | NOs | Growth in PCN medium to OD600 0.3 (pH 5.8, 0.4 mM Pi); then addition of 250 µM Spermine NONOate for 20 min |
| RNA from macrophages | MAC | RNA isolated from macrophages 8 h post infection |

## 2.9 RNA-associated methods

### 2.9.1 Growth conditions for RNA extraction

Transcriptomic data for prophage BTP1 (Chapter 5.2) was taken from the complete D23580 transcriptome dataset including identification of transcription start sites (TSS) using differential RNA-seq (Canals et al., 2018; Hammarlöf et al., 2017), where experiment design was based on that used in Kröger et al. (Kröger et al., 2012, 2013).. A description of the experimental conditions associated with the data presented in this thesis is given in Table 2.8. For analysis of prophage-regulated genes (Chapter 5.3), only the conditions ESP, inSPI2 and anaerobic growth were used.

### 2.9.2 RNA extraction

The protocol is essentially as described (Kröger et al., 2013). Four or 5 $OD_{600}$ units were removed from bacterial cultures, and cellular transcription was stopped using 0.4X culture volume of a 5% phenol (pH 4.3) 95% ethanol "stop" solution (Sigma P4557 and E7023, respectively). Cells were stabilised on ice in stop solution for at least 30 minutes before cells were harvested at 7,000 x g for 10 minutes at 4°C. At this point pellets were either stored at -80°C, or RNA was immediately extracted. To isolate RNA, pellets were resuspended in 1 ml of TRIzol™ Reagent (Invitrogen). 400 µL of chloroform was added and the samples were immediately and thoroughly mixed by inversion. Samples were moved to a Phase-lock tube (5 Prime) and the aqueous and organic phases were separated by centrifugation at 13,000 rpm for 15 minutes at room temperature in a table top centrifuge. The aqueous phase was moved to a new 1.5 ml tube and the RNA was precipitated using isopropanol for 30 minutes at room temperature followed by centrifugation at 21,000 x g for 30 minutes at room temperature. The RNA pellet was rinsed with 70% ethanol followed by centrifugation at 21,000 x g rpm for 10 minutes at room temperature. The RNA pellet was air-dried for 15 minutes and resuspended in DEPC-treated water at 65°C with shaking at 900 rpm on a Thriller thermoshaker (Peqlab) for 5 minutes with occasional vortexing. RNA was kept on ice whenever possible and RNA was stored at -80°C. RNA concentration was measured using a nanodrop ND-1000 Spectrophotometer and RNA quality was inspected visually using a 2100 Bioanalyser (Agilent).

### 2.9.3 cDNA library preparation and RNA-seq

Strand-specific cDNA library preparation and high throughput cDNA sequencing (RNA-seq) of wild-type D23580 and isogenic prophage mutants was performed on DNase I digested total RNA by Vertis Biotechnologie AG (Freising, Germany). Strains and conditions in which they were grown before RNA extraction are detailed in Table 5.1. RNA was not depleted for ribosomal RNA. RNA was fragmented by sonication

and poly(A)-tails were added to each fragment by poly(A) polymerase. The 5' end of each fragment was de-phosphorylated using tobacco acid pyrophosphatase (TAP). An RNA adaptor, containing a 6-10 nucleotide bar-code was ligated to the 5' end of each fragment to allow for identification of each RNA fragment after sequencing. First strand cDNA synthesis was performed using oligo(dT) priming and Moloney murine leukaemia virus reverse transcriptase (M-MLV RT). The resulting cDNA was amplified by PCR to approximately 20-30 ng/µL using a high fidelity DNA polymerase. cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and analysed by capillary electrophoresis.

cDNA prepared by Vertis Biotechnologie AG was sequenced on an Illumina Nextseq 500 platform using 75 bp read lenghts. The strains and conditions under which they were grown before RNA was extracted are detailed in Table 5.1.

### 2.9.4 Generating Digoxigenin-labelled riboprobes

Digoxigenin (Dig)-labelled riboprobes were generated by *in vitro* transcription using T7 RNA polymerase using a Dig Northern Starter kit (Roche). A linear DNA template incorporating the antisense sequence to the transcript of interest and a T7 promoter sequence was first generated by PCR using primers shown in Table 2.4 and a high fidelity enzyme as previously described. The DNA template was purified by gel extraction according to the manufacturer's instructions and 100-200 ng of DNA template was used in each *in vitro* transcription reaction. Linear DNA template was combined with 1X labelling mix, which contains unlabelled nucleotides and DIG-11-UTP, 1X transcription buffer and 40 units of T7 polymerase in a final volume of 20 µL. Labelling transcription mixes were incubated at 42°C for 1 hour. 20 U of DNase I was added to remove template DNA and the reactions were incubated at 37°C for 15 minutes. The reaction was stopped by addition of 400 mM EDTA (pH 8.0) and labelled riboprobes were stored at -20°C.

### 2.9.5 Northern Blotting

Total RNA after extraction as described in 2.9.2 was separated based on its size by electrophoresis through an 8.3 M Urea, 1X TBE (Tris Borate EDTA) 7% polyacrylamide gel and a denaturing 20 mM guanidine thiocyanate 1.5% agarose gel. Generally 1-10 µg of RNA was mixed with an equal volume of 2X Urea-Blue denaturing buffer (0.025% xylene cyanol, 0.025% bromophenol blue and 50% urea) and samples were heat-denatured at 90°C for 5 minutes and chilled on ice before loading. 4 µl of Low Range ssRNA ladder (NEB) or 5 µl RNA molecular weight marker I DIG-labeled (Roche) were treated in the same way as the RNA samples in order to

allow approximation of detected transcript length. Samples were run in 1X TBE running buffer at a constant voltage of 120V for denaturing polyacrylamide gels or 80 V (at 4°C) for denaturing agarose gels.

Separated RNA was transferred from polyacrylamide gels to positively charged nylon membranes (Roche, cat. 11 209 272 001) using the Trans-Blot SD Semi-Dry Electrophoretic Transfer Cell (BioRad) at a constant amplitude of 125 mA for 30 minutes at 4°C. For denaturing agarose gels, separated RNA was transferred to positively charged nylon membranes using overnight capillary transfer in 20X saline-sodium citrate (SSC) buffer as described in the DIG Application Manual (Roche).

RNA was UV-crosslinked to the membranes in a CL-1000 UV-crosslinker (UVP) set to 3600 (360,000 $\mu$J/cm$^2$). The membrane was equilibrated in hybridisation buffer for 1 hour at 68°C in pre-warmed DIG Easy Hyb solution (Roche) in a rotating hybridisation oven. 5 µl (approximately 1.25 µg) of riboprobe was heat denatured in 5 ml of DIG Easy Hyb solution at 68°C for 30 minutes and added to the membrane for hybridisation overnight at 68°C in the rotating hybridisation oven. The membrane was washed twice for a total of 10 minutes in low stringency wash buffer 1 (2X SSC buffer, 0.1% SDS) at room temperature with rocking on a see-saw rocker (Stuart) at room temperature, followed by 2 washes for a total of 30 minutes in high stringency wash buffer 2 (0.1X SSC buffer, 0.1% SDS) with rocking at 68°C. Non-specific sites on the membrane were blocked using 1X blocking buffer (casein-based blocking buffer supplied by Roche and diluted 10-fold in maleic acid buffer (0.1 M maleic acid, 0.15 M NaCl adjusted to pH 7.5 using NaOH pellets)) for 30 minutes at room temperature with rocking. Alkaline phosphatase conjugated polyclonal anti-digoxigenin Fab-fragment was diluted 1:10,000 in 1X blocking buffer and immunological detection of the membrane proceeded for 30 minutes at room temperature with rocking. The membrane was then washed twice for a total of 30 minutes in wash buffer (maleic acid buffer, 0.3% Tween-20). The membrane was incubated for 5 minutes in detection buffer (0.1M Tris-HCl, 0.1M NaCl, pH 9.5) and CDP-*star*$^{TM}$ (Tropix) was used as the chemiluminescent substrate. Enzymatic de-phosphorylation of CDP-star by alkaline phosphatase results in light emission which was visualised using an ImageQuant Las 4000 Imager.

## 2.10 Methods pertaining to infection models

### 2.10.1 Murine macrophage infection assays

RAW 264.7 (ATCC) murine macrophage cells were maintained in Dulbellcco's Minimal Essential Medium (DMEM) supplemented with 5% fetal bovine serum & L-

glutamine (2 mM final concentration) without antibiotics, incubated at 37°C in 5% $CO_2$. For each infection assay, a total of 106 RAW 264.7 cells were seeded in a six well plate and infected with overnight grown, complement-opsonized *Salmonella* strains at a multiplicity of infection of 10:1 (bacteria:macrophage). The plates were centrifuged at 500 x g for 5 minutes to maximise the uptake and designated as t = 0. Thirty minutes after t = 0, the medium was replaced with medium containing 100 µg/ml gentamicin, incubated for one hour to kill the extracellular bacteria and then replaced with maintenance medium containing 10 µg/ml gentamicin. After t = 1.5 and t = 15.5 cells were lysed using ice cold 0.1% triton-X100, and intracellular bacterial counts determined after plating the dilutions. All strains were tested independently in triplicate. The ratio of the intracellular bacterial population at 15 hours 30 minutes to 1 hour 30 minutes was designated as fold replication, and was used as the determinant of virulence.

### 2.10.2 Chicken infection experiments

Chicken infection experiments were carried out with Lizeth Lacharme, Paul Wigley and the Author. All work was conducted in accordance with United Kingdom legislation governing experimental animals under project license PPL 40/3652 and was approved by the University of Liverpool ethical review process prior to the award of the license. All birds were checked a minimum of twice daily to ensure their health and welfare. Birds were housed in accommodation meeting UK legislation requirements. 1-day old Lohmann Brown Layers were obtained from a commercial hatchery, separated into groups on arrival and given *ad libitum* access to water and a laboratory-grade vegetable protein-based pellet diet. Chicks were housed at a temperature of 30°C. At 7 days of age chickens were inoculated by oral gavage with $10^8$ CFU of D23580 wild type or D23580 ΔBTP1::FRT. Birds were killed by cervical dislocation at 3 days post infection. The liver, spleen and the caecal contents were removed aseptically from each bird and diluted 1:5 (wt/vol.) in sterile 1X PBS. Tissues were then homogenized in a Colworth 80 microstomacher. Samples were serially diluted and dispensed onto Brilliant Green agar to quantify numbers of *Salmonella* as described previously (Salisbury et al., 2011). Data from caecal content and liver are not shown.

## 2.11 Next Generation sequencing of *Salmonella* isolates referred to Public Health England

### 2.11.1 Illumina short read sequencing & assembly

Illumina sequencing of Public Health England (PHE) *Salmonella* isolates was carried out at PHE, Collindale, London, UK as described by Ashton et al., 2016. DNA extraction for Illumina sequencing of *Salmonella* isolates was carried out using a modified protocol of the Qiasymphony DSP DNA midi kit (Qiagen). In brief, 0.7 ml of an overnight *Salmonella* culture in a 96 deep well plate was harvested. Bacterial cells were pre-lysed in 220 µl of ATL buffer (Qiagen) and 20 µl Proteinase K (Qiagen) and incubated with shaking for 30 mins at 56**°**C. Four microliters of RNase (100 mg/ml; Qiagen) was added to the lysed cells and re-incubated for a further 15 minutes at 37°C. This step increased the purity of the DNA for downstream sequencing. DNA from the treated cells was then extracted on the Qiasymphony SP platform (Qiagen) and eluted in 100 µl of sterile water. DNA concentration was derived using the GloMax system (Promega) and quality (optimal OD260/230 = 1.8 - 2.0) was determined using the LabChip DX system (Perkin Elmer). Extracted DNA was prepared using the NexteraXT sample preparation method and sequenced with a standard 2x100 base pair protocol on a HiSeq 2500 instrument (Illumina, San Diego). Raw FASTQs were processed with Trimmomatic (Bolger et al., 2014) and bases with a PHRED score of less than 30 removed from the trailing end.

### 2.11.2 PacBio sequencing of the U2 UK-ST313 reference strain

Single molecule sequencing was performed on the PacBio RS II instrument at the Centre for Genomic Research, University of Liverpool. DNA was extracted from strain U2 using the Zymo Research Quick-DNA$^{TM}$ Universal Kit (cat# D4069) as per the Biological Fluids & Cells protocol. Extracted DNA was purified with Ampure beads (Agencourt) and the quantity and quality was assessed by Nanodrop and Qubit assays. In addition, the Fragment analyzer (VH Bio), was used to determine the average size of the DNA, using a high sensitivity genomic kit. DNA was sheared to approximately 10kb using a Covaris g-tube and spinning at 5400rpm in an Eppendorf centrifuge. The size range was checked on the Fragment Analyzer. DNA was treated with exonuclease V11 at 37°C for 15 minutes. The ends of the DNA were repaired as described by Pacific Biosciences. Samples were incubated for 20 minutes at 37°C with damage repair mix supplied in the SMRTbell library kit (Pacific Biosciences). This was followed by a 5-minute incubation at 25°C with end-repair mix. DNA was cleaned using 1:1 volume ratio of Ampure beads and 70% ethanol washes. DNA was ligated to adapters overnight at 25°C. Ligation was terminated by incubation at 65°C for 10

minutes followed by exonuclease treatment for 1 hour at 37°C. The SMRTbell library was purified with 1:1 volume ratio of Ampure beads. The library was size-selected on the Blue Pippin (Sage) in the range 7kb-20kb. The DNA was recovered and the quantity of library and therefore the recovery was determined by Qubit assay and the average fragment size determined by Fragment Analyser. SMRTbell libraries were annealed to sequencing primers at values predetermined by the 'Binding Calculator' software (Pacific Biosciences) and complexes made with the DNA polymerase (P6/C4 chemistry). The complexes were bound to Magbeads and loaded onto 3 SMRT cells. Sequencing was done using 360-minute movie times. Sequence data from the 3 SMRT cells was assembled using the HGAP3/Quiver assembler. This resulted in 2 contigs representing the chromosome and the pSLT virulence plasmid. Terminal repeats were manually trimmed to represent circular molecules and the chromosome assembly was reordered so that the sequence started at the *thrL* locus in accordance with convention for *Salmonella* finished genomes. The closed sequences for the U2 chromosome and pSLT virulence plasmid were 4,811,399bp and 93,862bp respectively. Prokka (Seemann, 2014) was used to annotate the two sequences, using the –force flag to preferentially annotate CDS from reference databases FN424405 for the chromosome and AE006471 for the virulence plasmid. The finished U2 genome and annotation were submitted to Genbank and can be accessed using the Genbank accession number LT855376 (chromosome) and LT855377 (virulence plasmid).

## 2.12 Bioinformatic analyses

### 2.12.1 Genome sequences used in this study

All genome sequences used in this thesis are publically available, either as short read sequence data, or finished reference genomes. Accession numbers for all genome sequences used can be found in Appendix i along with the corresponding database where the genome sequence data can be accessed.

### 2.12.2 Prophage annotation

Prophage locations in the D23580 genome were extracted from the Genbank annotation (Accession number: FN424405) and confirmed by submitting the D23580 genome sequence to PHAST (Zhou et al., 2011). The genome of *S.* Typhimurium strain 4/74 (Accession number: CP002487) was used as a comparator to identify D23580-specific prophages. Genome comparison was done using blastn alignment with default parameters and visualised using the Artemis Comparison Tool (ACT) (Carver et al., 2008). Prophage ORFs were characterized by uploading individual prophage sequences to the RAST server, and visualised using the SEED viewer in

the context of the closest phage relatives in the RAST database (Aziz et al., 2008; Overbeek et al., 2014). Protein functions were inferred based on homologies to known proteins by the blastp and psiblast server of NCBI using the UniProt/Swiss-Prot database. Morphological classification was made based on similarity to other well-studied phages. Clustering of functional gene groups was based on a similar analysis by Zou et al., 2010. Assessment of the conservation of the functionally degradative mutations identified in Gifsy-1$^{D23580}$, ST64B$^{D23580}$ and Gifsy-2$^{D23580}$ across other ST313 isolates was achieved using blastn. Genome assemblies were downloaded from Enterobase (http://enterobase.warwick.ac.uk).

### 2.12.3 Genomic analysis of Public Health England *Salmonella* Typhimurium isolates

Genome analysis of sequences of PHE *S.* Typhimurium isolates was performed by Phil Ashton, Public Health England and the Author. Genome sequences from a total of 363 *Salmonella* Typhimurium isolates dating from 2012 and 3,014 *Salmonella* Typhimurium from January 1$^{st}$ 2014 to March 14th 2016 were analyzed. For simplicity, if isolates derived from blood culture, the infection was classed as extra-intestinal. If only a faecal isolate was received for a patient, the infection was classed as gastrointestinal (though this is only suggestive, not conclusive, data that the infection was restricted to the gastrointestinal tract). Of these, 7/363 (1.9%) and 79/3,014 (2.6%) were *Salmonella* Typhimurium ST313, respectively. Full strain metadata can be found in Appendix i. Sequence data (FASTQs) from 23 representative ST313 sequenced by Okoro et al*.* (Okoro et al., 2012) were downloaded from the European Nucleotide Archive (accessions available in Appendix i) and analyzed in the same way as sequence data generated from UK-isolated ST313 isolates.

The multi-locus sequence type (ST) was determined using a modified version of SRST (Inouye et al., 2012). For phylogenetic analysis, processed sequence reads were mapped to the *S*. Typhimurium LT2 reference genome (GenBank: AE006468) using BWA mem (Li and Durbin, 2009). SNPs were called using GATK2 (DePristo et al., 2011) in unified genotyper mode. Core genome positions that had a high quality SNP (>90% consensus, minimum depth 10x, MQ >=30) in at least one strain were extracted and IQ-TREE with parameters –m TVM+ASC –bb 1000 was used to construct a maximum likelihood phylogeny (Nguyen et al., 2015). The TVM model was chosen after using the model test functionality built into IQ-TREE.

To examine the evolutionary history of ST313, four timed phylogenies were constructed using BEAST v1.8.0 (Drummond and Rambaut, 2007), with varying clock

rate models and tree priors. The resulting models were compared in terms of their tree likelihood and posterior and the strict exponential and strict constant models were found to be superior. A comparison using AICM calculated with Tracer v1.6.0 showed that the models had very similar values, tree topologies and branching support in terms of posterior probability were similar between the models. The 95% HPD for the exponential growth rate estimate was -0.0026 to 0.0006; the strict, constant growth model was selected as the estimate of growth rate from the exponential model was around 0 (i.e. constant).

Accessory genome analysis was performed using *de novo* assemblies of quality processed FASTQs produced using SPAdes v2.5.1 using default parameters except –careful and –k 22, 33, 55, 77 (Bankevich et al., 2012). Whole genome assemblies were compared to the reference ST313 strain D23580 using BRIG (Alikhan et al., 2011). Antibiotic resistance loci present in the genome sequences were identified using ResFinder 2.1 (Zankari et al., 2012)

### 2.12.4 RNA-seq based transcriptomic analysis

Investigation of prophage gene expression used data from Canals et al. (2018) and Hammarlöf et al. (2017). Normalised mapped reads for the D23580 prophage regions were visualised in the Integrated Genome Browser (IGB) (Nicol et al., 2009), and transcripts per million (TPM) values (calculated were obtained from Canals et al. (2018) based on the approach of Wagner et al. (Wagner et al., 2012, 2013) representing a normalised expression value for each gene were used to compare gene expression between conditions. As in Kröger et al., 2013, a TPM cut-off score of 10 was used to define gene expression, so that only genes with a TPM value of >10 were considered to be expressed. Heat maps showing absolute expression were made using TPM values and conditional formatting in Microsoft Excel. Prophage transcriptome maps were generated by visualisation of sequence reads in the Integrated Genome Browser (IGB) (Nicol et al., 2009). For display in IGB, the read depth was adjusted in relation to the cDNA library with the lowest number of reads (Skinner et al., 2009). IGB was used to generate prophage transcriptome maps shown in Chapter 5.

For analysis of prophage-regulated core genes (Chapter 5.3), RNA-seq data quality was assessed using FastQC (version 0.11.5) (Andrews, 2010) and then processed with Trimmomatic (version 0.36) (Bolger et al., 2014) to remove Illumina TruSeq adapter sequences, leading and trailing bases with a Phred quality score below 20 and trim reads with an average base quality score of 20 over a 4bp sliding window.

All reads less than 40 nucleotides in length after trimming were discarded from further analysis.

The remaining reads of each library were aligned to the PacBio resequenced D23580 genome (D23580_liv), which differs from the Genbank version by 1 SNP and 2 single nucleotide indels (Canals et al., 2018), using Bowtie2 (version 2.2.9) (Langmead and Salzberg, 2012) and alignments were filtered with Samtools (version 1.3.1) (Li et al., 2009) using a MAPQ cut-off of 15. There was an average of 1,998,911 uniquely mapped sequence reads per sample (Table 2.9).

A custom genome annotation for the D23580_liv genome including sRNAs (Canals et al. 2018; Hammarlöf et al., 2017) was used to define gene expression. Reads were assigned to genomic features using featureCounts (version 1.5.1) (Liao et al., 2014) and count data was imported into R (version 3.2) (R Core Team, 2013). Count data was then filtered using a Counts Per Million (CPM) threshold of 1.2 (corresponding to a count of between 10-15 for library sizes in this study). Trimmed Mean of M values (TMM) scaling to adjust libraries for composition bias was performed using EdgeR (Robinson and Oshlack, 2010). Normalised count data was converted to logCPMs and variance stabilised using VOOM (Law et al., 2014) prior to differential gene expression testing using Limma (Ritchie et al., 2015). The analysis and visualisation tool Degust (Powell, 2017) was used to identify differentially expressed genes, using a minimum gene read count of 10 in at least one replicate (Law et al., 2014), and a minimum CPM of ≥1 in more samples than the size of the smallest group, in this case, 3 (Law et al., 2016). Differentially expressed genes were defined using a Benjamini-Hochberg adjusted P value (false discovery rate- FDR) of ≤0.01 and a fold change value of ≥2.

### 2.12.5 Sequence analysis and sequence identity analysis

Nucleotide sequence identity between single gene or genomic regions was computed using blastn. Protein sequence identity was computed using blastp. Either the command line version (blastall 2.2.17) or the web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used. Sequence identity was reported as a factor of query coverage and query identity, calculated as [query cover x query identity (decimal)]. Sequence alignments computed using BLAST were visualised using the artemis comparison tool (ACT) where required and the artemis genome browser was used to interrogate genome sequences and genome annotations (Carver et al., 2008).

**Table 2.9 Mapping statistics for RNA-seq data described in Chapter 5.6.** Read mapping statistics are shown for all experimental samples (Canals et al., 2018.

| Strain | Condition | % mapped reads | Total no. reads | Uniquely mapped | Not aligned | Multi-mapped |
|---|---|---|---|---|---|---|
| D23580 WT | ESP | 99.82 | 7,106,622 | 2,246,063 | 12,529 | 4,848,030 |
| D23580 WT | ESP | 99.73 | 6,317,700 | 2,459,055 | 17,102 | 3,841,543 |
| D23580 WT | ESP | 99.65 | 6,393,126 | 2,128,022 | 22,254 | 4,242,850 |
| D23580 WT | InSPI2 | 99.78 | 7,233,754 | 1,780,649 | 15,807 | 5,437,298 |
| D23580 WT | InSPI2 | 99.51 | 6,459,816 | 1,462,484 | 31,590 | 4,965,742 |
| D23580 WT | InSPI2 | 99.81 | 6,160,957 | 1,508,700 | 11,477 | 4,640,780 |
| D23580 WT | NoO2 | 99.84 | 7,123,525 | 2,303,151 | 11,305 | 4,809,069 |
| D23580 WT | NoO2 | 99.85 | 8,099,708 | 2,111,827 | 11,836 | 5,976,045 |
| D23580 WT | NoO2 | 99.85 | 8,667,400 | 2,534,019 | 12,934 | 6,120,447 |
| D23580 ΔBTP1 | ESP | 99.88 | 7,374,123 | 2,474,245 | 8,895 | 4,890,983 |
| D23580 ΔBTP1 | ESP | 99.79 | 7,608,355 | 2,691,343 | 15,825 | 4,901,187 |
| D23580 ΔBTP1 | ESP | 99.65 | 6,894,681 | 3,153,897 | 24,414 | 3,716,370 |
| D23580 ΔBTP1 | InSPI2 | 99.82 | 7,375,387 | 2,035,307 | 13,143 | 5,326,937 |
| D23580 ΔBTP1 | InSPI2 | 99.86 | 6,616,691 | 1,746,450 | 9,138 | 4,861,103 |
| D23580 ΔBTP1 | InSPI2 | 99.79 | 7,503,062 | 2,023,998 | 15,983 | 5,463,081 |
| D23580 ΔBTP1 | NoO2 | 99.73 | 7,843,238 | 1,985,195 | 21,417 | 5,836,626 |
| D23580 ΔBTP1 | NoO2 | 99.79 | 6,967,213 | 1,814,589 | 14,931 | 5,137,693 |
| D23580 ΔBTP1 | NoO2 | 99.80 | 6,484,729 | 1,868,969 | 13,021 | 4,602,739 |
| D23580 ΔBTP1 ΔBTP5 | ESP | 99.75 | 6,648,857 | 1,763,987 | 16,884 | 4,867,986 |
| D23580 ΔBTP1 ΔBTP5 | ESP | 99.74 | 5,598,561 | 1,708,604 | 14,561 | 3,875,396 |
| D23580 ΔBTP1 ΔBTP5 | ESP | 99.79 | 7,171,494 | 2,279,589 | 14,961 | 4,876,944 |
| D23580 ΔBTP1 ΔBTP5 | InSPI2 | 99.78 | 6,896,453 | 1,760,546 | 15,141 | 5,120,766 |
| D23580 ΔBTP1 ΔBTP5 | InSPI2 | 99.75 | 6,674,203 | 1,624,967 | 16,994 | 5,032,242 |
| D23580 ΔBTP1 ΔBTP5 | InSPI2 | 99.81 | 7,741,768 | 1,930,829 | 14,441 | 5,796,498 |
| D23580 ΔBTP1 ΔBTP5 | NoO2 | 99.64 | 5,187,790 | 1,302,735 | 18,563 | 3,866,492 |
| D23580 ΔBTP1 ΔBTP5 | NoO2 | 99.61 | 5,770,089 | 1,777,678 | 22,638 | 3,969,773 |
| D23580 ΔBTP1 ΔBTP5 | NoO2 | 99.68 | 5,601,759 | 1,264,197 | 17,921 | 4,319,641 |
| D23580 ΔBTP5 | ESP | 99.70 | 7,189,101 | 2,038,582 | 21,827 | 5,128,692 |
| D23580 ΔBTP5 | ESP | 99.82 | 6,718,987 | 2,117,221 | 11,763 | 4,590,003 |
| D23580 ΔBTP5 | ESP | 99.80 | 7,823,044 | 2,269,812 | 15,794 | 5,537,438 |
| D23580 ΔBTP5 | InSPI2 | 99.88 | 8,297,362 | 2,022,696 | 9,674 | 6,264,992 |
| D23580 ΔBTP5 | InSPI2 | 99.89 | 8,785,241 | 2,178,345 | 9,657 | 6,597,239 |
| D23580 ΔBTP5 | InSPI2 | 99.79 | 7,278,270 | 1,879,197 | 15,007 | 5,384,066 |
| D23580 ΔBTP5 | NoO2 | 99.81 | 6,352,867 | 1,748,201 | 12,141 | 4,592,525 |
| D23580 ΔBTP5 | NoO2 | 99.67 | 5,949,282 | 1,909,743 | 19,578 | 4,019,961 |
| D23580 ΔBTP5 | NoO2 | 99.86 | 7,732,496 | 2,055,886 | 11,054 | 5,665,556 |
| Average | - | 99.77 | 6,990,214 | 1,998,911 | 15,617 | 4,975,687 |

# Chapter 3

## The Prophage Repertoire of *S.* Typhimurium ST313

## 3.1 Introduction

All complete genomes of *Salmonella* encode prophages, and therefore it is likely that prophages are harboured by every member of the *Salmonella* genus (Bobay et al., 2014; Lemire et al., 2008; Villafane et al., 2008). The prophage sequences found in *Salmonalla* genomes may be classified into 3 classes: functional, defective and remnant.

Functional prophages encode all necessary faculties to complete each stage of the prophage lifecycle, including infection, lysogeny, induction and infection. Defective prophages arise when mutations occur which impede any stage of the prophage life cycle. Mutations that produce defective prophages include those that affect gene expression such as disruption of a promoter sequence; or mutations that inhibit protein function, such as frame-shift mutations in open reading frames (ORFs) or single nucleotide polymorphisms (SNPs) resulting in non-synonomous amino acid substitutions. Remnant prophages are prophage-derived sequence islands that have lost many of the genes required for prophage function. Remnant prophages may be described as 'domesticated', because they frequently contain only the subset of genes from the ancestral functional prophage which were beneficial to the host (Bobay et al., 2014). For example, in *Salmonella,* many virulence genes are carried on prophage remnant sequence islands (Casjens, 2011). The distinction between defective and remnant prophages is important, and these terms have been frequently used erroneously. Strictly, a defective prophage may be resuscitated by spontaneous reversion of the causative mutation, whilst a remnant prophage generally results from large deletions and extensive gene loss over a long period of evolutionary time, rendering resuscitation very unlikely. Defective prophages may therefore be thought of as recent evolutionary events which can be isogenic in closely-related strains, whereas remnant prophages are generally conserved across many strains (even those that are distantly related), and may be considered to be evolutionarily 'ancient'. Theoretically, a remnant prophage could be pseudo-resuscitated by recombination with a functional prophage, however this process would require the introduction of a substantial amount of new genetic material, which is not required for resuscitation of a defective prophage.

Due to the subtleties of the distinctions between functional, defective and remnant prophages, it is difficult to define prophage status *in silico*, though prophage remnants are generally distinguishable due to their short length and lack of essential phage genes, for example structural genes. Functional and defective prophages can only be definitively distinguished experimentally, however, a prophage with no overt genetic

mutations that does not form plaques on putatively susceptible strains can never be definitively proven to be defective, due to the possibility the prophage being studied is simply intractable to standard phage-detection methods. This chapter will characterise the prophage sequences of ST313 representative strain D23580 in terms of gene content and functional status, and describe the phage biology of the functional prophages.

### 3.1.1 Acknowledgement of the specific contribution of collaborators to the results described in Chapter 3.

Much of the content of this chapter is published in the research article Owen et al., (2017). I acknowledge the following contribution of collaborators to the results described in this chapter. Unless specified below, all work was completed by the Author.

| | |
|---|---|
| **Nicolas Wenner**<br>University of Liverpool, UK | Construction of suicide plasmids for deletion of the prophages (2.8.2), construction of single nucleotide substitution mutant D23580 $P_{dinI\text{-}gfoA}^{14028}$ (2.6.2) |
| **Angela Makumi, Abram Aertsen**<br>KU Leuven, Belgium | Construction (2.6.1) of $RecA^-$ strains and assay of RecA-independent induction of prophages BTP1 & P22 (3.3.4) |

## 3.2 The ST313 prophage repertoire

Six prophages were originally annotated in the genome of ST313 representative strain D23580 (Kingsley et al., 2009). I conducted a more detailed analysis, and found that in fact, D23580 contains a total of 9 phage-derived sequence islands (Figure 3.1A, Table 3.1). Four short prophage-remnants were named Def1-4, using the nomenclature defined for *S.* Typhimurium strain LT2 (Casjens, 2011). Def1-4 are found in all complete *S.* Typhimurium genomes currently available on NCBI (accessed 28th July 2016) (data not shown) and are likely to be ancestral to the serovar (Casjens, 2011). One feature designated 'BTP6' was included in the original annotation of D23580 (accession FN424405) (Kingsley et al., 2009) and is located at the position of prophage remnant Def4. The remaining five prophage regions, designated BTP1-5 by Kingsley et al. (2009), are full-length prophages (Table 3.1) that contain all the functional gene clusters required for putative phage function such as integration/excision, immunity, lysis, capsid and tail gene clusters. My analysis determined that three represent well-characterised prophages, commonly found in *S.* Typhimurium genomes: BTP2 is Gifsy-2 $^{D23580}$, BTP3 is ST64B $^{D23580}$ and BTP4 is Gifsy-1 $^{D23580}$ (Figueroa-Bossi et al., 2001; Tucker and Heuzenroeder, 2004). In contrast, BTP1 and BTP5 are novel prophages that are only found in ST313 and not present in the comparator ST19 strain 4/74 (Figure 3.1B). When compared to ST19 strain 4/74, the sopEΦ prophage is absent from D23580, however this prophage is variably present in ST19 isolates; sopEΦ is not present in most of the reference ST19 strains such as LT2 or 14028 (Mirold et al., 1999). Together, the phage-derived sequence islands found in D23580 harbour many characterised virulence-related genes in *S.* Typhimurium including genes encoding effector proteins and O-antigen modification genes (*gtr* genes) (Table 3.1).

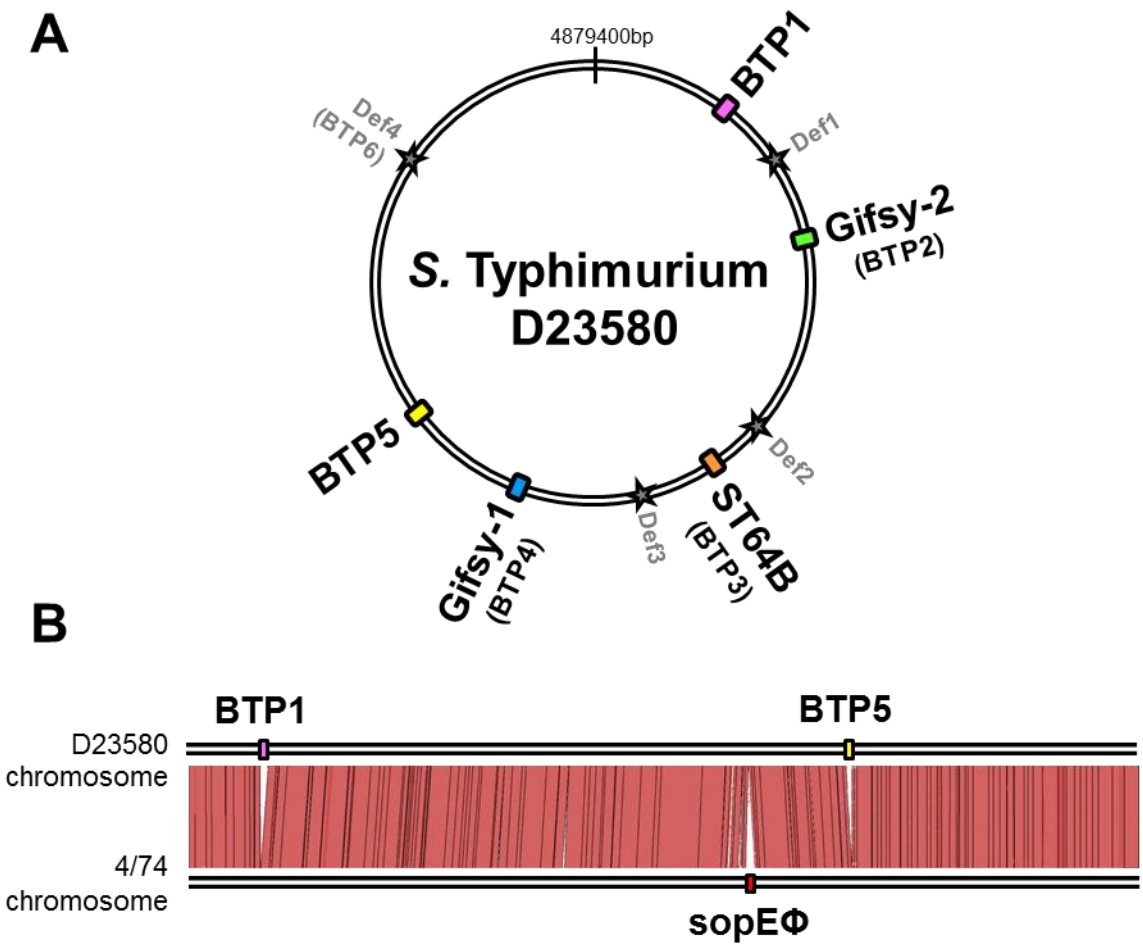



**Figure 3.1 The prophage repertoire of S. Typhimurium ST313 strain D23580, and genome comparison to ST19 strain 4/74.** A. Full length (functional or defective) prophages are labelled in bold and represented by coloured rectangles, whilst remnant prophages are labelled in grey and represented by grey stars. B. Alignment of the D23580 chromosome with comparator ST19 strain 4/74 using blastn. Red coloured blocks indicate individual blastn hits of more than 90% identity. White regions represent sequence that is unique to either genome. The unique prophages in both genomes are annotated.

**Table 3.1 Location of all prophage or remnant prophage elements in the D23580 chromosome and the encoded virulence-related genes.** All elements are present in ST19 representative strain 4/74, apart from BTP1 and BTP5 (highlighted in red).

| Prophage / remnant | Location on D23580 chromosome | Length (bp) | Virulence-related genes |
|---|---|---|---|
| **BTP1** | 368797 - 409321 | 40,525 | ***st313-td*** (*STMMW_03531*), ***gtrCc*** (*STMMW_03911*), ***gtrAc*** (*SMMTW_03921*) |
| **Def1** (SPI-16) | 652441 - 656532 | 4,091 | ***gtrCa*** *(STMMW_06231),* ***gtrBa*** (*STMMW_06241*), ***gtrAa*** (*STMMW_06251*) |
| **Gifsy-2** $^{D23580}$ (BTP2) | 1094166 - 1140420 | 46,254 | ***gtgA*** (*STMMW_10381*), ***sodCI*** (*STMMW_10551*), ***sseI*** (*STMMW_10631*-pseudogene), ***gtgE*** (*STMMW_10681*) ***gtgF*** (*STMMW_10691*) |
| **Def2** | 1945242 - 1951433 | 6,191 | ***sopE2*** (*STMMW_18441*) |
| **ST64B** $^{D23580}$ (BTP3) | 2065743 - 2105830 | 40,087 | ***sseK3*** (*STMMW_19812*) |
| **Def3** (inc. SPI-12) | 2354483 - 2369638 | 15,155 | ***sspH2*** (*STMMW_22721*) |
| **Gifsy-1** $^{D23580}$ (BTP4) | 2753352 - 2803478 | 50,126 | ***gogB*** (*STMMW_26001*), ***gipA*** (*STMMW_26191*), ***gtgA*** (*STMMW_26331*), |
| **BTP5** | 3370278 - 3401458 | 31,181 | **-** |
| **Def4** | 4440630 - 4461045 | 20,415 | ***gtrCb*** (STMMW_41431), ***gtrBb*** (*STMMW_41541*), ***gtrAb*** (*STMMW_41551*) |

### 3.2.1 Functionality of the D23580 prophages

Prophages were initially defined as functional if viable (infectious) phage particles could be produced, with or without chemical induction by the DNA-damaging agent mitomycin C. To test the functionality of each of the D23580 prophages, a series of naive indicator strains of D23580 were created using scarless genome editing to delete each prophage (Chapter 2.8.2). This resulted in a collection of five isogenic D23580 derivative strains, each lacking a single prophage (Table 3.2A) and harboring intact phage attachment sites (*attB*). As Gifsy phages have previously been shown to yield visible plaques only on hosts with short LPS, due to the receptor OmpC being blocked by the O-antigen (Ho and Slauch, 2001), an additional *galE496* mutation was introduced to facilitate detection of Gifsy-phage plaques (Chapter 2.6.1). Specific indicator strains used to detect each phage are shown in Table 3.2.

The supernatant from a 16 h LB culture of strain D23580 yielded circular plaques of ~1 mm in diameter with turbid centres on the specific indicator strain D23580 ΔBTP1 at a remarkably high abundance of ~$10^9$ PFU/ml (plaque-forming units per ml). Mitomycin C induction increased the phage titer 10-fold to ~$10^{10}$ PFU/ml (Table 3.2A). PCR from isolated plaques confirmed that the plaques were from the BTP1 phage (data not shown). However, plaques for the remaining four prophages were not detected using the specific indicator strains, with or without mitomycin C induction.

Prophages may contain superinfection exclusion systems that give cross-immunity to other prophages. To control for the possibility that other D23580 prophages may be excluding the target phage in the specific indicator strains, a strain with all five full-length prophages deleted was constructed (designated D23580 ΔΦ). When supernatant from D23580 ΔBTP1 was plated onto a lawn of D23580 ΔΦ, no plaques were observed, supporting the conclusion that of the five D23580 prophages, only BTP1 is capable of producing viable phages (Table 3.2B).

**Table 3.2 Detection of active prophage in D23580 by plaque formation on specific indicator strains for each D23580 prophage.** 'Not detected' indicates spontaneous phage induction at less than the detection threshold of $1x10^2$ PFU/ml. A. D23580 was included as an indicator strain to confirm that the strain does not support plaques from induced phages present in its own supernatant. B. To control for the possibility that other prophages may be super-excluding the target phage in the specific indicator strains, supernatant from D23580 ΔBTP1 was plated on an indicator strain with all 5 prophages removed (D23580 ΔΦ), yielding no detectable plaques. Methods describing plaque assay can be found in Chapter 2.4.2.

**A**

| | | Specific phage titer (PFU/mL) in D23580 WT supernatant * | |
|---|---|---|---|
| **Specific indicator strain** | **Target prophage** | **Non-induced** | **+ Mitomycin C** |
| D23580 ΔBTP1 | BTP1 | $1x10^9$ | $2x10^{10}$ |
| D23580 ΔGifsy-2 *galE496* | Gifsy-2 (BTP2) | Not detected | Not detected |
| D23580 ΔST64B | ST64B (BTP3) | Not detected | Not detected |
| D23580 ΔGifsy-1 *galE496* | Gifsy-1 (BTP4) | Not detected | Not detected |
| D23580 ΔBTP5 | BTP5 | Not detected | Not detected |
| D23580 ΔGifsy-1 *galE496* | Gifsy-1 $P_{dinI\text{-}gfoA14028s}$ | $1x10^4$ | $2x10^6$ |
| D23580 WT | Putative superimmune phage (control) | Not detected | Not detected |

**B**

| | | Specific phage titer (PFU/mL) in D23580 ΔBTP1 supernatant * | |
|---|---|---|---|
| **Specific indicator strain** | **Target prophage** | **Non-induced** | **+ Mitomycin C** |
| D23580 ΔΦ (All 5 prophages deleted) | Gifsy-1, Gifsy-2, ST64B or BTP5 | Not detected | Not detected |

* Supernatant from an overnight cultures prepared as described in 2.3.1.

## 3.3 BTP1

### 3.3.1 Gene content

The gene content and synteny of BTP1 resembles that of model *Salmonella* phage P22, exhibiting an overall 50% similarity at the nucleotide level, suggesting that BTP1 belongs to the *Podoviridae* family of short-tailed phages (Ackermann, 2009). Functional annotation of BTP1 genes identified a number of putative cargo loci on the prophage: independent transcriptional units that are not required for the prophage lytic or lysogenic cycles (Cumby et al., 2012). Like P22, BTP1 contains a *gtrAC* operon (*STMMW_03911-03921*) (Figure 3.2) which encodes a glycosyltransferase enzyme (GtrC) that modifies the O-antigen of the bacterial lipopolysaccharide (Kintz et al., 2015). This glycosyltransferase mediates the acetylation of the rhamnose residue of the O-antigen and is classified as a family 2 GtrC protein, which is typically associated with invasive *Salmonella* serovars such as Typhi and Paratyphi A (Kintz et al., 2015). Previously it has been shown that a BTP1 lysogen is highly resistant to superinfection by P22 (Kintz et al., 2015), suggesting the two phages contain cross-protective superinfection immunity systems, likely the *gtr* glycosyltransferase.

The *st313-td* gene (*STMMW_03531*) is encoded immediately downstream of the *cI* repressor locus. According to a recent paper, deletion of *st313-td* decreases survival within murine macrophages, and results in attenuation in a mouse infection model (Herrero-Fresno et al., 2014). At the time of description of the *st313-td* gene, prophage BTP1 had not been characterised and therefore the fact that *st313-td* was BTP1-encoded has not been previously explored. Additionally, the *pid* ORFan locus (*STMMW_03751*) is present in BTP1. The P22-encoded Pid regulatory protein de-represses the host *dgo* operon in *S.* Typhimurium LT2 (Cenens et al., 2013). The complete functionally annotated map of BTP1 is shown in Figure 3.2. The chromosomal context of the BTP1 attachment site and the core *att* sequence for BTP1 is shown in Figure 3.3.

A comparative genomic analysis assigned putative function to 43/64 (62.5%) of the annotated open reading frames (ORFs) of BTP1 (Figure 3.2). A comparison with model phages P22 and Lambda revealed that BTP1 ORFs can be classified into 10 functional blocks; *int/xis, ea* region, Recombination, Immunity, Replication, *nin* region, Lysis, Capsid & tail and *gtr* (glycosyltransferase) (Figure 3.2).

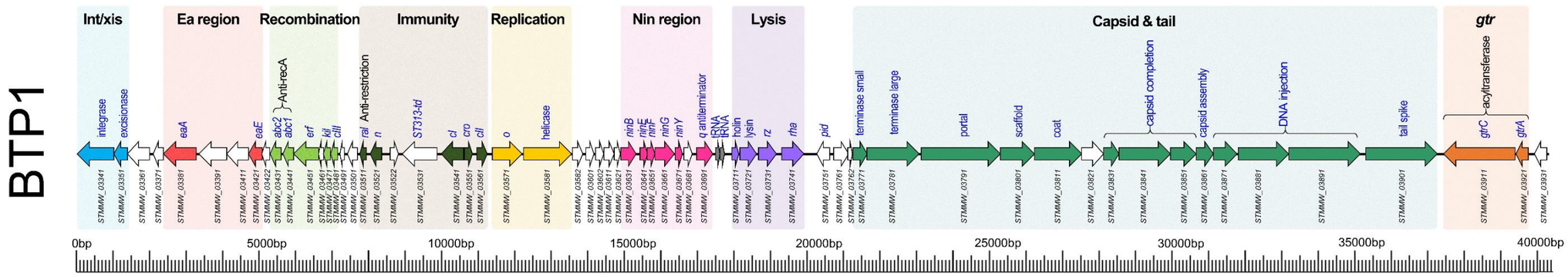
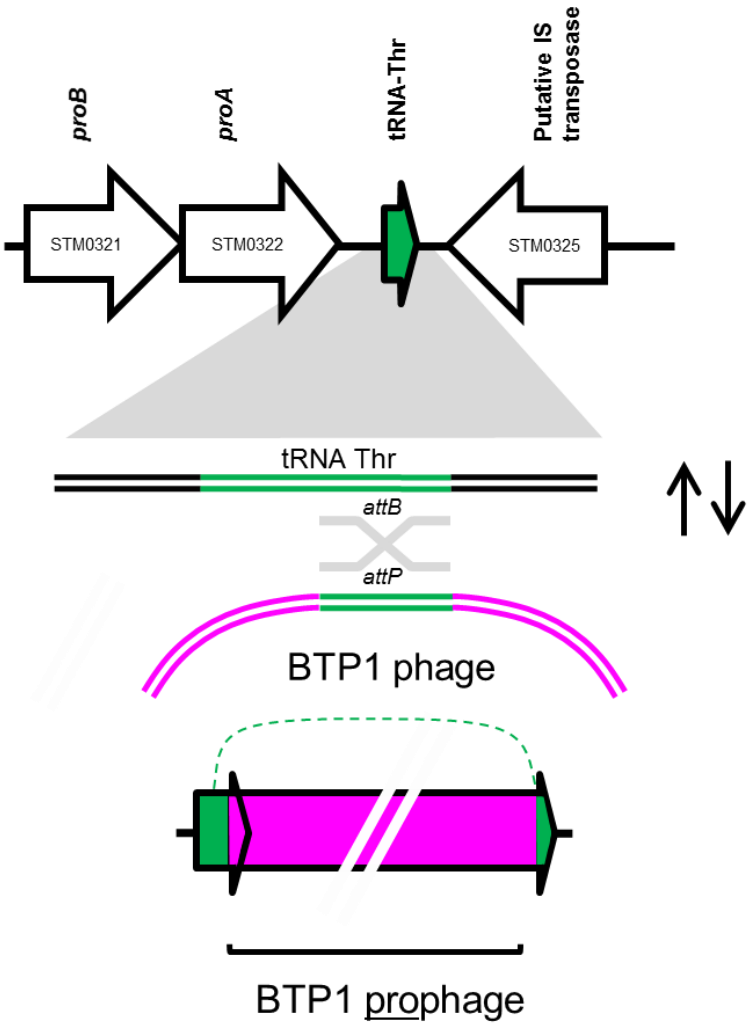



**Figure 3.2 Functionally annotated map of the BTP1$^{D23580}$ prophage sequence.** Coloured arrows represent ORFs for which a function can be inferred based on sequence. White arrows represent ORFs for which no function can be inferred.

A




B

| Site | Sequence |
|---|---|
| BTP1 *att* core | ATTCGTAATGCGAAGGTCGTAGGTTCGACTCCTATTATCGGCACCA |

**Figure 3.3 Genomic context and core sequence of the BTP1 attachment site**

### 3.3.2 BTP1 homology to other sequenced prophages and phages

To further understand the genetic context of BTP1 in phage sequence space, the nucleotide sequence of the BTP1 prophage was compared to other sequenced phages and prophages (Figure 3.4). The functional blocks of BTP1 show mosaic conservation in other phage and prophage genomes suggesting functional overlap and recombination. Similarity to other phages extends beyond phages that infect the *Salmonella* genus*,* as BTP1 shares sequence identity with P22-like phages of *E. coli* and *Shigella* (CUS-3, HK620 and Sf6). Some BTP1 functional blocks are rarer among the sequences sampled in this analysis than others. For example, the BTP1 *gtr* block is only found elsewhere in *Salmonella* phage ε34. Others, such as the capsid & tail block, *int/xis* block and recombination block, are found in the majority of sampled sequences. The immunity regions of the P22-like phages and prophages also show mosaicism, with some phages sharing an almost identical immunity region, and other phages completely lacking homology to the immunity region of BTP1. For example, there is no significant nucleotide homology between the immunity regions of BTP1 and P22. In contrast, prophage ST104 contains a comparable immunity region to BTP1 (Figure 3.5). The differences between the immunity regions suggest the regulation of lysogeny may differ between these prophages. Notably, the *sieB* gene, encoding an asRNA based superinfection exclusion system (Ranade and Poteete, 1993), is present in the immunity region of P22 and ST104, and is absent from the immunity region of BTP1.

Comparison of the BTP1 and P22 late genes using tblastx shows that other structural regions share extensive homology, with the notable absence of the P22 immunity region (*imml*) in BTP1 (Figure 3.5). Due to the high level of sequence identity and the detailed three-dimensional structural information that is available for many P22 proteins, it was possible to map the protein coding genes of BTP1 onto a structural model of P22 (Figure 3.6). The capsid and tail genes of BTP1 show a high level of similarity to P22, with the majority of proteins sharing >90% amino acid identity. The genes involved in DNA injection (*STMMW_03871-91*) show less homology, putatively suggesting subtle differences between the mechanisms of DNA entry of the phages. The tailspike protein of P22-like phages consists of a short head-binding domain (HBD) which attaches the tailspike to the capsid, and a receptor-binding domain (RBD) responsible for recognition and binding to the bacterial receptor (Steinbacher et al., 1997). The HBD of the BTP1 tailspike shares 97% sequence identity with P22, however the receptor-binding domain is more divergent and is just 65% identical. This

may indicate differences in the enzymatic activity of the tailspike protein, which in P22 is known to have endorhamnosidase activity (Iwashita and Kanegasaki, 1973).

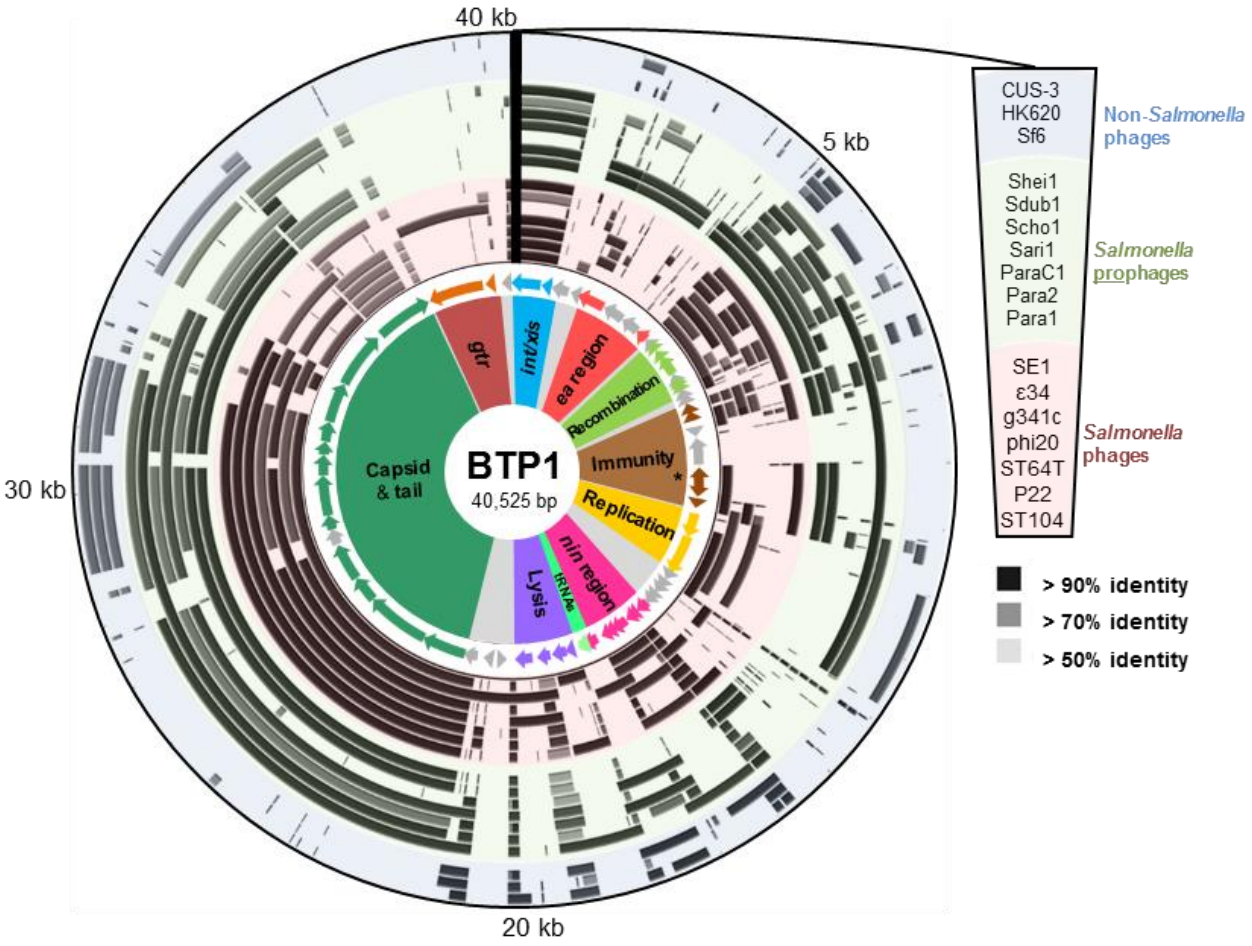



**Figure 3.4 Functional genetic architecture of BTP1 and homology to sequenced *Salmonella* phages and prophages, and non-*Salmonella* phages.** Starting from the inside, circular panels represent functional clusters of genes: ORFs colored according to functional cluster with functionally-unknown ORFs colored grey; BLAST homology to other *Salmonella* phage sequences (red-shaded ring), *Salmonella* prophage sequences (green-shaded ring) and non-*Salmonella* phages (blue-shaded ring). The order of sequences represented by each concentric ring of the BLAST homology panel are shown on the right.

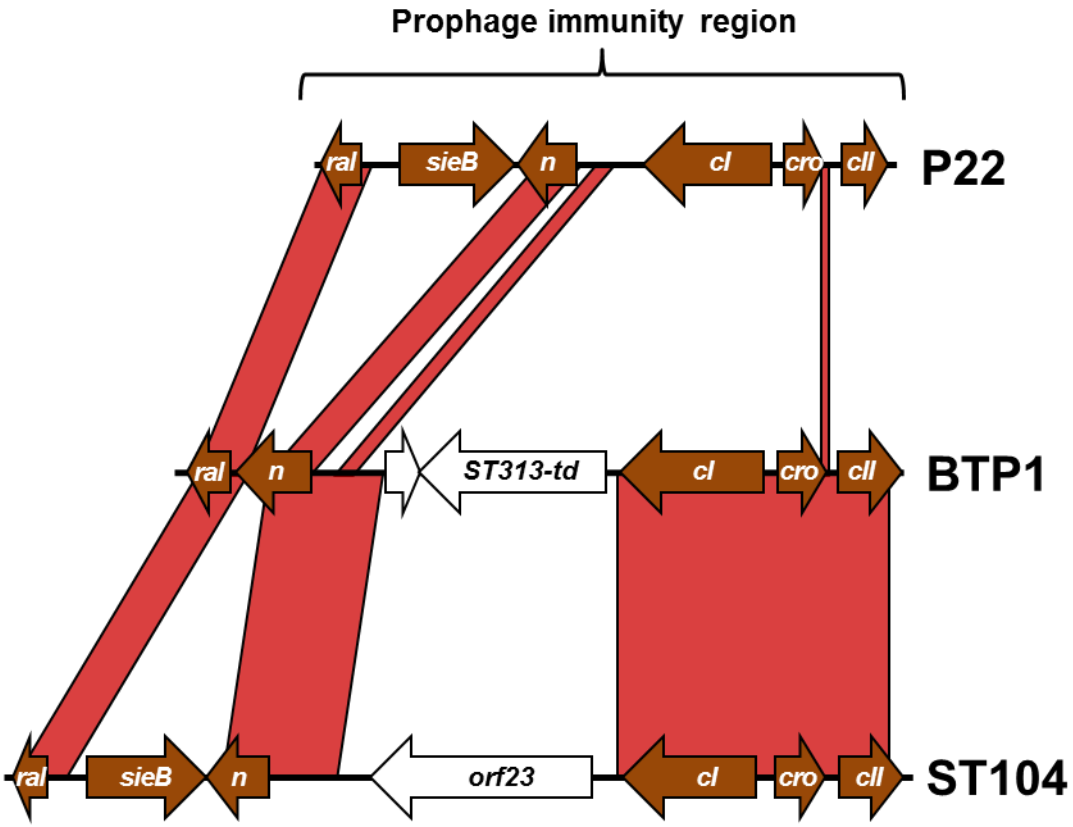



**Figure 3.5 The immunity region of BTP1 compared to P22 and ST104.** Comparison of nucleotide similarity (computed using blastn) of the immunity regions of phage P22, BTP1 and ST104. Red blocks indicate individual blastn hits of >90% identity.

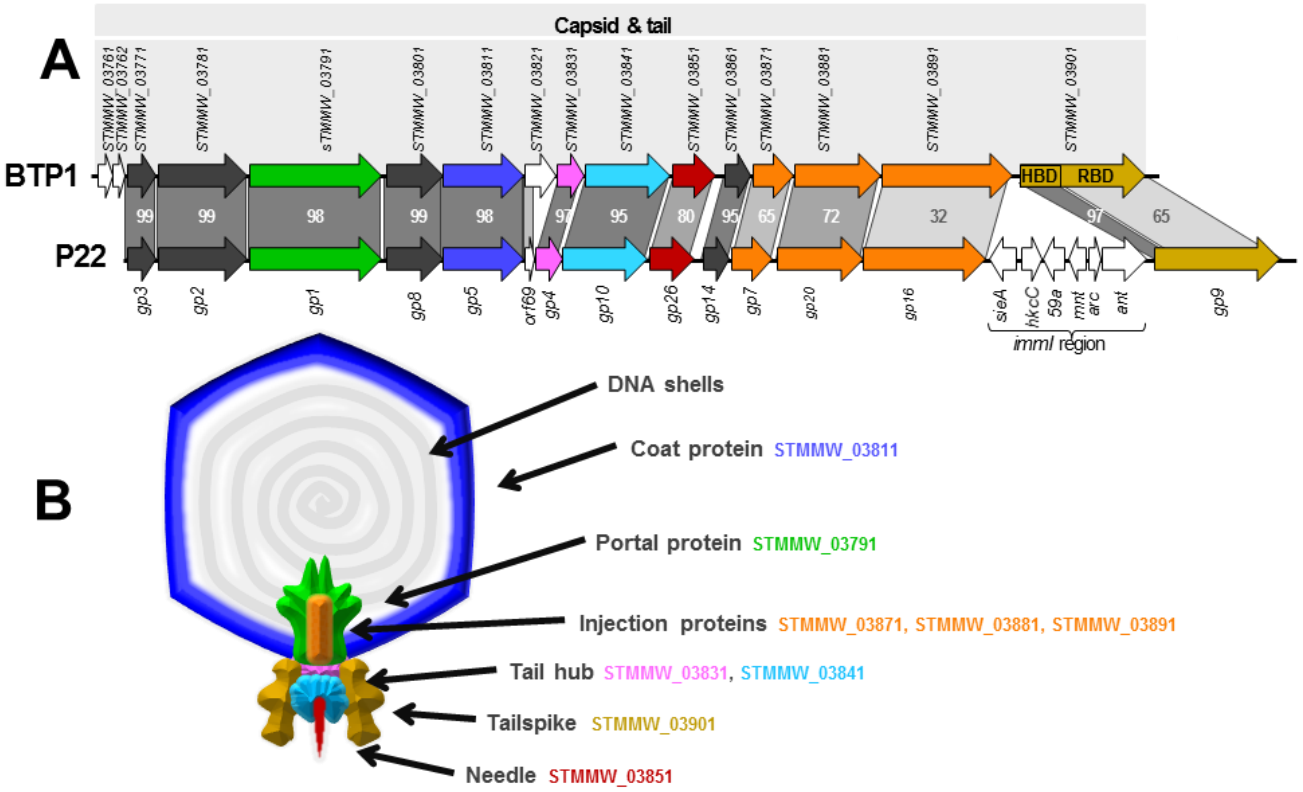



**Figure 3.6 Structural gene cluster of BTP1 and virion protein model based on phage P22.** (A) The structural gene cluster of prophage BTP1 consists of 13 ORFs of known function in virion structure and assembly (shown as colored arrows). Additionally, there are three ORFs of unknown function (shown as white arrows). Model phage P22 contains an almost identical structural cluster, but with the major addition of the *imml* region. Grey bars show the percentage amino acid identity of gene homologues. Published CDS annotation is shown for both regions. The gene encoding the tailspike (*STMMW_03901*) is separated into its two functional domains: the head-binding domain (HBD) and the receptor-binding domain (RBD). (B) Diagram of predicted structure and location of structural proteins in BTP1, based on cryoEM reconstruction of phage P22 by Tang *et al.,* 2011. BTP1 shares near identical structural gene content with P22 and colored protein segments correspond to the protein coding genes in (A).

The complete BTP1 prophage was not detected in any other complete *Salmonella* genomes on publically available databases. However, analysis of draft genome sequences on NCBI revealed that a prophage highly similar to BTP1 was present in a genome of a *S.* Typhimurium isolate belonging to the sequence type ST568. Strains belonging to ST568 are commonly isolated from wild passerine (garden) birds in the UK and frequently belong to the DT56v. phage type (Hughes et al., 2010; Mather et al., 2016). Interestingly, African ST313 isolates carrying the BTP1 prophage most often belong to the DT56v. phage type (Kingsley et al., 2009). In order to understand the presence and distribution of the BTP1 prophage in *S.* Typhimurium ST568, strains belonging to this sequence type were investigated using the open-access genome sequence and metadata database, Enterobase, which contains over 100,000 *Salmonella* genome sequences (https://enterobase.warwick.ac.uk/). 75 ST568 genomes are available on Enterobase, and metadata is available for of the 36 genomes. The majority originated in the UK, and ST568 has also been isolated in Ireland, France and Myanmar. Though the sequence type has previously been associated with passerine birds in the UK, metadata available in Enterobase suggest that this sequence type can also cause disease in humans, as the majority of genomes with metadata available are of human origin. However, epidemiological conclusions cannot be drawn from this data, as sampling of *Salmonella* is hugely biased towards human isolates (which are routinely sequenced by Public Health England, and Public Health England are one of the largest contributors to Enterobase).

The Enterobase ST568 genome with the best genome assembly (as judged by highest N50 value) was selected as a representative genome for further analysis of the BTP1$^{ST568}$ prophage. The genome assembly chosen was for a strain called 201102250 which was isolated in 2011 and uploaded to Enterobase by F.X. Weill (Institut Pasteur, Paris); though no country or isolation source are provided. The intact BTP1 prophage sequence was located on a large contig of 419,117 bp. The prophage was 99.995% identical to BTP1$^{D23580}$, representing a difference of just 2 SNPs (Figure 3.7A). Analysis of the chromosomal sequence surrounding the prophage on the contig indicated that the attachment site of BTP1$^{201102250}$ was the same as BTP1$^{D23580}$ (tRNA-Thr) (data not shown).

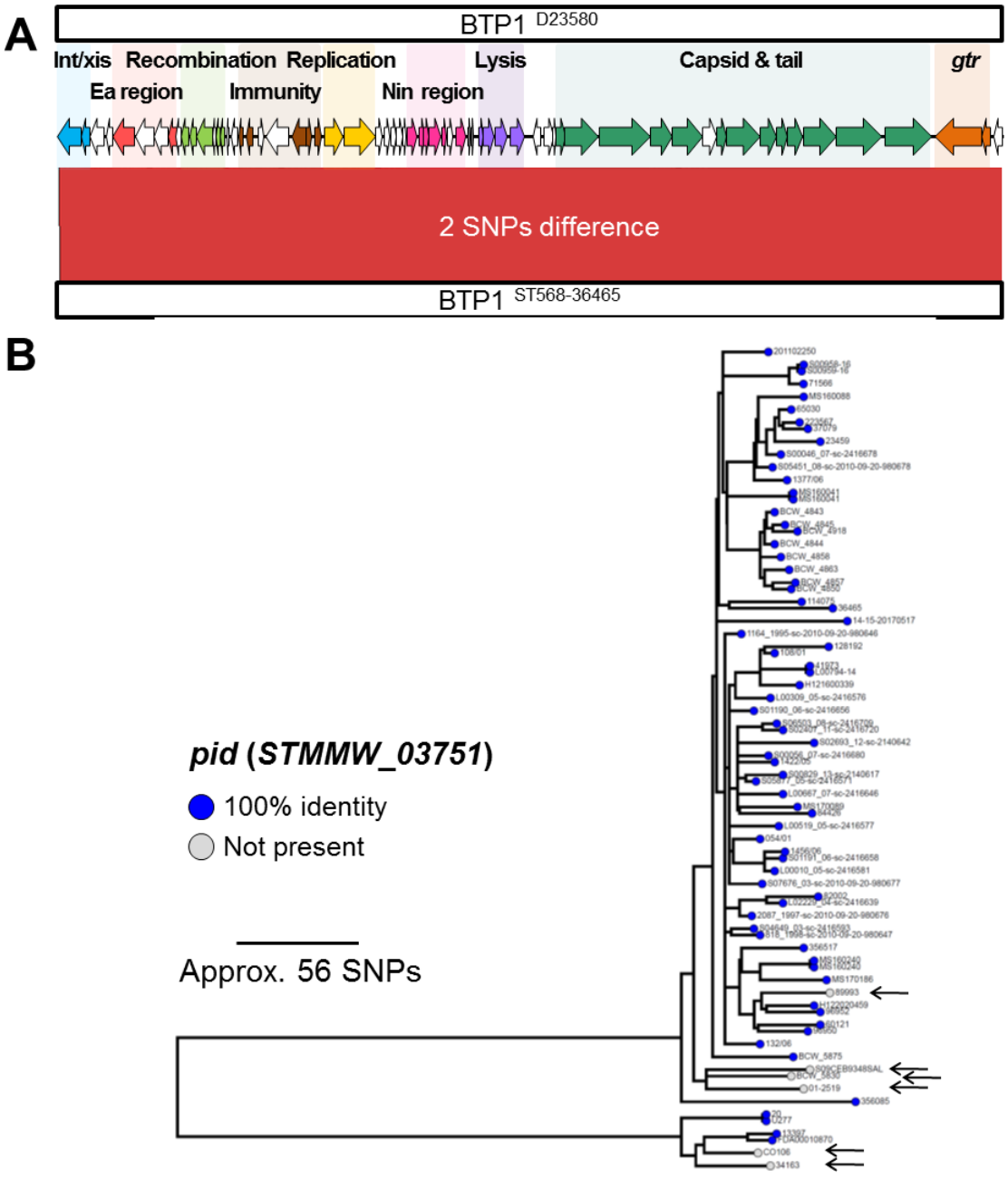



**Figure 3.7 Prophage BTP1 is found in *S.* Typhimurium ST568.** A. BTP1 is 99.995% identical to a prophage from ST568 strain 201102250 and differs by just 2 SNPs. B. The *pid* gene, representing BTP1 is present in the majority of ST568 genomes sequences available on Enterobase. Black arrows highlight the genomes where the *pid* gene (and, by inference, BTP1) is not present (grey nodes).

The *pid* gene was used as a marker of the BTP1 prophage, in order to screen all 75 ST568 genomes on Enterobase and these data were visualised in the context of a core SNP phylogeny of the ST568 isolates (Figure 3.7B). The BTP1 prophage (indicated by the presence of the *pid* (*STMMW_03751*) gene) was present in 69 of the 75 ST568 genomes. As both the ST313 and ST568 sequence types of *S.* Typhimurium that carry the BTP1 prophage are associated with the phage type DT65v., it is speculated that the BTP1 prophage is the key molecular determinant of this phage type i.e. the presence of the BTP1 prophage in *S.* Typhimurium genomes results in the DT56v. phage type. Furthermore, the presence of BTP1 in a clade of *S.* Typhimurium associated with wild birds raises interesting questions concerning the transmission of BTP1 between the two distinct sequence types of *S.* Typhimurium. As many wild bird species follow annual migratory cycles, which for some species may include sub-Saharan Africa, it is possible that BTP1 was transmitted to ST313 in Africa from ST568 in the UK, or *vice versa*, to ST568 in the UK from ST313 in Africa, via wild birds. Though such a scenario cannot be verified from available data, it may become an important surveillance consideration should the BTP1 prophage be implicated in the disease tropism of *S.* Typhimurium ST313.

### 3.3.3 Structural morphology and taxonomic classification of the BTP1 phage

The morphology of the BTP1 phage was determined by Transmission Electron Microscopy (TEM). The purified BTP1 phage preparation contained short-tailed phages with icosahedral capsids of approximately 65 nm in diameter (Figure 3.8B). The addition of purified BTP1 phages to susceptible *S.* Typhimurium 4/74 cells allowed the same short-tailed phages to be seen adsorbing to the surface of the bacterial cells (Figure 3.8A). Empty ‘ghost’ virions were also observed, suggesting that some phages had already adsorbed and injected DNA into the host cell at 20 minutes post-infection.

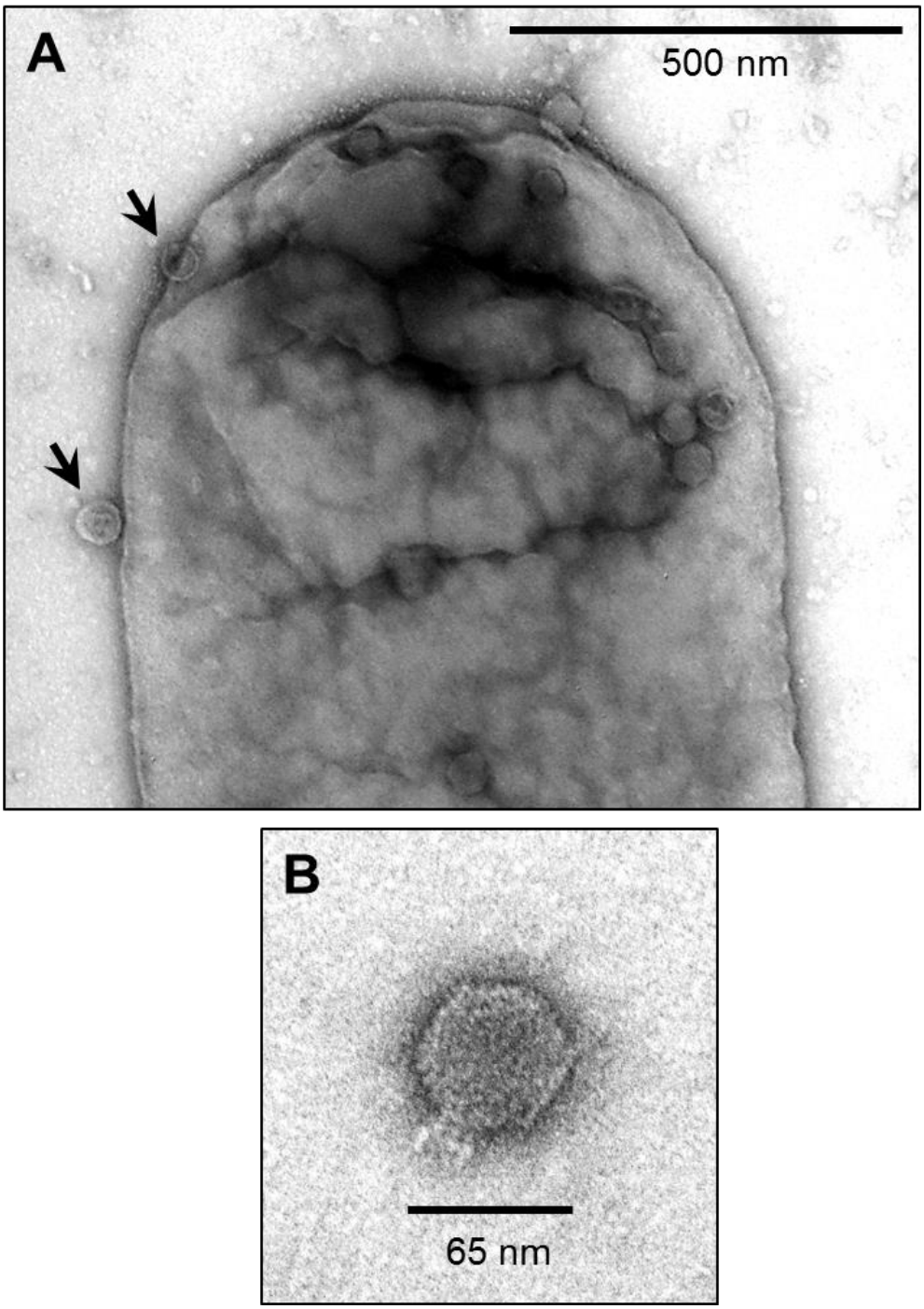



**Figure 3.8 TEM of BTP1 virions infecting *S.* Typhimurium 4/74.** BTP1 virions from D23580 overnight supernatant (Chapter 2.3.1) were added to mid-exponential phase 4/74 ($OD_{600}$ =0.3) cells in liquid culture. After 20 minutes samples were prepared for TEM with negative staining using uranyl acetate (Chapter 2.4.8). (**A)** BTP1 virions adsorbing to a 4/74 cell (black arrows). (**B)** An isolated BTP1 virion.

I propose that BTP1 represents a new species designated *Salmonella* virus BTP1 within the *P22virus* genus (*taxonomic proposal pending acceptance*). This is consistent with the International Committee on Taxonomy of Viruses (ICTV) criterion that members of a proposed species differ from those of other species by more than 5% at the DNA level as confirmed with the blastn algorithm, and the phage has been verified by electron microscopy. The BTP1 phage genome was sequenced and re-ordered to begin at the small terminase subunit, to conform to the reference sequence for the *P22virus* genus, *Salmonella virus P22* (accession: BK000582). No sequence differences were found between the BTP1 prophage and phage sequences (data not shown). The annotated genome and sequencing reads for the BTP1 phage have been deposited at the DDBJ/EMBL/GenBank database under the accession LT714109 (URL: http://www.ebi.ac.uk/ena/data/view/LT714109) and study accession PRJEB18919. The proposed species *Salmonella* virus BTP1 has been assigned the taxon ID 1934252.

### 3.3.4 Spontaneous induction of prophage BTP1

BTP1 was the only phage that could be recovered from the D23580 culture supernatant using plaque assay methodology (3.2.1) and the uninduced culture yielded a titre of ~$10^9$ PFU/ml which was increased to ~$10^{10}$ PFU/ml with the addition of mitomycin C (Table 3.2A). The titer of BTP1 in D23580 culture supernatants was 10,000-fold greater than the titer of P22 in the supernatant of D23580 ΔΦ [P22] (Figure 3.9A). To determine whether the unusually high level of spontaneous induction (Table 3.2) is an intrinsic characteristic of the BTP1 prophage, a prophage-cured derivative of *Salmonella* Typhimurium LT2 (MA8505) was lysogenized with BTP1, and compared with LT2 lysogenized with P22. The phage titer of non-induced culture supernatants was determined (Figure 3.9D), showing that a high frequency of spontaneous induction is an inherent characteristic of phage BTP1, independent of the D23580 host. The burst size of BTP1 in D23580 was around 600 (Figure 3.9B), resembling the burst size of phage P22 (Aramli and Teschke, 1999). The titer of $10^9$ PFU/ml in the D23580 non-induced 16 h culture (Table 3.2A) suggests that the fraction of the population undergoing phage-mediated lysis is approximately 0.2%. In LT2 [BTP1], the fraction of the population undergoing phage-mediated lysis is approximately 0.02% (this calculation is explained in Chapter 2.4.5).

When measured over the course of a growth curve, it was found that BTP1 spontaneous induction was not phase-specific. Instead, induction occurred constitutively so that the increase in PFU/ml was approximately proportionate to the increase in CFU/ml over the growth curve (Figure 3.9C). To determine if the high-level

of BTP1 spontaneous induction was dependent on the SOS-response, the spontaneously induced phage titer in a RecA$^-$ BTP1 lysogen was measured, using a RecA$^-$ P22 lysogen as a comparator (Figure 3.9D). In host strain LT2 (MA8505), P22 spontaneous induction is virtually abolished in a Δ*recA* background, and there is a 3-log reduction to the order of only 10 PFU/ml. In a LT2 (MA8505) lysogen, BTP1 spontaneous induction was also dramatically reduced in the Δ*recA* background by 5-log but remained greater than that of P22 at around 500 PFU/ml. Considering that LT2 is not the native host of BTP1, the role of RecA on spontaneous induction of BTP1 in D23580 was also assessed, and a 4-log reduction was observed. These results show that the majority of BTP1 spontaneous induction is RecA-dependent, and likely induced by the SOS-response. However, there is still considerable RecA-independent induction of the BTP1 prophage, and notably, there is 1 log more RecA-independent induction in the native D23580 background than in an LT2 (MA8505) background.

To understand how unusual the spontaneous induction of the BTP1 prophage was in context of other characterised phages, a review of published spontaneously-induced phage titers was conducted (Table 3.3). The highest reported spontaneously-induced phage titers were in the order of $10^7$ PFU/ml, for prophages ϕadh and φLC3 in *Lactobacillus gasseri* and *Lactococcus lactis*, respectively. Consequently, it was concluded that BTP1 exhibits the highest spontaneously-induced phage titer of any published prophage/host system.

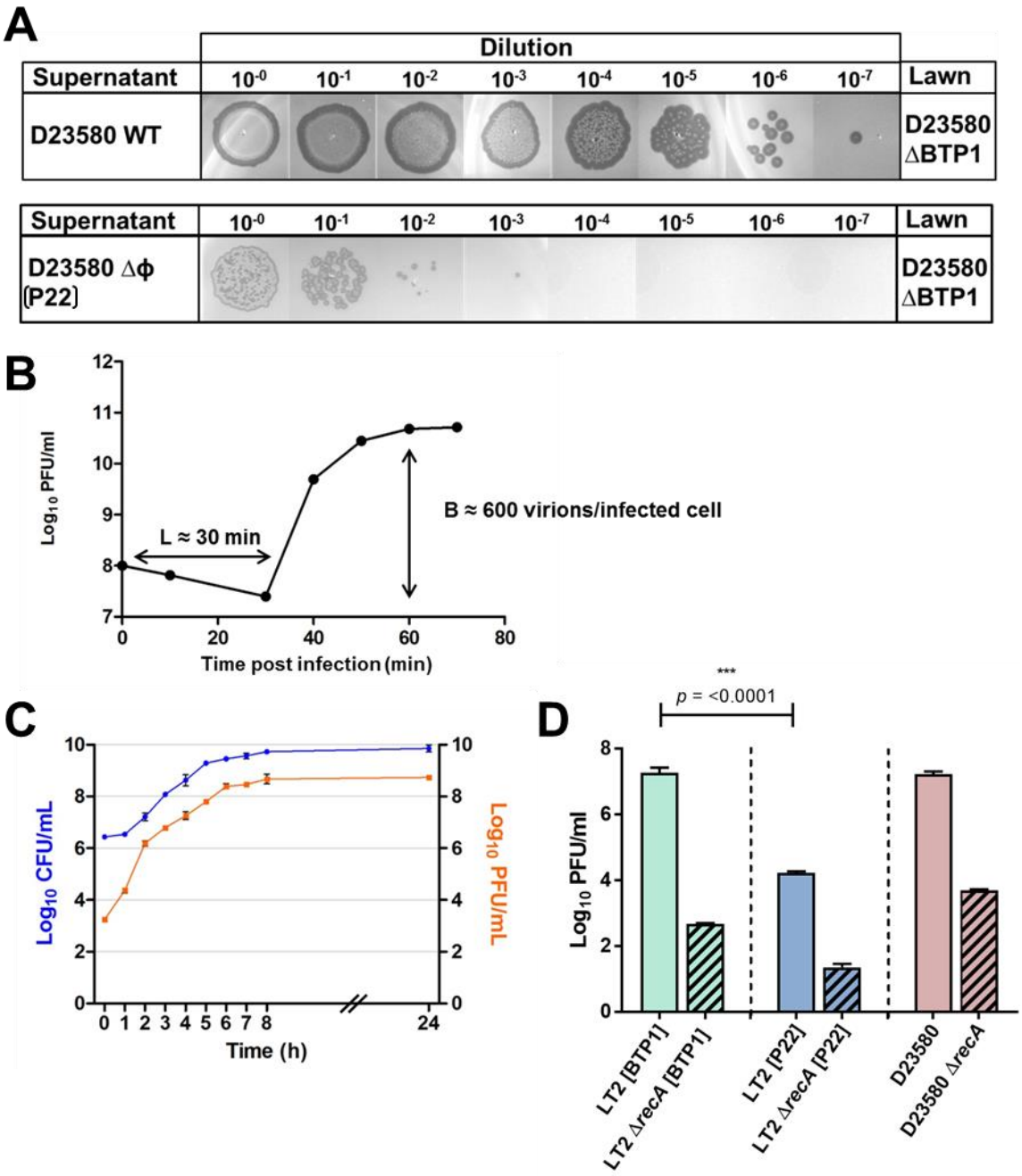



**Figure 3.9 BTP1 spontaneous induction and growth parameters.** A. Visualisation of BTP1 PFU/ml in the overnight supernatant of a D23580 culture (Chapter 2.3.1), compared to a P22 lysogen. B. Single step growth curve of BTP1. 'B' indicates burst size and 'L' indicates latent period. Burst size was calculated as the total number of virions produced in the burst divided by the number of infected bacterial cells. C. Growth curve of *S.* Typhimurium D23580 showing CFU/ml on the left y-axis and PFU/ml on the right y-axis indicating BTP1 titer. D. Spontaneous induction of BTP1 and P22 in a RecA$^-$ background. RecA-independent spontaneously induced phage titers are shown in strain LT2 for P22, and strains LT2 and D23580 for BTP1.

**Table 3.3 Overview of the order of magnitude of spontaneous induced phage titers from published literature and this study.** PFU/ml values are ordered by smallest to largest. Values resulting from the results of this thesis are shown in red text.

| Prophage | Resident host | PFU/ml | Reference |
|---|---|---|---|
| SPO2 | *Bacillus subtilis* | $10^4$ | (Garro and Law, 1974) |
| ST64B | *Salmonella* Typhimurium | $10^4$ | (Figueroa-Bossi and Bossi, 2004) |
| P1 | *Enterococcus faecalis* | $10^{4.5}$ | (Matos et al., 2013) |
| P22 | *Salmonella* Typhimurium | $10^5$ | This study |
| MuSo2 | *Shewanella oneidensis* | $10^5$ | (Gödeke et al., 2011) |
| ϕ105 | *Bacillus subtilis* | $10^5$ | (Garro and Law, 1974) |
| ϕMMP02 | *Clostridium difficile* | $10^{5.5}$ | (Meessen-Pinard et al., 2012) |
| ϕMMP04 | *Clostridium difficile* | $10^{5.5}$ | (Meessen-Pinard et al., 2012) |
| Mu | *Escherichia coli* | $10^6$ | (Howe and Bade, 1975) |
| SV1 | *Streptococcus pneumoniae* | $10^6$ | (Carrolo et al., 2010) |
| ϕadh | *Lactobacillus gasseri* | $10^7$ | (Baugher et al., 2014) |
| φLC3 | *Lactococcus lactis* | $10^7$ | (Lunde et al., 2003) |
| BTP1 | *Salmonella* Typhimurium | $10^9$ | This study |

### 3.3.4.1 Spontaneous induction of BTP1 at the single-cell level

Studying spontaneous prophage induction by measuring the titer of phages in culture supernatants is a simple and effective method that has been employed for decades (Lwoff, 1953). However, the titer of induced phage is an indirect measurement of spontaneous induction, as the result relies on the secondary infection of an indicator strain by the induced phages, rather than the induction event itself. For instance, defective prophages may be inducible and capable of cell lysis, but their induction would be missed by phage-titering based methods as the resulting phage-particles are uninfectious. In order to better understand the prolific spontaneous induction phenotype of the BTP1 prophage, a fluorescence reporter system was developed to allow study of BTP1 induction at the single-cell level.

As BTP1 is a lambda-like prophage, the transcriptional architecture of prophage induction could be predicted based on the Lambda molecular switch paradigm (summarised in Figure 1.8). Upon induction, loss of repressor function putatively allows transcription to initiate from the promoters $P_L$, upstream of the gene encoding for the *n* antitermination gene (STMMW_03521), and $P_R$, upstream of the gene encoding for the *cro* regulatory gene (STMMW_03551). Theoretically it would be possible to use a plasmid-based system involving a fluorescent marker under the control of the $P_L$ or $P_R$ promoters in order to visualise prophage induction. However, plasmid-based systems are highly likely to affect prophage induction dynamics, as multiple copies of the $P_L$ or $P_R$ promoters would titrate the repressor protein away from the native promoter. Additionally, as a number of tightly controlled molecular events have to occur following the initiation of prophage induction by transcription at $P_L$ or $P_R$, this event in itself is not necessarily indicative of successfully completed prophage induction. In light of these drawbacks, a plasmid-based system was not adopted to study the induction of BTP1 and instead, a chromosomal system was developed.

After initial transcription from $P_L$ and $P_R$, it is predicted based on studies in Lambda that the N protein antiterminates both the *n* and *cro* transcripts to allow transcription of the 'early' genes, from *n* to *int* (STMMW_03521 to STMMW_03341) and from *cro* to *q* (STMMW_03551 to STMMW_03691). Subsequently, the Q protein is thought to antiterminate a small transcript (pR') occurring prior to the lysis genes. In the BTP1 prophage, there are 2 tRNA genes between the *q* gene and the lysis genes, therefore it is unclear how and where Q-mediated antitermination occurs. However, Q protein synthesis likely results in the transcription of the lysis and structural genes, known collectively as the phage 'late' genes (STMMW_03711 to STMMW_03901). The late

genes are the final genes transcribed in prophage induction, and were chosen as the best marker for prophage induction. Initially, a system was developed as shown in Figure 3.10. A *gfp* construct consisting of a promoter-less *gfp*$^+$ gene from the pZEP08 plasmid (Hautefort et al., 2003) and the FRT-flanked kanamycin resistance gene from the pKD4 plasmid (Datsenko and Wanner, 2000) was designed (Figure 3.10B). The construct was amplified to contain 50 bp flanking ends corresponding to an intergenic position in the late genes (Chapter 2.6.4), recombined into the chromosome using the lambda-red recombineering methodology (Datsenko and Wanner, 2000), resulting in transcription of *gfp*$^+$ with the late gene operon (Figure 3.10A). Importantly, the kanamycin gene was orientated in the opposite direction to the *gfp*$^+$ to prevent the aberrant transcription of late genes downstream of the construct in non-induced cells. Once the construct was inserted into the chromosome the kanamycin gene was removed by FLP-mediated recombination of the FRT sites. Description of the methods used to construct the reporter systems can be found in Chapter 2.6.4.

Though the system functioned correctly, and fluorescent cells were visualised upon chemical induction, the fluorescence signal was transient and lost around 1 hour after chemical induction (data not shown). It is likely that the loss of fluorescence signal shortly after induction is the result of cell lysis and phage release at the end of the induction process (schematic shown in Figure 3.11). Consequently, this construct could only be used to measure the number of cells undergoing induction at any one time, rather than to measure the total number of cells in the population that have spontaneously induced.

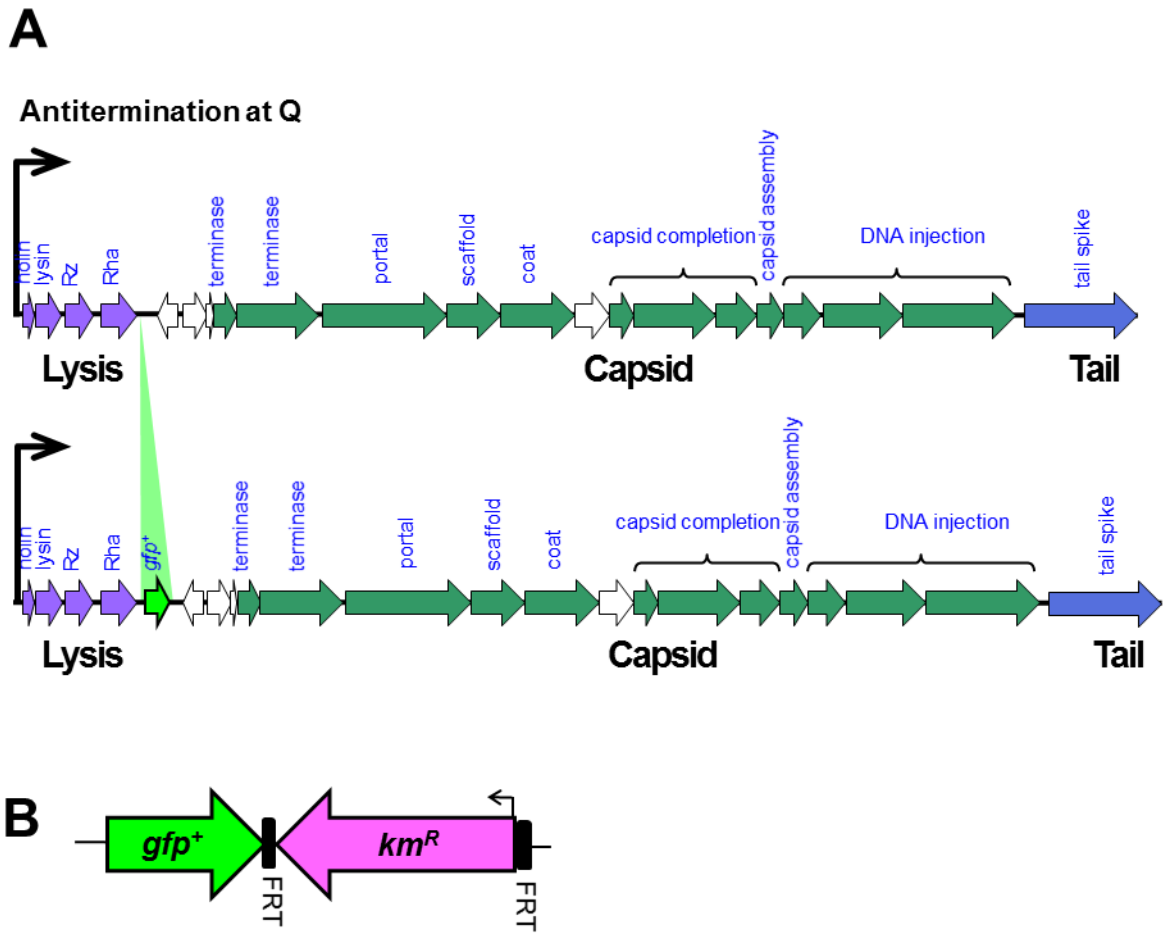



**Figure 3.10 Illustration of initial GFP reporter system for BTP1 induction.** A. Genetic context of the *gfp*$^{+}$ in the final construction. B. Architecture of the *gfp*+ kanamycin resistance construct used to create the florescence reporter system on the chromosome of D23580. Methods used to develop the construct can be found in Chapter 2.6.4.

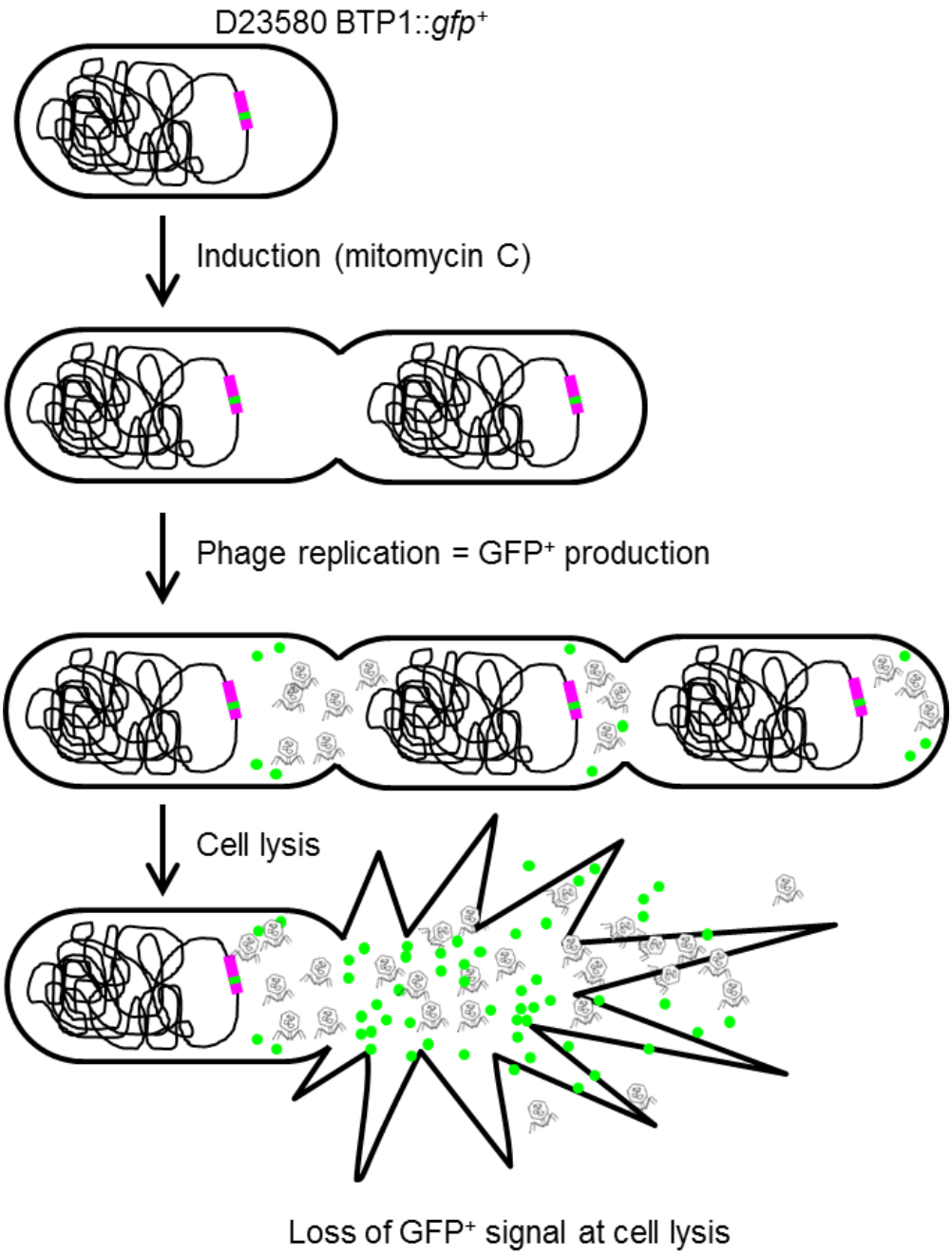



**Figure 3.11 Schematic illustration of the expected fate of induced cells containing the initial GFP reporter construct.**

A second system was developed using the same construct, but instead of using an intergenic position, the entire lysis gene cluster was replaced with the *gfp*$^+$ construct (Figure 3.12). As the synthesis of the lysis proteins are likely to be essential for cell lysis, loss of these genes would allow induced cells to remain fluorescent in the culture, without loss of signal (schematic shown in Figure 3.13). A time-course of chemical induction of the D23580 ∆lysis::*gfp*$^+$ strain is shown in Figure 3.14. This system was used to study the natural spontaneous induction of D23580 cultures.

Overnight cultures of D23580 ∆lysis::*gfp*$^+$ were screened for fluorescent cells. As expected, the majority of the population were non-fluorescent, short rods, resembling healthy stationary phase *Salmonella* bacteria. However, it was also possible to find highly fluorescent cells in the culture (Figure 3.15). The fluorescent cells in the overnight culture of D23580 ∆lysis::*gfp*$^+$ were observed to be highly heterologous in size; some resembled the non-fluorescent bacteria in size, whilst others resembled large filaments (an example is shown in Figure 3.15). As it was previously observed that BTP1 spontaneous induction occurs throughout the growth curve, this dimorphism in size may correspond to the growth-phase the cell was in when it became induced; exponentially growing cells that become induced may form large filaments, whilst cells that become induced towards stationary phase may remain as single cells.

The visualisation of fluorescent cells in the overnight culture of the D23580 ∆lysis::*gfp*$^+$ shows that this approach can be used to study the cell population in which the BTP1 prophage becomes spontaneously induced, without subsequent cell lysis. In the future these constructs could be used to accurately enumerate the induced fraction of the population using flow cytometry techniques, though optimisation will be required to control for the impact of heterogeneity of fluorescent cell sizes upon cell detection and sorting. In addition, similar constructs could be made in other prophages, allowing comparison of the rate of spontaneous induction of the lysogenic population, and the factors that contribute to spontaneous prophage induction.

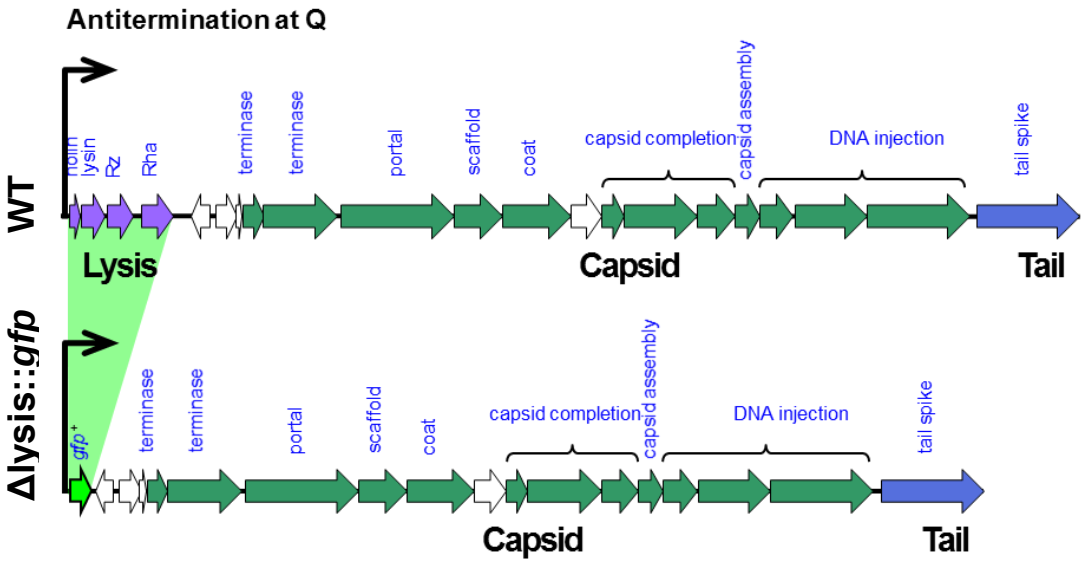



**Figure 3.12 Illustration of the Δlysis::*gfp*⁺ construct for BTP1 induction.** Genetic context of the *gfp*⁺ in the final construction. The *gfp*⁺ gene replaces the entirety of the lysis gene cluster.

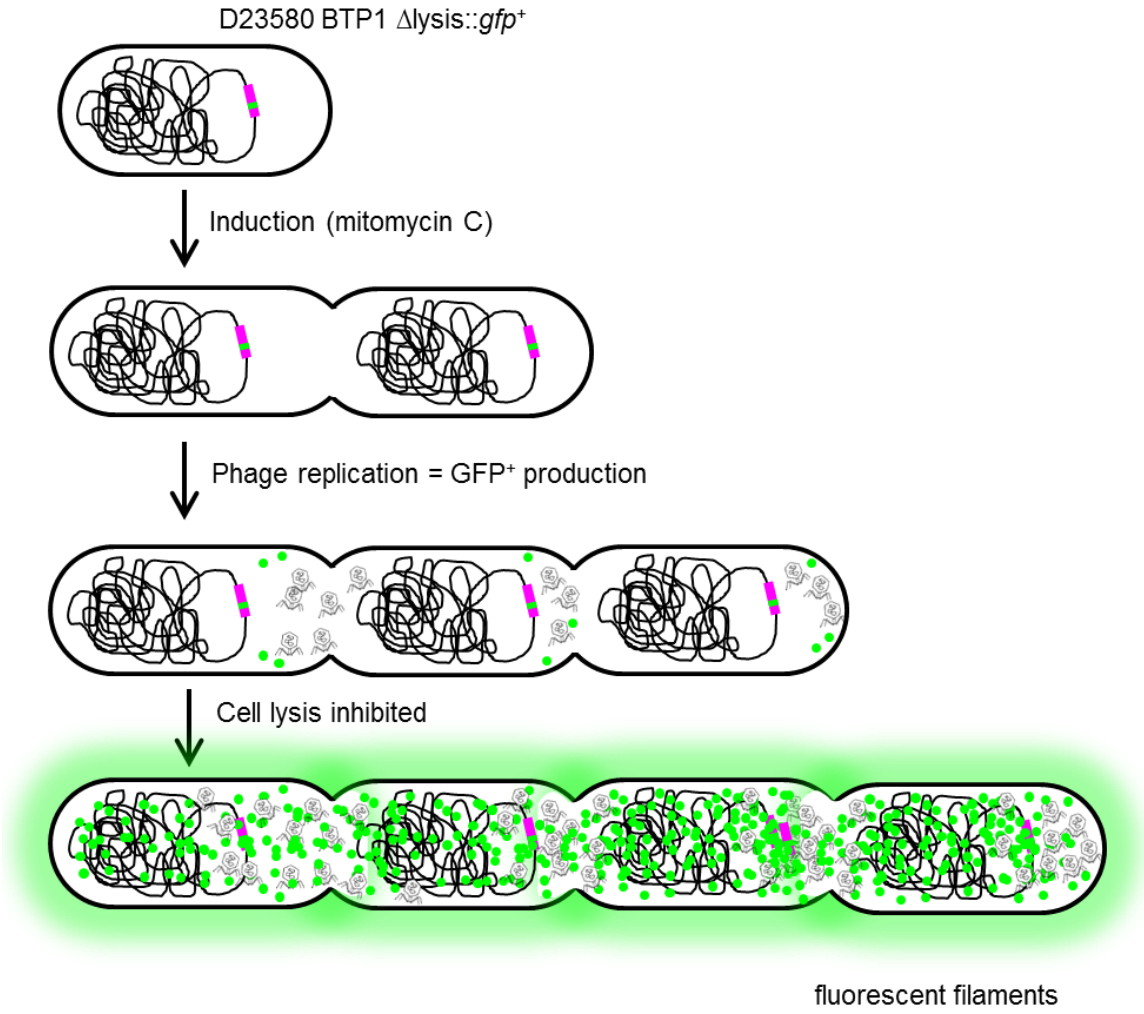



**Figure 3.13 Schematic illustration of the expected fate of induced cells containing the Δlysis::*gfp*$^+$ reporter construct.**

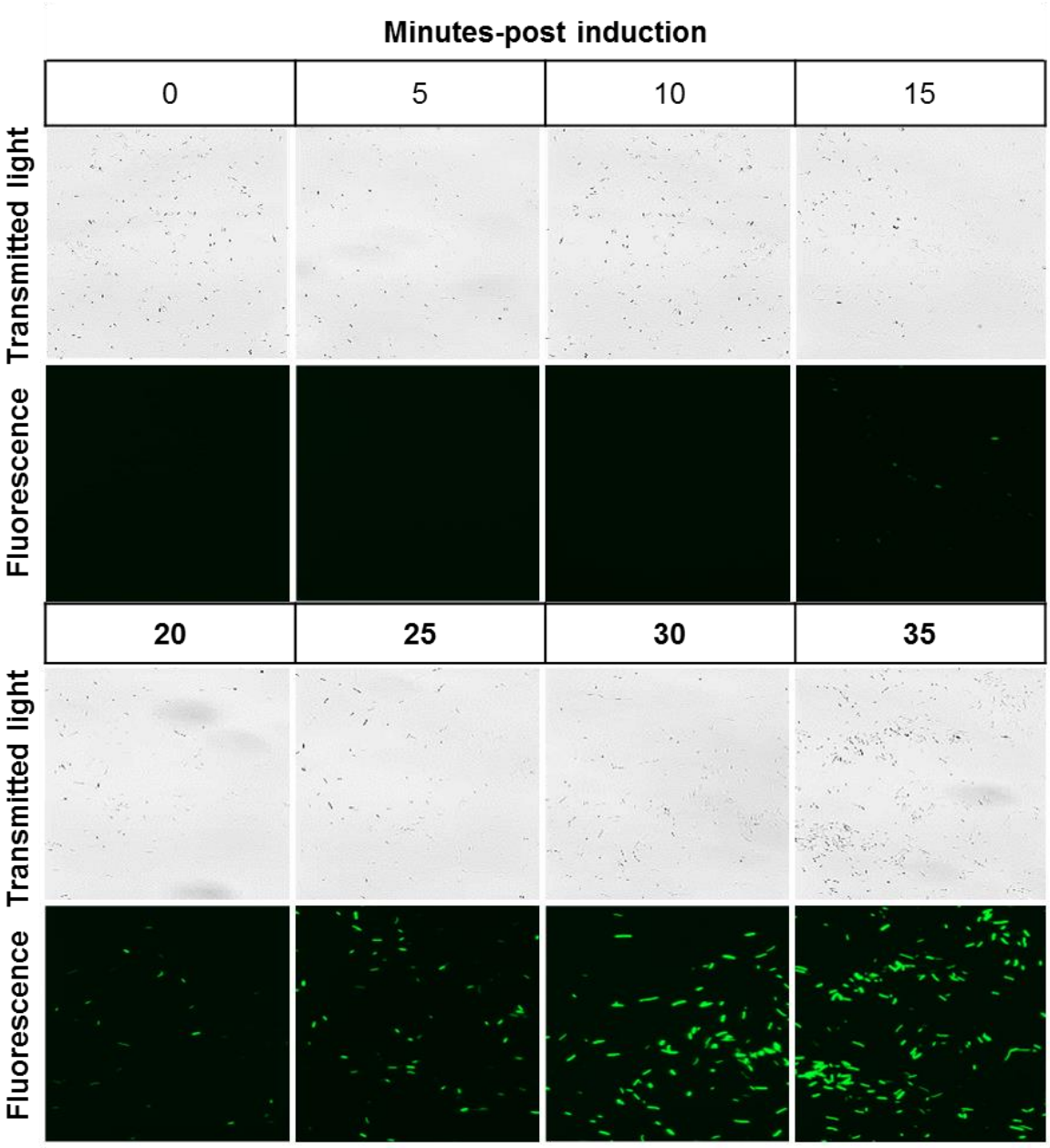



**Figure 3.14 Time course of D23580 BTP1 ∆lysis::*gfp*$^{+}$ cell fluorescence post chemical induction.** Samples were collected and fixed at 5 minute intervals after addition of mitomycin C (Chapter 2.3.6). Transmitted light and fluorescent light images of the same field visualised using a 40X objective are shown (Chapter 2.3.6).

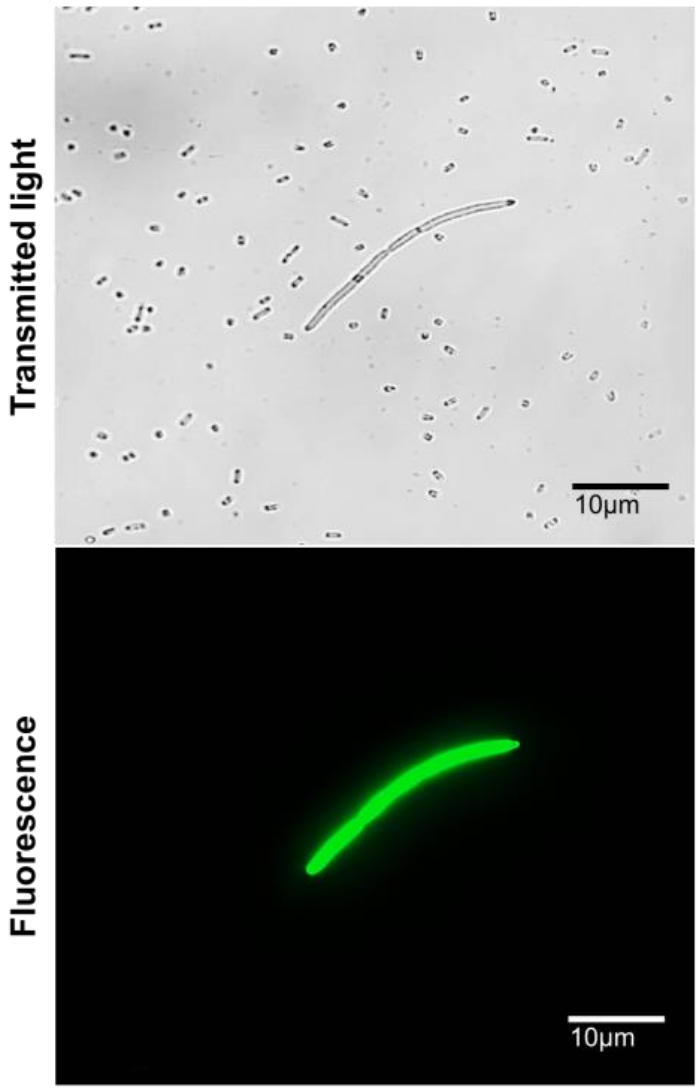



**Figure 3.15 Example of elongated fluorescent cell morphology found in overnight culture of D23580 ∆lysis::*gfp*$^{+}$.** Both transmitted light and fluorescent light images of the same field using a 100X oil immersion objective are shown. Microscopy methods can be found in Chapter 2.3.6.

## 3.3 Gifsy-2 (BTP2)

The Gifsy prophages are lambdoid prophages present in the majority of *S.* Typhimurium strains (Lemire et al., 2008). However, of the Gifsy-family of prophages, Gifsy-2 is the most widely distributed and can be found in a number of *Salmonella* serovars apart from *S.* Typhimurium (Lemire et al., 2008; Thomson et al., 2004). Gifsy-2 is required for *S*. Typhimurium virulence in the mouse infection model, and a Gifsy-2 cured strain of *S.* Typhimurium was found to be attenuated 100-fold relative to the wildtype (Ho et al., 2002). Gifsy-2 contains a number of virulence genes which include genes encoding for the periplasmic superoxide dismutase SodCI and the secreted effector proteins GtgE and SseI (Bossi and Figueroa-Bossi, 2005) (Figure 3.16).

The Gifsy-2 prophage of D23580 is 98% identical to the Gifsy-2 prophage of ST19 strain 4/74 at the nucleotide level (data not shown). Gifsy-2 is located in the same intergenic chromosomal site in D23580 and 4/74, between the *pncB* and *pepN* genes. Notably, the prophage is orientated in the opposite direction in D23580 relative to 4/74, suggesting that the prophage has not remained as a stable lysogen over the course of the evolution of these two strains, but has excised and reintegrated. My comparative analysis of the Gifsy-2 nucleotide sequence showed that the gene encoding the O replication protein of Gifsy-2$^{D23580}$ was a pseudogene; a 71 bp deletion in the middle of the coding sequence caused a frame shift at position 1,102,739 on the D23580 chromosome, and the generation of a truncated O protein (Figure 3.17). The O protein of lambdoid phages is required for replication of the phage chromosome during lytic growth and contains the phage origin of replication (Taylor and Wegrzyn, 1995). As the replication of the phage chromosome is an essential part of the lytic cycle, this pseudogene is likely to be responsible for the lack of functional Gifsy-2 $^{D23580}$ viruses detectable in the supernatant when D23580 is grown with and without mitomycin C.

Additionally, a putative tail fibre gene (STMMW_10601) is ~ 200 bp longer in D23580 relative to the ortholog in 4/74, though analysis of more Gifsy-2 sequences would be required to determine if this represents an insertion event in Gifsy-2 $^{D23580}$ or a deletion event in Gifsy-2 $^{4/74}$. Lastly, the *sseI* gene of Gifsy-2 $^{D23580}$ (STMMW_10631) has been disrupted by the insertion of a transposase gene (STMMW_10641), which has caused the loss of function of the SseI effector protein in D23580 (Carden et al., 2017).

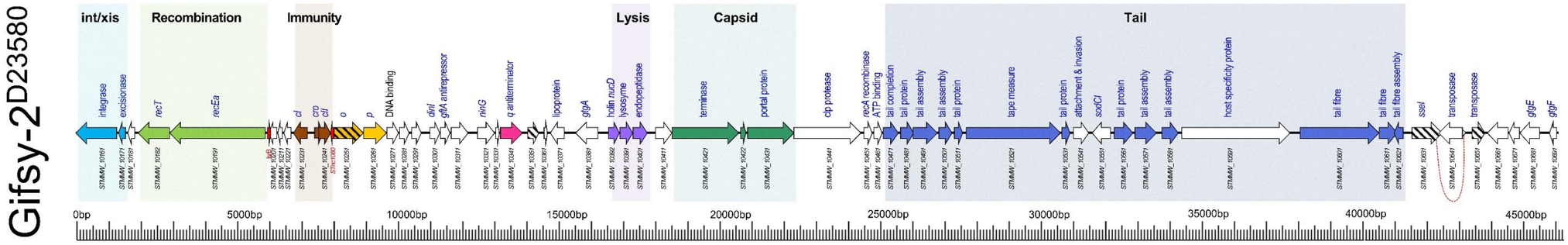



**Figure 3.16 Functionally annotated map of the Gifsy-2$^{D23580}$ prophage sequence.** Coloured arrows represent ORFs for which a function can be inferred based on sequence. White arrows represent ORFs for which no function can be inferred. Arrows filled with slanted black stripes represent pseudogenes. Red dotted lines connect the separated parts of pseudogenised genes which have been disrupted by early truncation or insertion.

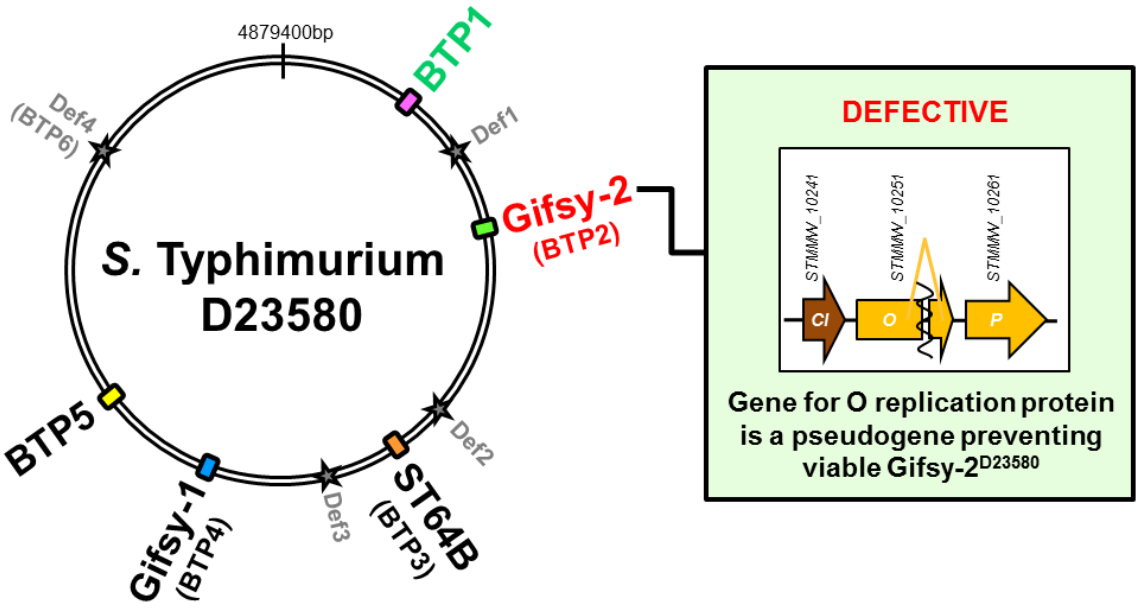



**Figure 3.17 Gifsy-2$^{D23580}$ is likely to be defective due to a frame shift mutation in the *o* replication gene.**

## 3.4 ST64B (BTP3)

ST64B is lambdoid siphovirus present in a number of *Salmonella* genomes, in particular those of epidemiological relevance, such as the multidrug resistant DT104 clone (Tucker and Heuzenroeder, 2004). Though the prophage was not required for *Salmonella* virulence in a murine model (Bossi and Figueroa-Bossi, 2005.), ST64B harbours the *sseK3* gene (Figure 3.18) which has been shown to be an important secreted effector protein modulating the host immune response (Günster et al., 2017; Yang et al., 2015).

ST64B$^{D23580}$ is almost 100% identical to ST64B$^{4/74}$, with just 6 SNPs separating the two prophages (data not shown). ST64B is located in the same position and orientation in D23580 and 4/74, within tRNA$^{Ser}$ gene. ST64B is defective in *S.* Typhimurium 14028, due to a frame-shift mutation in a tail assembly gene that prevents assembly of viable phage particles, and creates two short gene fragments annotated *sb21* and *sb22* (Figueroa-Bossi and Bossi, 2004). The tail gene frame-shift mutation is conserved in ST64B$^{D23580}$ (D23580 genes *STMMW_19871* and *STMMW_19861* are the homologs of *sb21* and *sb22,* respectively). The frameshift mutation that prevents viable ST64B phage production in ST19 strain 14028 can spontaneously revert to a functional form, particularly under conditions that positively select for resuscitation of the functional prophage, such as when a strain containing the defective prophage is co-cultured with a sensitive host (Figueroa-Bossi and Bossi, 2004).

Together with the failure of D23580 supernatant to produce plaques on a ST64B-sensitive host strain, and the conservation of the frame-shift mutation of strain 14028, these results indicate that ST64B is defective in *S.* Typhimurium ST313 (Figure 3.19).

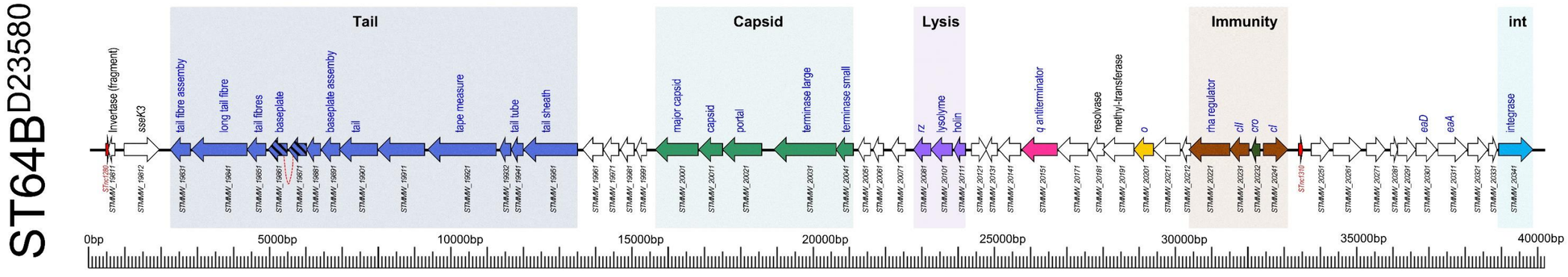



**Figure 3.18 Functionally annotated map of the ST64B$^{D23580}$ prophage sequence.** Coloured arrows represent ORFs for which a function can be inferred based on sequence. White arrows represent ORFs for which no function can be inferred. Arrows filled with slanted black stripes represent pseudogenes. Red dotted lines connect the separated parts of pseudogenised genes which have been disrupted by early truncation or insertion.

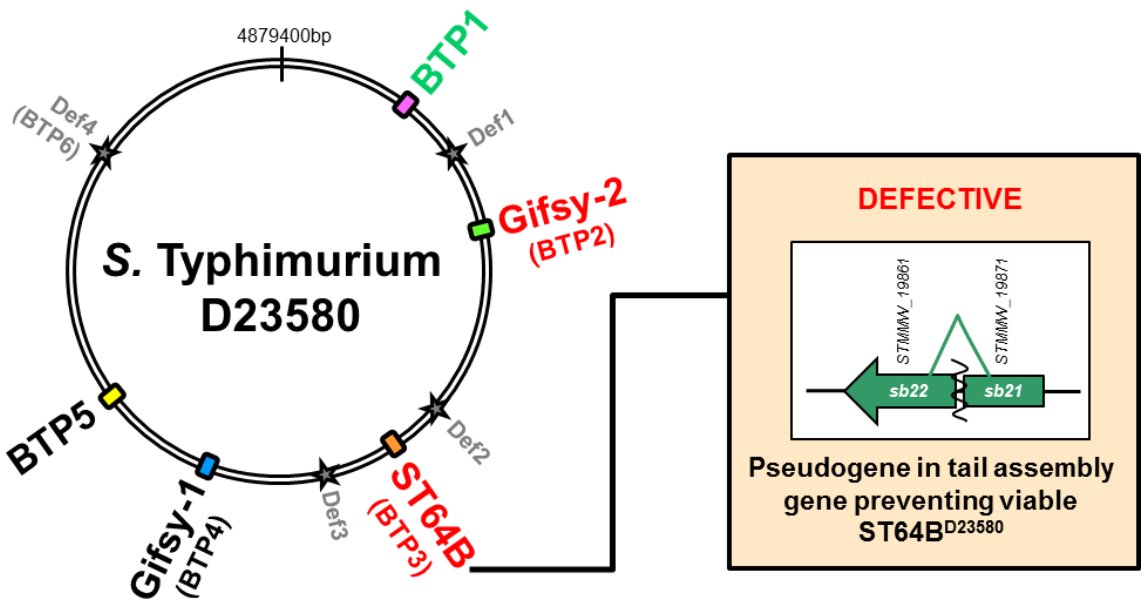



**Figure 3.19 ST64B$^{D23580}$ is likely to be defective due to a frame shift mutation in a tail assembly gene.**

## 3.5 Gifsy-1 (BTP4)

Gifsy-1 is a lambdoid siphovirus like its relative Gifsy-2 (Chapter 3.3). Gifsy-1$^{D23580}$ carries the important virulence factors *gipA* (*STMMW_26192*), *gogB* (*STMMW_26001*, and *gtgA* (*STMMW_26331*) (Figure 3.20) (Coombes et al., 2005; Moreno Switt et al., 2012). The chromosomal attachment site for Gifsy-1 is the *Salmonella lepA* gene (STM2583, STMMW_25991). The Gifsy-1 phage genome contains a small piece of the *lepA* gene which reconstitutes the intact *lepA* locus upon prophage integration (Casjens, 2011). Gifsy-1 is known to be more genetically diverse than Gifsy-2 (Hiley et al., 2014). Whilst the location and attachment site of Gifsy-1 is conserved in D23580, there is considerable genetic divergence relative to Gifsy-1$^{4/74}$ in a ~5kb region towards the right terminus of the prophage sequence (Figure 3.21) which includes the replication and immunity genes as well as a number of ORFs of unknown function. However, the function of Gifsy-1 in comparator strain 4/74 has not been defined, so the sequence of Gifsy-1$^{14028}$ (a strain where Gifsy-1 is known to be functional) was used as a comparator to identify potential degradative mutations (Figueroa-Bossi and Bossi, 1999).

All coding genes remain intact in Gifsy-1$^{D23580}$ compared to Gifsy-1$^{14028}$, with no evidence for functional degradation *via* pseudogene formation. However, a single SNP was detected in the -10 promoter region of the *dinI* gene of Gifsy-1$^{D23580}$ compared to the same region in *S.* Typhimurium 14028 (Gifsy-1$^{14028}$), where Gifsy-1 is functional (Bossi et al., 2003) (Figure 3.22A). It has previously been shown that the anti-repressor gene of Gifsy-1, *gfoA*, is located immediately downstream of *dinI* and is co-transcribed from the *dinI* promoter ($P_{dinI\text{-}gfoA}$), making *gfoA* essential for the induction of Gifsy-1 phage (Lemire et al., 2011). This finding raised the possibility that the promoter SNP could be responsible for the lack of detectable Gifsy-1 in the D23580 supernatant (Table 3.2). Single nucleotide allelic exchange of $P_{dinI\text{-}gfoA}^{D23580}$ with $P_{dinI\text{-}gfoA}^{14028}$ yielded viable (infectious) Gifsy-1 phages that formed plaques on D23580 ΔGifsy-1. Furthermore, D23580 $P_{dinI\text{-}gfoA}^{14028}$ produced a Gifsy-1 titer consistent with that reported for 14028, both with and without mitomycin C induction (Figueroa-Bossi and Bossi, 2004) (Table 3.2A; Figure 3.22B). The fact that a single modified nucleotide caused the production of functional phage shows that the $P_{dinI\text{-}gfoA}$ SNP is responsible for lack of Gifsy-1 viability in D23580.

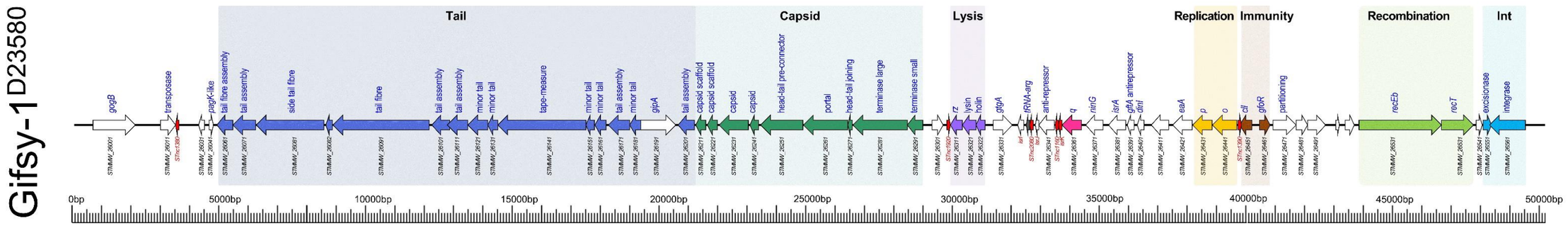



**Figure 3.20 Functionally annotated map of the Gifsy-1$^{D23580}$ prophage sequence.** Coloured arrows represent ORFs for which a function can be inferred based on sequence. White arrows represent ORFs for which no function can be inferred.

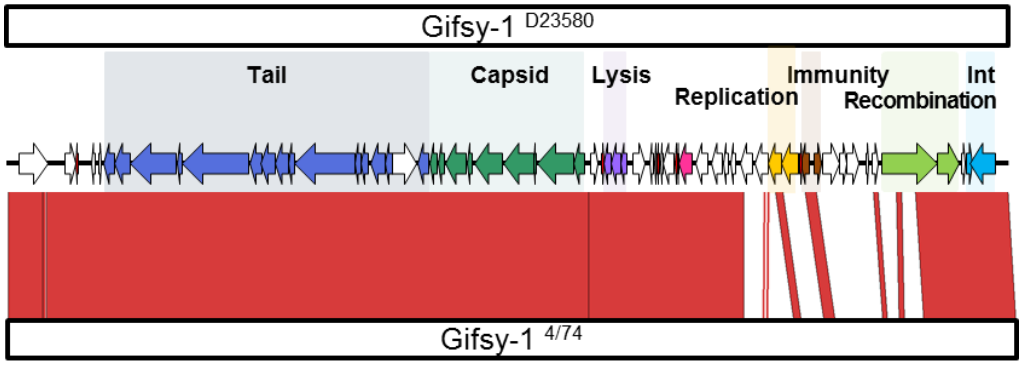



**Figure 3.21 Comparison between the Gifsy-1 prophage of strains D23580 and 4/74.** Red coloured blocks indicate individual blastn hits of more than 90% identity. White regions show represent sequence that is unique to either genome.

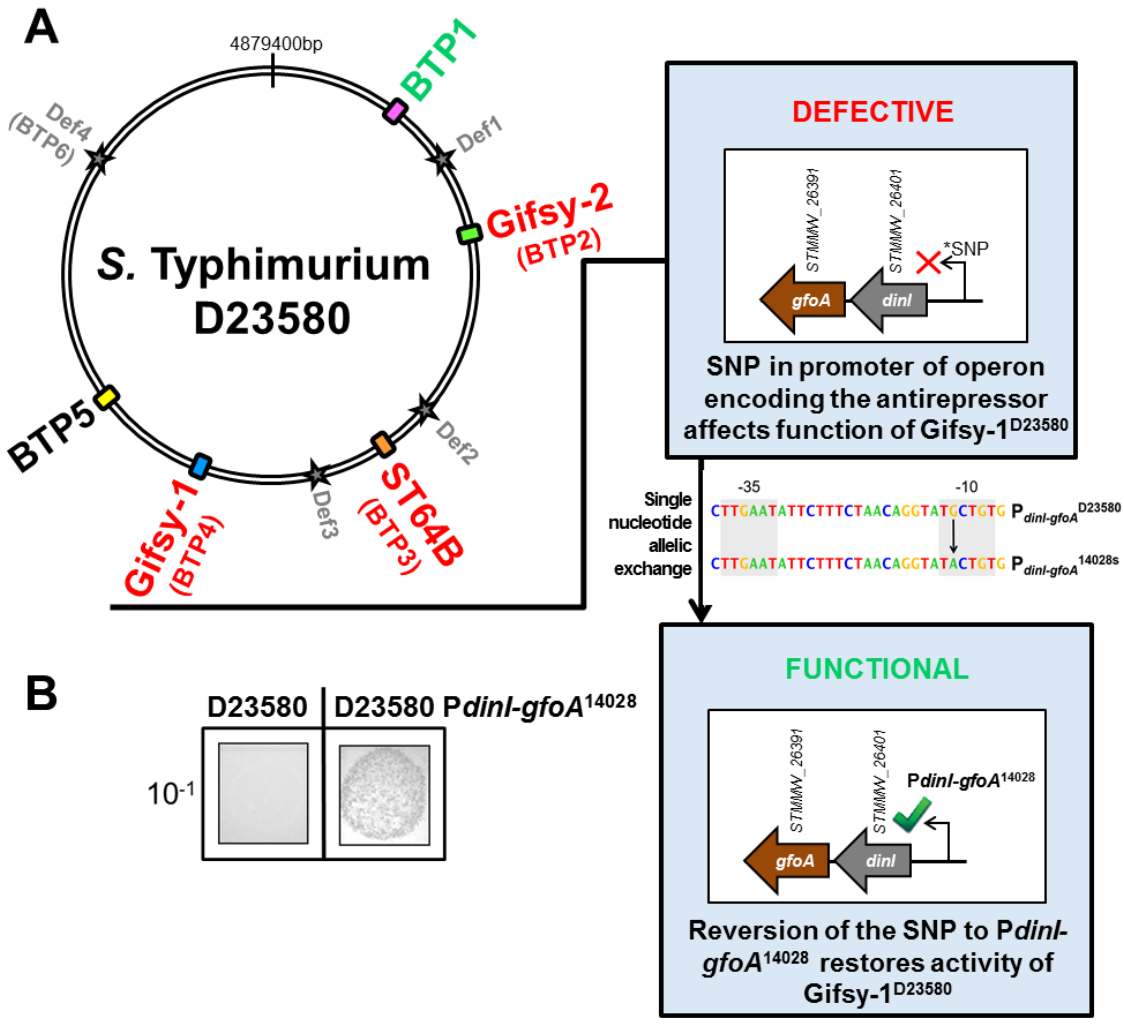



**Figure 3.22 Gifsy-1$^{D23580}$ is defective due to a single SNP in the promoter of the antirepressor gene.** A. A SNP in the -10 region of the *dinI-gfoA* promoter is likely to abolish transcription of these genes. B. Restoration of the SNP to the genotype found in Gifsy-1$^{14028}$ restored the formation of functional phages by Gifsy-1. Methods used for the single nucleotide allelic exchange are found in Chapter 2.6.2.

## 3.6 BTP5

### 3.6.1 Gene content of BTP5

The BTP5 structural genes suggest that the phage is a member of the Myoviridae family of long contractile-tailed phages and it shares sequence similarity with P2-like phages including P2, WΦ, and 186 which infect *E. coli*. P2 has been described as the prototype for the non-inducible class of temperate phages, as its repressor protein lacks an identified RecA cleavage site (Kalionis et al., 1986), and P2-like phages are known to be associated with satellite phages (Nilsson and Haggård-Ljungquist, 2006). However, BTP5 contains two genes that show homology to the SOS operon of coliphage 186 (Brumby et al., 1996): *tum* (STMMW_32041) and *orf97* (STMMW_32031), which function as an antirepressor system and facilitate induction of coliphage 186 (Shearwin et al., 1998). No function could be inferred based on sequence homology to other studied phages for 7/42 of BTP5 open reading frames (ORFs). The complete functionally annotated map of BTP5 is shown in Figure 3.23. The chromosomal context of the BTP5 attachment site (tRNA-Met) and the core *att* sequence for BTP5 are shown in Figure 3.24.

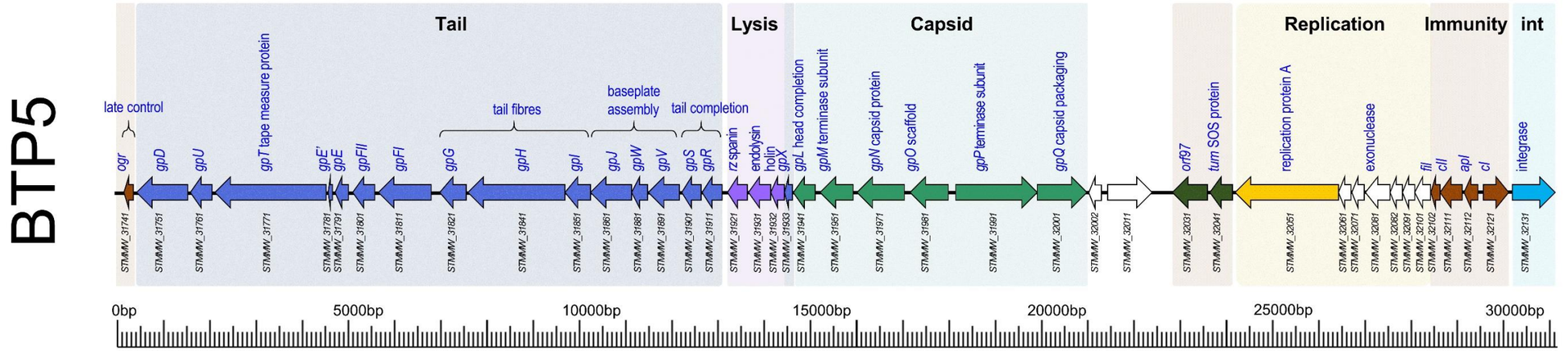
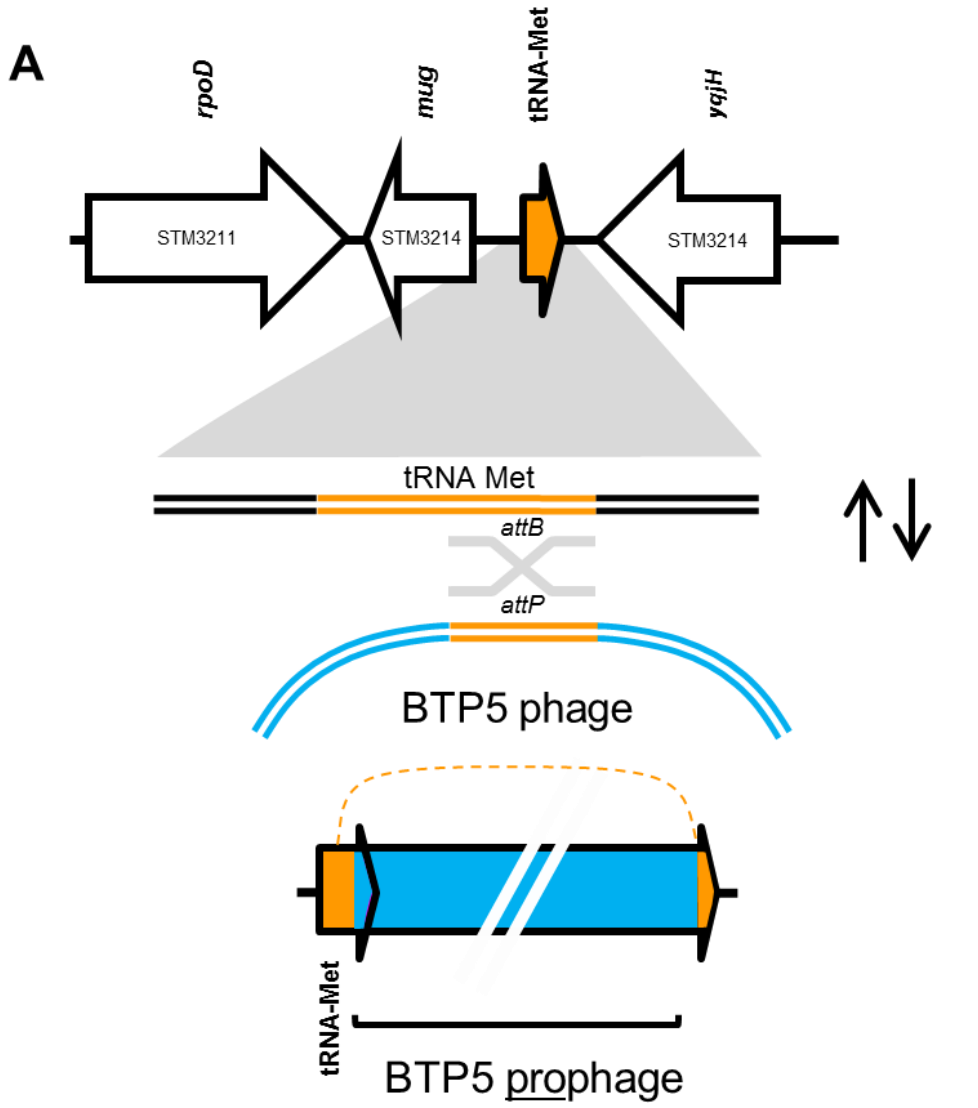



**Figure 3.23 Functionally annotated map of the BTP5$^{D23580}$ prophage sequence.** Coloured arrows represent ORFs for which a function can be inferred based on sequence. White arrows represent ORFs for which no function can be inferred.




B

| Site | Sequence |
|---|---|
| BTP5 *att* core | ACTCATAATCGCTTGGTCGCTGGTTCAAGTCCAGCAGGGGCCACCAAA<br>TTTTAGCTTTAAAATCATATAATTAAGCCACTCTA |

**Figure 3.24 Chromosomal genomic context and core sequence of the BTP5 attachment site.**

### 3.6.2 BTP5 homology to other sequenced phages and prophages

Comparison of BTP5 to other P2-like phages and prophages suggests that there is less genetic mosaicism than in the P22-like group described for the BTP1 prophage (Figure 3.25). The majority of the P2-like phages and prophages show conservation of nearly the entire capsid and tail gene clusters, which make up approximately 60% of the phage chromosome. Structural gene homology is not restricted to prophages of the genus *Salmonella*, as BTP5 shares >70% sequence identity to the structural genes of phages and prophages from other Enterobacteriaceae including *Escherichia, Klebsiella, Serratia* and *Enterobacter*. However, outside of the structural gene clusters, there is limited homology to other phages or prophages, suggesting that BTP5 may represent novel phage biology.

Analysis of publically available draft genome assemblies identified a putative BTP5-like prophage in a genome of a single strain of *S.* Typhi called 238675, that was isolated from blood in South Africa in 2008 and sequenced (short read accession number: ERR338011) as part of a study to describe the global population structure of *S.* Typhi (Wong et al., 2015). Strain 238675 has subsequently been classified as belonging to *S.* Typhi clade 2.4 (Wong et al., 2016). The prophage was located on a short contig of 31,288 bp, which spanned the entire length of the BTP5 prophage. As the prophage assembled as a whole contig, rather than as part of a larger chromosomal contig, it was not possible to ascertain from the genome assembly whether the attachment site of $BTP5^{S.Typhi\ 238675}$ is the same as the attachment site of $BTP5^{D23580}$. Alignment of the $BTP5^{S.Typhi\ 238675}$ with $BTP5^{D23580}$ showed that the prophages were 98.27% percent identical at the nucleotide level, a difference of 537 SNPs (Figure 3.26). The location of the SNPs within the genome relative to $BTP5^{D23580}$ was determined, and showed that the SNPs are restricted to the structural and lysis regions of the prophage (Figure 3.26). This finding raises the intriguing possibility that BTP5 is a broad-host range phage capable of infecting both *S.* Typhimurium and *S.* Typhi, or that some of the detected SNPs may be responsible for restricting the host range of the phage to serovar Typimurium or Typhi.

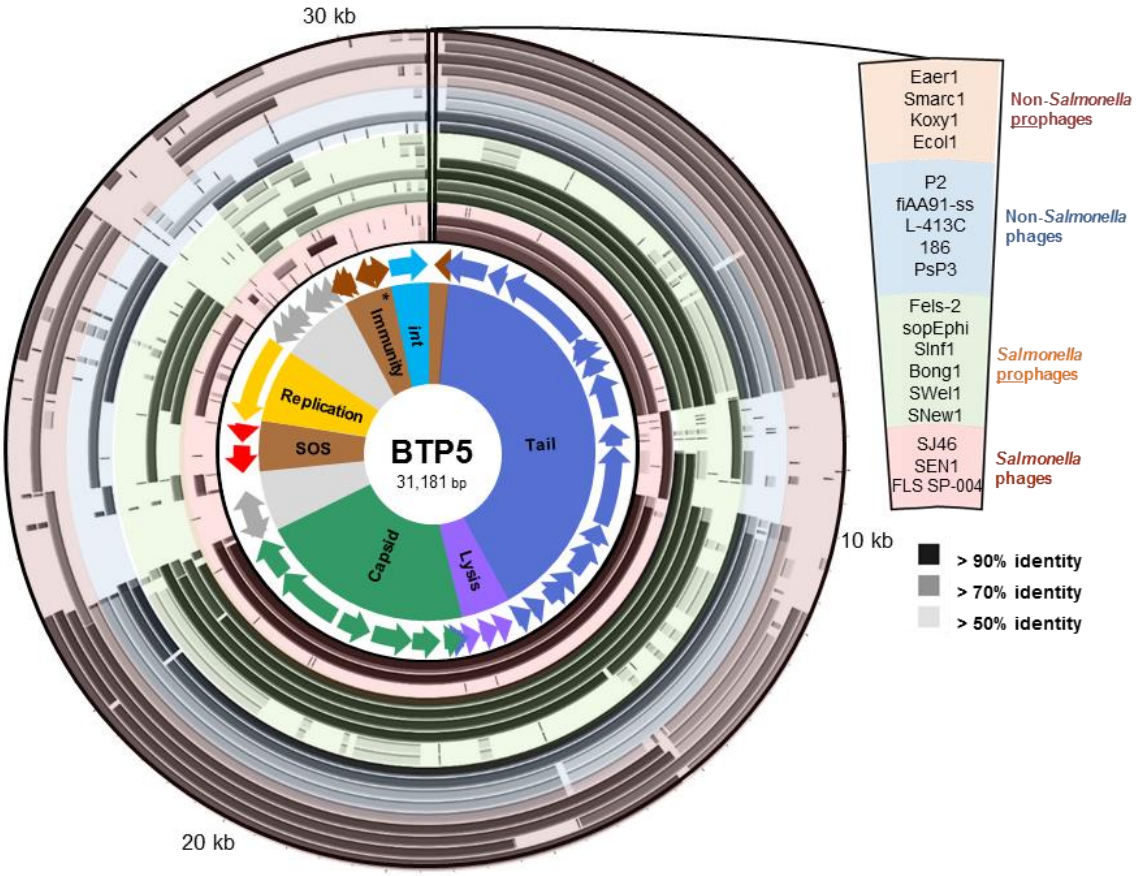



**Figure 3.25 Functional genetic architecture of BTP5 and homology to sequenced phages and prophages from *Salmonella* and other bacteria.** Starting from the inside, circular panels represent functional clusters of genes: ORFs coloured according to functional cluster with the functionally unknown ORFs colored grey; BLAST homology to other *Salmonella* phage sequences (red-shaded ring), *Salmonella* prophage sequences (green-shaded ring), non-*Salmonella* phages (blue-shaded ring) and non-*Salmonella* prophages (orange-shaded ring). The order of sequences represented by each concentric ring of the BLAST homology panel are shown on the right.

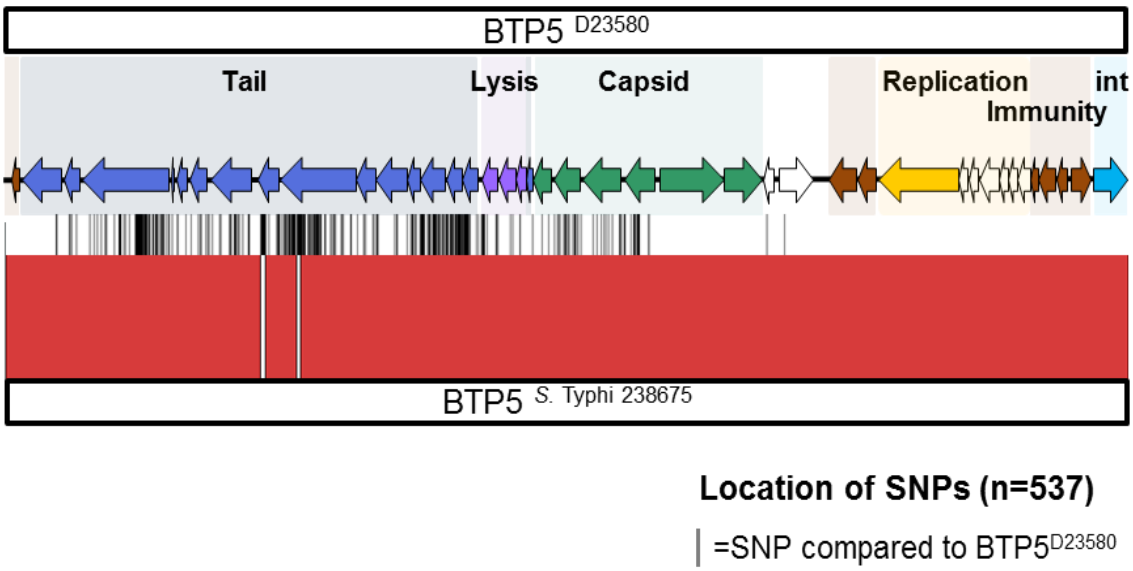



**Figure 3.26 Prophage BTP5 is found in *S.* Typhi strain 238675.** BTP5 is 98.27% identical at the nucleotide level to a prophage from ST568 strain 201102250, a difference of 537 SNPs. Red coloured blocks indicate individual BLAST hits of more than 90% identity. The location of the SNPs relative to BTP5$^{D23580}$ is shown by black lines at the top of the BLAST hits.

### 3.6.3 Functional assessment of the BTP5 prophage

Despite the presence of a very similar prophage in *S.* Typhi suggesting that BTP5 represents a functional, broad-host range phage, the supernatant from a 16 h culture of strain D23580 (Chapter 2.3.1) did not form plaques on the specific indicator strain D23580 ΔBTP5, even after mitomycin C exposure. Consequently, an alternative approach was used to determine whether putative BTP5 phages present in the culture supernatant were capable of injecting DNA into susceptible bacteria to form lysogens.

A D23580-derived strain that carried a kanamycin resistance gene (*Km*$^R$) inserted into a non-coding region of BTP5 was constructed (designated D23580 BTP5::*Km*$^R$) (methods described in Chapter 2.6.2). The presence of functional BTP5 phages in the culture supernatant was assayed by the formation of kanamycin-resistant BTP5 lysogens when culture supernatant was added to the putatively BTP5-sensitive strain D23580 ΔBTP5. The supernatant from a 16 h LB culture of D23580 BTP5::*Km*$^R$ yielded kanamycin-resistant colonies when mixed with the indicator strain D23580 ΔBTP5 (Figure 3.27). Lysogenization with the BTP5::*Km*$^R$ prophage at the BTP5 attachment site was confirmed by PCR (data not shown, primers given in Table 2.4). To confirm that lysogenization by BTP5 was not dependent on the presence of BTP1 (for example, transmission of *Km*$^R$ by generalised transduction by BTP1 phage particles), the experiment was repeated using supernatant from D23580ΔBTP1 BTP5::*Km*$^R$ to remove BTP1 phage (Figure 3.27). No effect on lysogen formation was observed, showing BTP5 is functional in the absense of BTP1. No increase in efficiency of lysogenization was observed with mitomycin C induction, consistent with reports that P2-like phages are not sensitive to induction by activated RecA protein (Kalionis et al., 1986).

Taken together, these data suggest that BTP5 is capable of infection and lysogenization of D23580 ΔBTP5, but does not form visible plaques. To check that the kanamycin-resistant colonies were not the result of a low-frequency resuscitation of a defective BTP5 by spontaneous mutation, as has been described for prophage ST64B (Figueroa-Bossi and Bossi, 2004), the experiment was repeated using a 16 h supernatant from one of the kanamycin-resistant colonies (putative BTP5::*Km*$^R$ lysogens). The number of kanamycin-resistant colonies produced after infection of strain D23580 ΔBTP5 did not increase, suggesting that no spontaneous resuscitation of BTP5 had occurred to produce the lysogens.

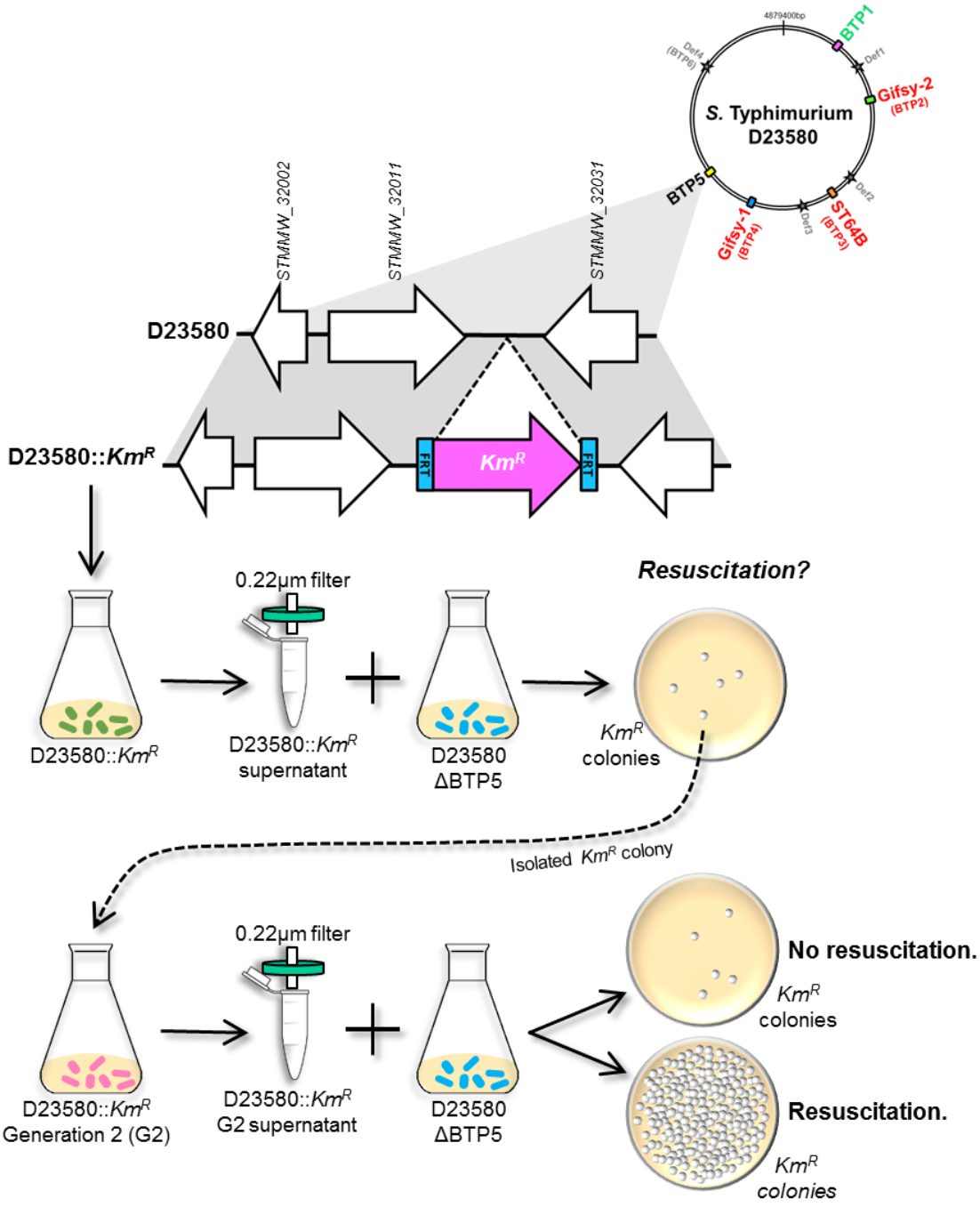



**Figure 3.27 Schematic describing experiments to determine the function of BTP5.** A kanamycin resistance locus was inserted into a non-coding position in the BTP5 prophage (D23580::*Km*$^R$). Supernatant from this strain was added to D23580 ΔBTP5 and plated on agar containing kanamycin. Resistant colonies indicated lysogeny of BTP5::*Km*$^R$. To test whether the colonies were the result of resuscitation of a defective BTP5, the experiment was repeated with donor supernatant from one of the kanamycin-resistant colonies.

Taken together these experiments suggest that BTP5 is a functional prophage, but the frequency of lysogenisation by BTP5 in the D23580 culture supernatant was extremely low. As enumeration of the phage by titration could not be achieved, empirical calculation of lysogenisation efficiency was not possible. However, on average, 50 µl of culture supernatant yielded just 8 $Km^R$ colonies (putative BTP5::$Km^R$ lysogens). It is unclear whether the low level of lysogenisation is simply a result of very low-level spontaneous induction of BTP5, or of inefficient infection of the indicator strain D23580 ΔBTP5.

Without a way to enumerate BTP5 phage particles, or visualisation by electron microscopy, the function of BTP5 remains unclear. Further exploratory work with BTP5 must be undertaken in order to determine its functionality definitively.

## 3.7 Prophage remnant elements

The prophage remnant elements found in ST313 strain D23580 are also present in many other *S.* Typhimurium strains, and therefore their gene content has been well-described (Casjens, 2011). All four prophage remnants are highly conserved in strain D23580, probably because they all contain important virulence associated genes. Even though the remnant prophages derive from horizontally acquired elements, their high conservation across *Salmonella* serovars means they are considered to belong to the core genome.

Remnant prophage Def1 (Figure 3.28) has also been designated SPI-16 in the *Salmonella* literature. Def1 is a very short remnant, which is thought to be of prophage origin because of its location adjacent to a tRNA (common prophage attachment sites) and the presence of its *gtr* (glycosyltransferase) operon. The *gtrABC* encodes a family III-type glycosyltransferase system, which mediates the addition of a glucose molecule to the galactose moiety of the repeating sub-units of the O-antigen, and is thought to be a molecular determinant for the O-antigen serotype of Typhimurium (Davies et al., 2013). Because *S.* Typhimurium genomes commonly harbour multiple *gtr* systems, the operons are conventionally given the suffix ‘a’, ‘b’ or ‘c’ to distinguish the genes. The *gtr* operon encoded by Def1 (STMMW_06231- STMMW_06251) is commonly known as *gtr*ABCa, and is important for persistence in a murine infection model (Bogomolnaya et al., 2008).

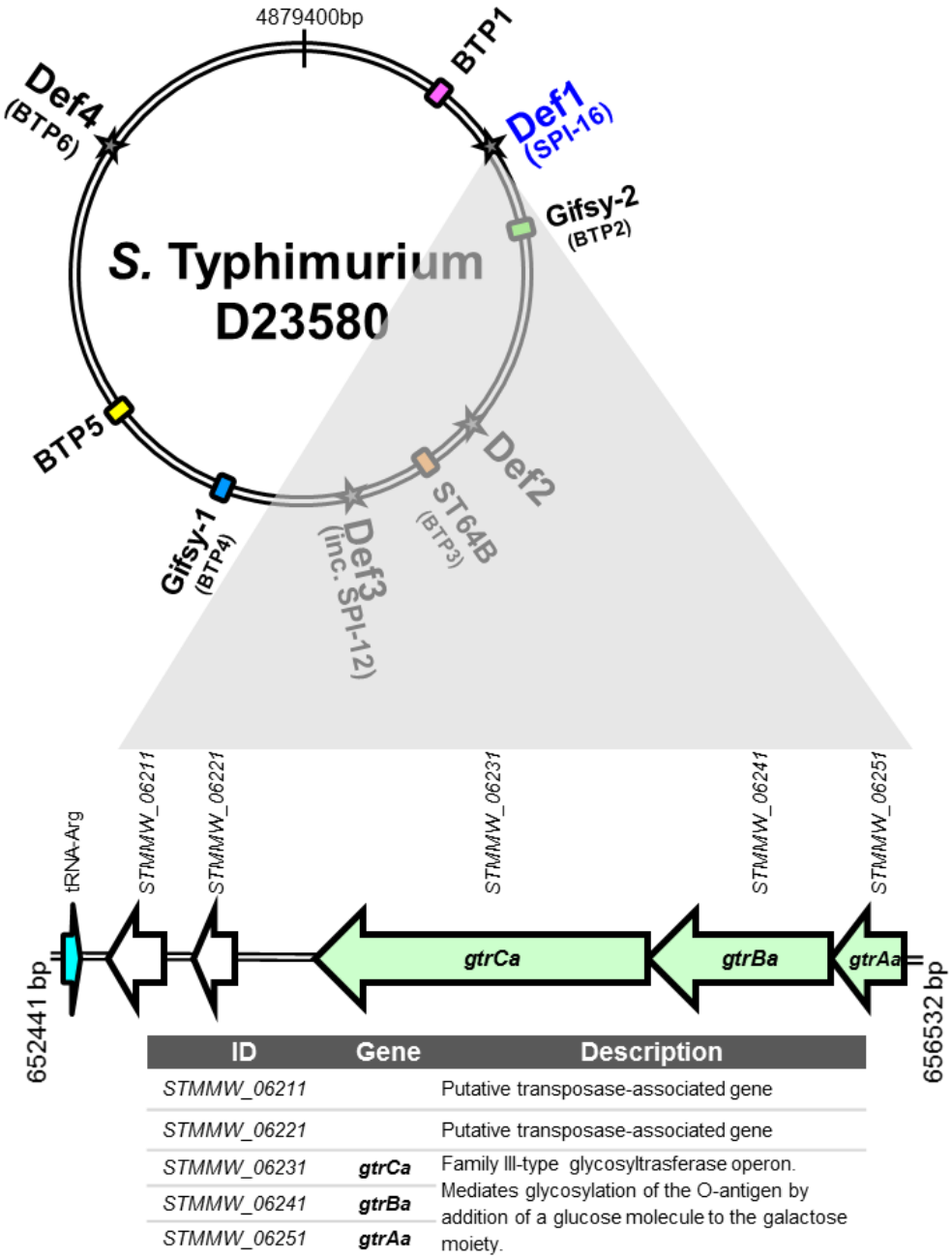
Remnant prophage Def2 encodes the type III secretion system effector protein SopE2 which affects mammalian Rho-GTPases (Figure 3.29) and a putative acyltransferase (STMMW_18481). The Def2 prophage remnant of ST313 strain D23580 is

considerably smaller than the strain in which it was first described, LT2 (Casjens, 2011), due to a ~9500 bp deletion that has been previously described (Kingsley et al., 2009). The deleted portion of Def2 includes the genes *pagO, pagM, pagK* and *mig-13* along with the small non-coding RNAs RyeA and SdsR. This large deletion, removing over half of Def2, illustrates how rapidly the process of genome degradation can create remnant prophage islands, and very few prophage-like genes now remain on Def2$^{D23580}$.

The third remnant prophage, Def3, is located adjacent to a tRNA-Pro gene (Figure 3.30). A subset of the genes in Def3 are known as SPI-12 in the *Salmonella* literature. Def3 contains the *oafA* gene, which acetylates the abequose moiety in the repeating sub-units of the O-antigen (Slauch et al., 1996). Along with a number of phage-like genes, Def-3 also encodes the type III effector protein SspH2, which can modulate host Nod1-mediated IL-8 secretion to affect inflammation (Bhavsar et al., 2013).

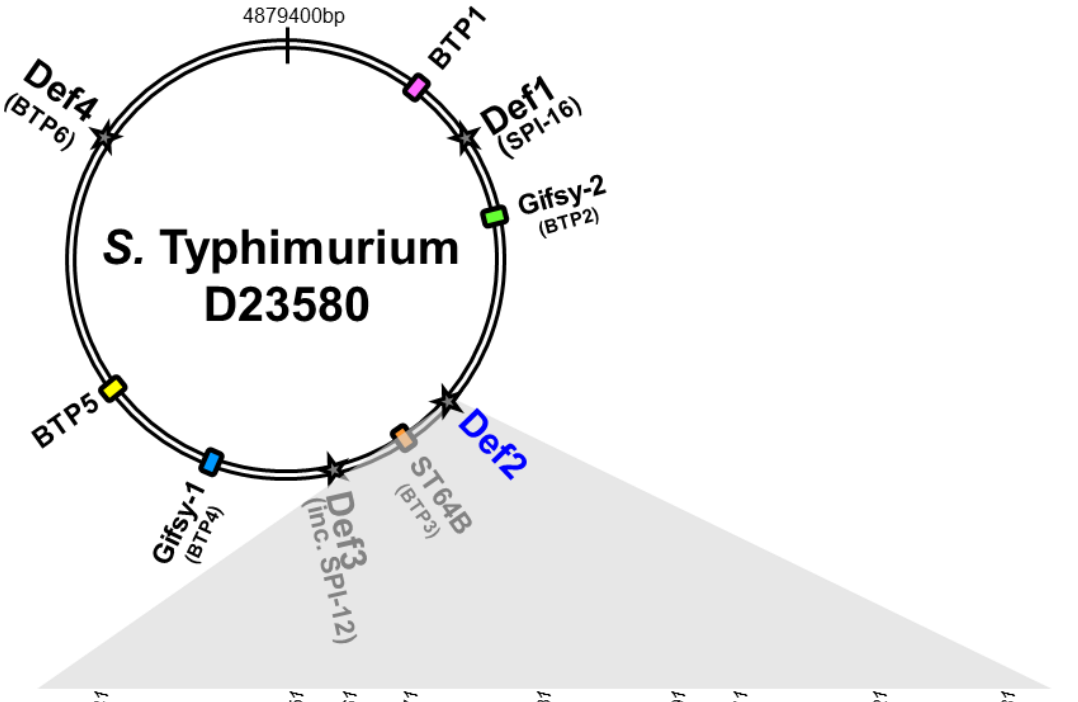
Finally, Def4, the largest remnant prophage in D23580, contains phage tail tube and sheath proteins that resemble phage P2, suggesting it is derived from a Myovirus (Casjens, 2011). Along with phage-like structural proteins, Def4 encodes a glycosyltransferase operon *gtrABC*b. This glycosyltransferase operon has not been mechanistically described, and the molecular target of GtrABCb is unknown (Davies et al., 2013). Though genome-wide mutagenesis studies have implicated some of the Def4 genes in long-term systemic infection in mice, these were the genes resembling phage structural proteins (Lawley et al., 2006), and the role of the *gtrABC*b system in *Salmonella* infection is unknown (Bogomolnaya et al., 2008)

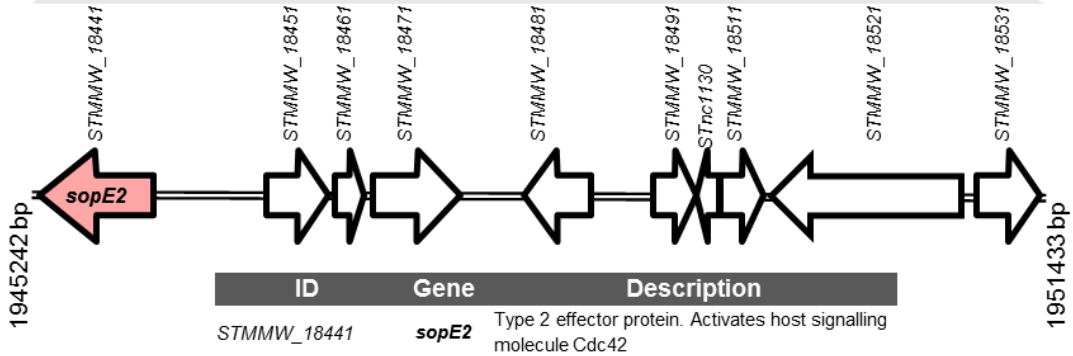
Overall the prophage remnants of D23580 present a compelling case for investigation of novel prophages as a source of virulence factors in *S.* Typhimurium ST313. All four of the prophage remnants described encode O-antigen modification systems, and two of the four carry type III-secreted effector genes.




| ID | Gene | Description |
|---|---|---|
| *STMMW_06211* | | Putative transposase-associated gene |
| *STMMW_06221* | | Putative transposase-associated gene |
| *STMMW_06231* | ***gtrCa*** | Family III-type glycosyltrasferase operon. Mediates glycosylation of the O-antigen by addition of a glucose molecule to the galactose moiety. |
| *STMMW_06241* | ***gtrBa*** | |
| *STMMW_06251* | ***gtrAa*** | |

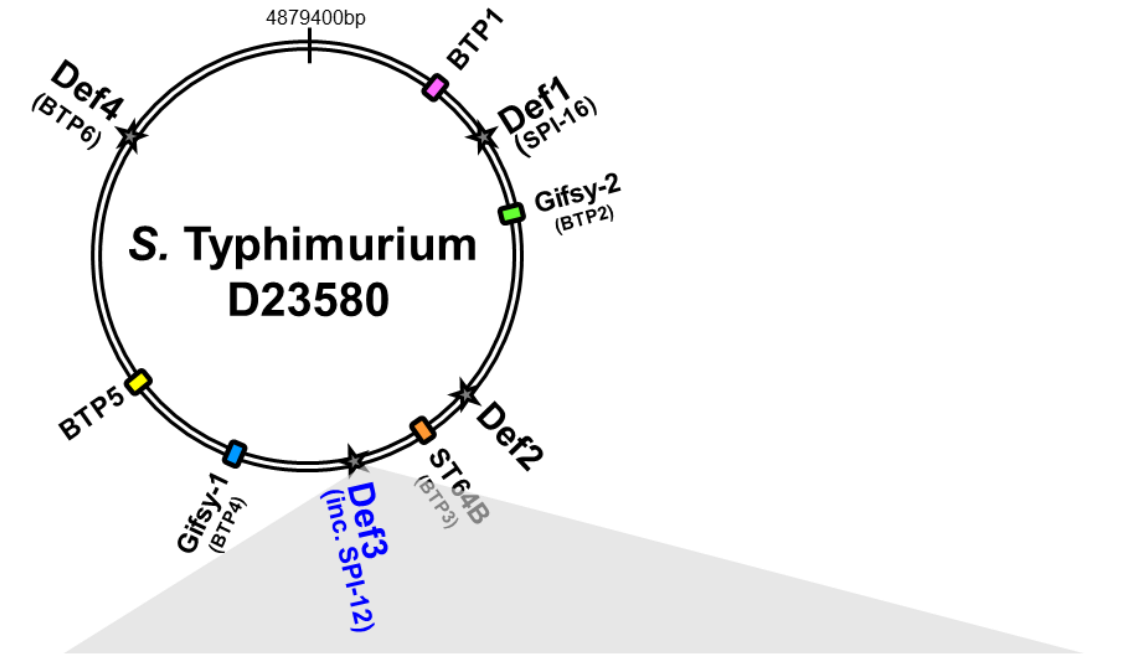
**Figure 3.28 Gene map and functional annotation of prophage remnant Def1$^{D23580}$.**

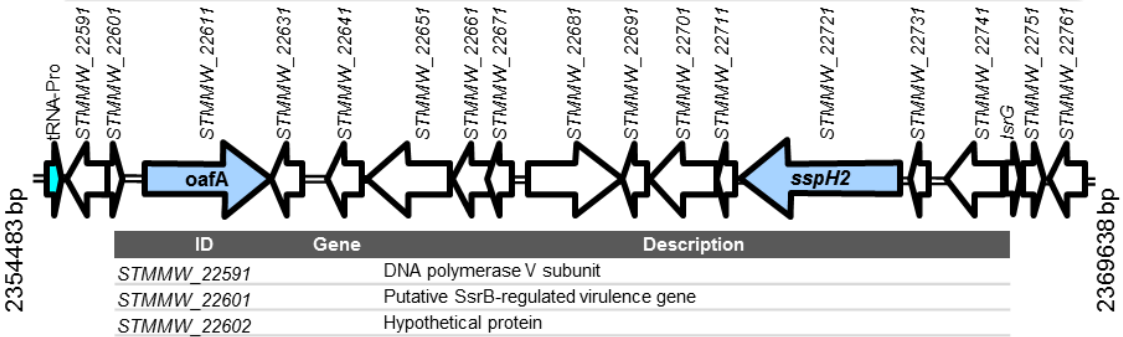



| ID | Gene | Description |
|---|---|---|
| *STMMW_18441* | ***sopE2*** | Type 2 effector protein. Activates host signalling molecule Cdc42 |
| *STMMW_18451* | | Putative cytoplasmic protein |
| *STMMW_18461* | | Hypothetical protein |
| *STMMW_18471* | | Putative membrane protein |
| *STMMW_18481* | | Putative acetyltranferase |
| *STMMW_18491* | | Hypothetical protein |
| *STnc1130* | | Non-coding RNA |
| *STMMW_18511* | | Hypothetical protein |
| *STMMW_18521* | | Putative transposase (pseudogene) |
| *STMMW_18531* | | Putative integrase |

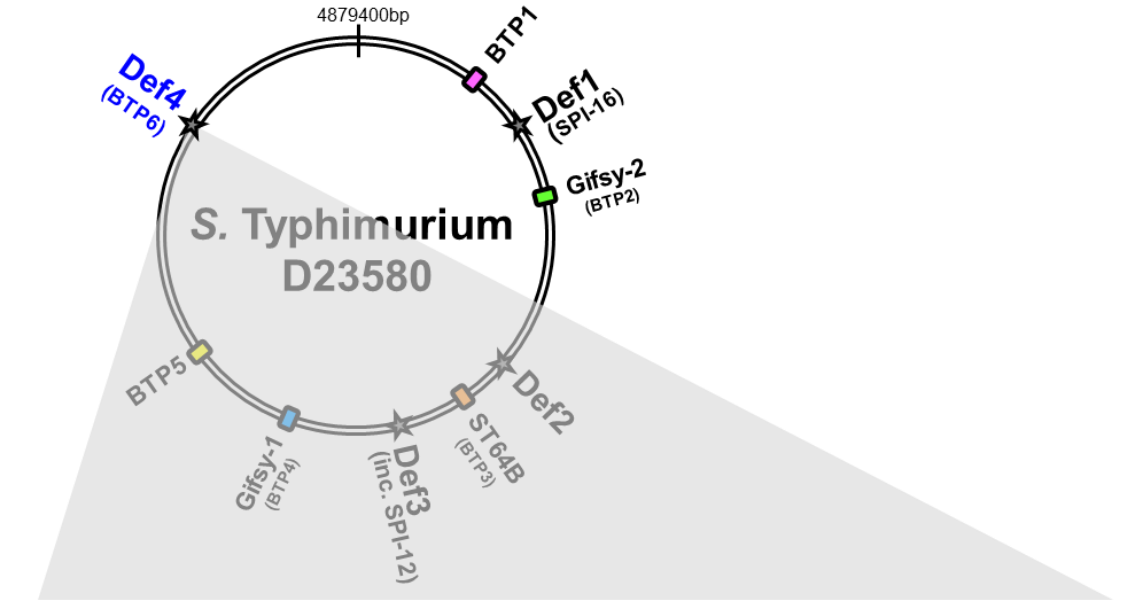
**Figure 3.29 Gene map and functional annotation of prophage remnant Def2$^{D23580}$.**

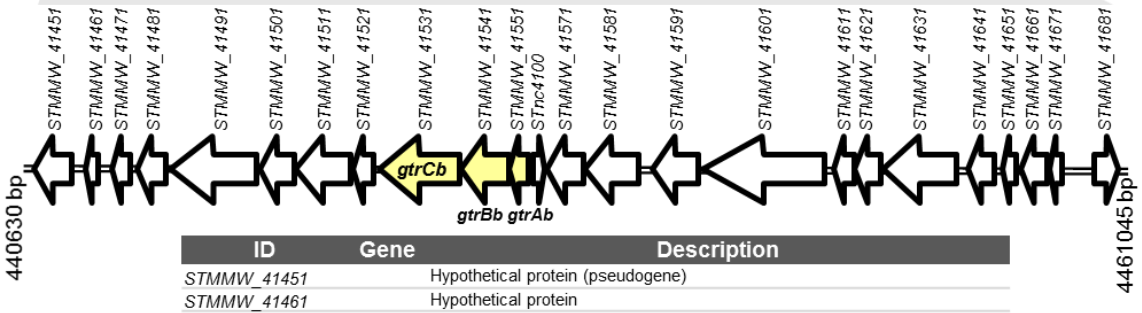



| ID | Gene | Description |
|---|---|---|
| *STMMW_22591* | | DNA polymerase V subunit |
| *STMMW_22601* | | Putative SsrB-regulated virulence gene |
| *STMMW_22602* | | Hypothetical protein |
| *STMMW_22611* | ***oafA*** | O-antigen acetylase |
| *STMMW_22631* | | Hypothetical protein |
| *STMMW_22641* | | Putative phage tail fibre assembly protein |
| *STMMW_22651* | | Putative phage protein |
| *STMMW_22661* | | Putative phage prohead protease |
| *STMMW_22671* | | Putative phage holing |
| *STMMW_22681* | | Putative phage protein (pseudogene) |
| *STMMW_22691* | | Putative phage protein |
| *STMMW_22701* | | Putative phage antitermination protein (Q) |
| *STMMW_22711* | | Putative phage-related protein |
| *STMMW_22721* | ***sspH2*** | Type 3 effector protein. Enhances Nod1-mediated IL-8 secretion |
| *STMMW_22731* | | Putative phage tail fibre protein |
| *STMMW_22741* | | Putative phage side tail protein |
| *STnc150* | ***IsrG*** | Non-coding RNA |
| *STMMW_22751* | | Homolog of virulence gene *msgA* |
| *STMMW_22761* | | Putative exported protein |

**Figure 3.30 Gene map and functional annotation of prophage remnant Def3$^{D23580}$.**




| ID | Gene | Description |
|---|---|---|
| *STMMW_41451* | | Hypothetical protein (pseudogene) |
| *STMMW_41461* | | Hypothetical protein |
| *STMMW_41471* | | Putative cytoplasmic protein |
| *STMMW_41481* | | Putative cytoplasmic protein |
| *STMMW_41491* | | Putative phage tail fibre protein |
| *STMMW_41501* | | Putative phage baseplate protein |
| *STMMW_41511* | | Putative phage baseplate protein |
| *STMMW_41521* | | Putative inner membrane protein |
| *STMMW_41531* | ***gtrCb*** | Family IV glycosyltransferase operon required for *Salmonella* persistence. This family of *gtr* operon has not been mechanistically described. |
| *STMMW_41541* | ***gtrBb*** | |
| *STMMW_41551* | ***gtrAb*** | |
| *STnc4100* | | Non-coding RNA |
| *STMMW_41571* | | Putative phage baseplate component protein |
| *STMMW_41581* | | Putative phage protein |
| *STMMW_41582* | | Putative inner membrane protein |
| *STMMW_41591* | | Putative phage protein |
| *STMMW_41601* | | Putative phage tail protein |
| *STMMW_41611* | | Putative phage protein |
| *STMMW_41621* | | Putative phage tail tube protein |
| *STMMW_41631* | | Putative phage tail sheath protein |
| *STMMW_41632* | | Putative cytoplasmic protein |
| *STMMW_41641* | | Putative phage protein |
| *STMMW_41651* | | Putative phage protein |
| *STMMW_41661* | | Putative phage lysin |
| *STMMW_41671* | | Putative inner membrane protein |
| *STMMW_41681* | | Putative DNA-binding/regulatory protein |

**Figure 3.31 Gene map and functional annotation of prophage remnant Def4$^{D23580}$.**

## 3.8 Discussion

Representative ST313 strain D23580 was reported to carry five prophages (Kingsley et al., 2009). Three of these, Gifsy-2, ST64B and Gifsy-1, are present in many *S.* Typhimurium genomes, whilst BTP1 and BTP5 are novel prophages. The prophage repertoire of D23580 was reported to be highly conserved in other ST313 genomes of both described ST313 lineages (Kingsley et al., 2009; Okoro et al., 2015). The presence of novel prophages in the genome of ST313 *S.* Typhimurium is not in itself surprising as prophages are known to be highly variable in *Salmonella* populations (Brüssow et al., 2004; Figueroa-Bossi et al., 2001). However, despite their intrinsic mobility and apparent transience, prophage carriage has also been strongly associated with *Salmonella* epidemics and the link between prophage carriage and *Salmonella* disease epidemiology is highlighted by the continual use of phage typing schemes for surveillance of *Salmonella* outbreaks since the 1970s (Rabsch, 2007). Consequently, the potential contribution to the virulence of ST313 by novel prophages BTP1 and BTP5 remains intriguing. Certainly, the presence of a number of putative accessory genes in novel phages BTP1 and BTP5 warrants further investigation.

One BTP1 gene, *st313-td,* has been reported to be associated with virulence of ST313 in a mouse infection model (Herrero-Fresno et al., 2014). Though no mechanism of action was described for the virulence effect of *st313-td*, if the gene proves to be important for bloodstream invasion, its high conservation among ST313 isolates would be strong evidence that prophage BTP1 contributes to the virulence of ST313. Additionally, the high level of similarity of BTP1 to the generalized-transducing phage P22 raises interesting implications for horizontal gene transfer among BTP1-susceptible *Salmonellae.* Due to the high level of conservation in ST313 genomes, it is tempting to speculate that prophages are important within the ecological niche of ST313, whether through immunity to environmental phage predation, permitting colonization of environmental reservoirs in Africa or by contributing to bacterial virulence through lysogenic conversion.

As well as putative cargo virulence factors, genetic analysis revealed that the BTP1 tailspike may be unusual. Amino acid sequence comparison of the BTP1 and P22 tailspike protein receptor-binding domains (RBD) showed considerable divergence (65% amino acid percent identify), suggesting the enzymatic activity of the BTP1 tailspike may differ from other P22-like phages. In their thorough review on the evolution of P22-like phages, Casjens & Thuman-Commike included the BTP1 prophage sequence in detailed phylogenetic analysis of capsid assembly genes

(Casjens and Thuman-Commike, 2011) where it is designated 'Typh1'. Figures 7, S1Z4 and S1Z5 in the aforementioned review show that whilst the BTP1 tailspike head-binding domain is very closely related to P22-like phages, the RBD falls on a deep branch amongst non-P22-like *Salmonella* phages, sharing more recent common ancestors with the RBDs of sipho- and myovirus tailspikes than P22-like phages. The significance of the unusual BTP1 tailspike RBD is unclear.

Investigation of the biology of the novel prophages in ST313 revealed that the titer of phage BTP1 in the 16 h culture supernatant is about $10^9$ PFU/ml without chemical induction, approximately equal to the bacterial CFU/ml, suggesting a high frequency of spontaneous induction. The spontaneous induction of BTP1 was visualized at the single-cell level using a fluorescent reporter system. In fact, BTP1 has the highest spontaneous induction titer of any prophage in the contemporary literature. The high frequency spontaneous induction was maintained in a different host strain (LT2), indicating it is an intrinsic characteristic of prophage BTP1. However, it should be noted that, in part due to the lack of functional prophage characterisation studies, spontaneously induced phage titers are not reported for the vast majority of known prophages. Additionally, some caution has to be taken when interpreting the spontaneously induced phage titers described in the literature. Spontaneous phage titers are likely to be frequently underestimated due to low efficiency of plating of the phage in question on non-optimal indicator strains. Accurate estimation of phage titer using plaque assay methodology relies on the efficiency of plating of the indicator strain, and indicator strains with maximal efficiency of plating may have to be made synthetically using genome editing approaches to remove any elements that confer phage resistance, as has been undertaken in this study.

Spontaneous prophage induction is a well-known phenomenon in bacteria and the presence of free phages in the cultivation media of lysogens has been reported as early as the 1950s (Lwoff, 1953). However, spontaneous induction and its potential effect on host biology has not been well documented among clinically relevant bacteria. Furthermore, the effect of spontaneous prophage induction specifically on *Salmonella* infection dynamics and virulence is not well understood. There is evidence that spontaneous phage induction directly contributes to virulence in a number of other bacterial pathogens (reviewed in (Nanda et al., 2015). Phage induction causes an increase in biofilm formation by releasing extracellular DNA in organisms including *Streptococcus pneumoniae* and *Shewanella oneidensis* (Carrolo et al., 2010; Gödeke et al., 2011). More recently, spontaneous prophage induction was shown to modulate

*Pseudomonas aeruginosa* biofilm formation by stimulating biogenesis of membrane vesicles (Turnbull et al., 2016). However, contrary to published associations between spontaneous prophage induction and biofilm formation, recent publications have shown ST313 to be defective at forming RDAR (red, dry, and rough) colonies and biofilms (Ramachandran et al., 2016; Singletary et al., 2016).

It has also been reported that phage-mediated lysis may act as a crude method of toxin delivery, facilitating human gingival fibroblast invasion by *Aggregatibacter actinomycetemcomitans* (Stevens et al., 2013), and enabling Stx toxin release in Shiga toxin-producing *E. coli* (Shimizu et al., 2009; Xu et al., 2012). *Salmonella* Typhi is known to produce a specific toxin, CDT (Galan, 2016), but no candidate toxin genes have been identified in the genome of ST313 strain D23580. It is possible that BTP1-mediated lysis causes increased release of cellular material such as endotoxin during infection, and this deserves further investigation.

Examples of prophages enhancing bacteria fitness by expressing advantageous accessory genes or by mediating spontaneous lysis are numerous (Fortier and Sekulovic, 2013; Nanda et al., 2015). However, the inevitable result of spontaneous induction of tailed phages is cell death, and therefore any benefit of lysogenic conversion or spontaneous induction itself is offset by the negative impact on bacterial growth dynamics. The high level of spontaneous BTP1 induction equates to the lysis of approximately 0.2% of the bacterial population in strain D23580. It seems likely that if BTP1-mediated cell death was deleterious, ST313 derivatives containing mutations in the BTP1 prophage would have been positively selected (Bobay et al., 2014; Lawrence et al., 2001), and in fact, the BTP1 prophage is highly conserved in ST313 from across Africa (Okoro et al., 2015). This leaves the intriguing possibility that phage-mediated lysis could be tolerated by the ST313 population if the benefit of lysogenic conversion factors such as immunity from BTP1-like phages and cargo loci were advantageous enough to outweigh the cost of spontaneous lysis. The functional conservation of BTP1 across all African ST313 isolates of both described lineages supports the hypothesis that the high rate of BTP1 spontaneous induction benefits the pathogen in some way. The impact of BTP1-mediated lysis on ST313 growth dynamics, ecological niche, virulence and fitness will be an important topic for further study.

Despite this, if there was a strong selection pressures to retain the accessory functions of BTP1, mutations could arise that inactivate prophage induction whilst retaining the lysogenic conversion phenotype, such as the mutations reported in this

chapter for the defect Gifsy-2, ST64B and Gifsy-1 prophages of ST313. The finding that the non-ST313-specific prophages Gifsy-1, ST64B and Gifsy-2 are defective in D23580 is intriguing. Although these prophages are present in most sequenced *S.* Typhimurium genomes in NCBI (accessed June 2016), this is the first report of mutational inactivation of all three prophages in the same chromosome. The significance of these inactivations remains to be established.

From a phage perspective, it could be argued that temperate phages exploit their capacity to transmit either horizontally (i.e. lytically as a virion) or vertically (i.e. lysogenically as a prophage) as a bet-hedging strategy to improve overall transmission. In fact, it was recently shown that during an active infection, phage P22 regularly engages in an asymmetrically segregating phage carrier state that results in the cultivation of a transiently P22 resistant subpopulation of host cells that can subsequently be lytically consumed (Cenens et al., 2015). As such, P22 likely manages to maintain an active infection with high virion titers in a host population that is progressively taken over by P22 lysogens. In the same bet-hedging context, the BTP1 prophage could have evolved to vastly increase its spontaneous induction rate in order to ensure high virion titers that increases the opportunity for the phage to mediate horizontal transmission whenever the BTP1 lysogen would encounter a niche with susceptible host cells. Furthermore, such BTP1-mediated lysis of susceptible competitors would likely contribute to the ecological fitness of BTP1 lysogens as well, as demonstrated previously with Gifsy-1 and Gifsy-2 lysogens (Bossi et al., 2003).

In summary, this thorough characterization of the prophage repertoire of *S.* Typhimurium ST313 poses interesting questions about the potential fitness costs and benefits of novel prophages in epidemic *S.* Typhimurium ST313. The high rate of BTP1-mediated spontaneous lysis represents novel biology in this clinically important sequence type, and could modulate the behavior of the pathogen in terms of ecological niche, host range or invasiveness in humans.

# Chapter 4

## Diverse UK lineages of *S.* Typhimurium ST313

## 4.1 Introduction

Multi locus sequence typing (MLST) is a molecular approach for typing micro-organisms and uses the allelic varation of seven highly conserved *Salmonella* housekeeping genes to infer bacterial phylogeny (Achtman et al., 2012). Whole genome sequence studies of isolates collected from patients with iNTS disease in sSA initially identified a novel sequence type (ST), ST313, of *S.* Typhimurium in 2009 (Kingsley et al., 2009). Subsequent genome-based studies confirmed two distinct phylogenetic lineages of ST313 associated with iNTS and spatio-temporal phylogenetic reconstruction suggested that lineage 1 emerged around 1960 in south west Africa, whereas lineage 2 emerged around 1977 in Malawi (Okoro et al., 2012). Both lineages are associated with antimicrobial resistance (AMR) mediated by differing Tn*21*-like integrons on the virulence plasmid pSLT (Kingsley et al., 2009) and it was proposed that clonal replacement of lineage 1 by lineage 2 had occurred, driven by the acquisition of chloramphenicol resistance in lineage 2 (Okoro et al., 2012). *S.* Typhimurium ST313 has recently been detected in Brazil (Almeida et al, 2017). Unlike African ST313, the nine Brazilian ST313 isolates were predominantly associated with gastro-intestinal infection and were antibiotic-susceptible.

It has been suggested that the link between *S*. Typhimurium ST313 and iNTS disease in sSA is that, compared with the generalist *S*. Typhimurium ST19, ST313 has adapted to an extra-intestinal/invasive lifestyle via genome degradation (Kingsley et al., 2009; Okoro et al., 2015). This would be consistent with the finding of an accumulation of pseudogenes in pathways associated with gastrointestinal colonization, as observed in host-restricted *Salmonella* serovars such as *S.* Typhi and in *Yersinia pestis, Shigella* spp, *Mycobacterium leprae* and *Bordetella pertussis* (Cole et al., 2001; McClelland et al., 2004; McNally et al., 2016; Nuccio and Bäumler, 2014; Parkhill et al., 2001; The et al., 2016; Yang et al., 2005). Another observation from comparative genomic studies, which supports the hypothesized enhanced virulence of ST313 includes the detection of novel prophages BTP1 and BTP5 (Owen et al., 2017; Chapter 3) including the reported BTP1 phage-encoded putative virulence gene, *st313-td* (Herrero-Fresno et al., 2014).

A number of phenotypic characteristics which distinguish ST313 from gastroenteritis-associated ST19 strains have been described, including a reduction in motility, flagellin expression, stationary-phase catalase activity and biofilm formation (Carden et al., 2015; Ramachandran et al., 2015; Singletary et al., 2016). Despite these

phenotypes, it remains to be proven whether ST313 are intrinsically capable of causing a higher level of systemic disease than ST19.

Since April 1st, 2014, every presumptive *Salmonella* isolate received by the *Salmonella* Reference Service (SRS) of Public Health England (PHE) has been whole genome-sequenced (WGS) to allow identification, characterization and typing in one laboratory process (Ashton et al., 2016). In this chapter, I investigate the prevalence of ST313 in cases of laboratory-confirmed *S*. Typhimurium infections reported in England and Wales, analyse clinical data regarding the origin of isolates (faeces or blood) and review information collected from patients regarding whether infection was associated with travel to high-incidence areas such as sSA. A phylogenetic approach is used to place the UK-isolated ST313 into the evolutionary context of ST313 lineages isolated in Africa (hereafter referred to as African ST313).

### 4.1.1 Acknowledgement of the specific contribution of collaborators to the results described in Chapter 4.

The majority of this chapter has been submitted for publication and makes up an article entitled "Public health surveillance in the UK revolutionises our understanding of the invasive *Salmonella* Typhimurium epidemic in Africa". The manuscript is available as an un-reviewed preprint (https://doi.org/10.1101/139576), and is currently in revision at a peer reviewed journal. I acknowledge the following contribution of collaborators to the results described in this chapter. Unless specified below, all work was completed by the Author.

| | |
|---|---|
| **Phil Ashton**<br>Public Health England, UK | Assembly, phylogenetic & BEAST analysis (described in 2.11.1 & 2.12.3) of UK-isolated ST313 strains |
| **Lukeki Kaindama & Lesley Larkin**<br>Public Health England, UK | Acquisition of enhanced surveillance data from the patients from whom UK-isolated ST313 strains were derived |
| **Will Rowe**<br>University of Liverpool, UK | Improved assembly of the $bla_{\text{CTX-M-15}}$ element from the genome of isolate U60 and read mapping to confirm insertion into *ompD*. |

## 4.2 Epidemiology of ST313 in the UK

Between January 2014 and May 2016, 2,888 *S*. Typhimurium isolates were whole-genome sequenced by Public Health England (all *Salmonella* isolates sequenced by PHE are available from the NCBI SRA BioProject PRJNA248064), of which 79 (2.7%) were of multi-locus sequence type ST313. Whole genome sequence data were available for a further 363 *S*. Typhimurium isolates from 2012, of which 7 (1.9%) were ST313. Of these 86 ST313 isolates, 75 were derived from human patients (5 patients had two isolates sequenced and one patient had 5 isolates sequenced), 1 from a dog and 1 was isolated from an unspecified raw food sample. The sample type was recorded for 72 of the 75 human patient isolates; 13 patients had one or more isolate from extra-intestinal sites (blood, pus or bronchial alveolar lavage) indicating iNTS disease and 59 isolates were from faeces alone (indicating gastrointestinal infection). Travel information was available for 51 of the 75 human patients, of whom 8 reported travel to sSA during the estimated disease incubation period and one adult male reported consuming food of West African origin in London.

Of the 51 patients with travel information, 48 had sample type recorded. Of the 8 patients who reported travel to Africa, 6 had extra-intestinal infections. In contrast, just 2 of 40 patients for whom travel information was available and did not report travel to Africa had extra-intestinal infection, showing that travel to Africa is significantly associated with iNTS disease in the UK (OR 57.0 [95% CI: 6.7, 484.8], p-value = 0.0002) (Table 4.1A).

**Table 4.1 Summary of key epidemiological features of ST313 sampled by PHE in 2012 and 2014-2016.** Odds ratios, 95% confidence intervals (CI) and p-values are shown for each association.

**A** **Association between isolation source and travel to sub-Saharan Africa. Total = 48 (51 individual patient isolates with travel info but 3 with source unknown).**

| | Extra-intestinal infection | Gastro-intestinal infection |
|---|---|---|
| **Reported travel to sub-Saharan Africa** | 6 | 2 |
| **No reported travel to sub-Saharan Africa** | 2 | 38 |
| | Odds ratio 57.0 [95% CI: 6.7, 484.8], p-value 0.0002 | |

**B** **Association between lineage 2 infection and travel to sub-Saharan Africa. Total= 51 (51 individual patient isolates with travel info)**

| | Lineage 2 | Non-lineage 2 |
|---|---|---|
| **Reported travel to sub-Saharan Africa** | 7 | 1 |
| **No reported travel to sub-Saharan Africa** | 2 | 41 |
| | Odds ratio 143.5 [95% CI: 11.4, 1802.9], p-value 0.0001 | |

**C** **Association between lineage 2 infection and extra-intestinal infection. Total = 72 (75 individual human patient isolates in total, but 3 with source unknown & one not included in phylogeny due to poor quality sequence so lineage unknown)**

| | Lineage 2 | Non-lineage 2 |
|---|---|---|
| **Extra-intestinal infection** | 10 | 3 |
| **Gastro-intestinal infection** | 1 | 57 |
| | Odds ratio 190.0 [95% CI: 17.9, 2014.0], p-value < 0.0001 | |

## 4.3 Phylogenetic analysis reveals unprecedented diversity of ST313

Sequence data quality was sufficient to permit whole genome single nucleotide polymorphism (SNP) phylogenetic analysis of the isolates from 76 of the 77 patient and non-human isolates. Within the wider phylogenetic context of *S*. Typhimurium, all ST313 isolates submitted to PHE formed a monophyletic group that clustered with

previously described African ST313 isolates (Okoro et al., 2012) (Figure 4.1). A second maximum likelihood ST313 phylogeny was generated with a closely-related ST19 strain also received by PHE (strain U21) as an outlier, to study phylogenetic relationships (Figure 4.2). Of the 76 isolates from distinct patients/sources, 12 belonged to the previously described lineage 2, and 64 did not fall within any ST313 lineage that has been reported to date. No lineage 1 isolates were identified. Furthermore, the African associated lineages 1 and 2 did not form a monophyletic group within the novel diversity observed. Both of the African lineages share more recent common ancestors with UK-associated strains than with each other. Neither the food nor the animal isolate belonged to lineage 2. All the UK-derived isolates in this study originated from diagnostic laboratories in England and Wales. To simplify the categorization/differentiation of the lineages for discussion purposes, isolates belonging to lineage 1 and 2 (including those isolated in the UK) will be referred to as African lineages and the non-lineage 1 and 2 isolates will be referred to as UK-ST313.

The UK-ST313 isolates do not themselves form a coherent monophyletic cluster, revealing an unappreciated level of genetic diversity within ST313 (Figure 4.1).To examine the evolutionary history of ST313, a maximum clade credibility tree was inferred using Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Figure 4.3) and the topology was largely congruent with respect to the Maximum Likelihood tree (Figure 4.1, Figure 4.3). The most recent common ancestor (MRCA) of ST313 is estimated to have been in approximately 1787 (95% highest posterior distribution (HPD), 1735 - 1836). Lineage 1 was estimated to have diverged from other ST313 sampled in this study in approximately 1796 (95% HPD 1744-1842), while lineage 2 diverged from other ST313 sampled here in 1903 (95% HPD 1876 - 1927). The lineage 1 MRCA dated to around 1984 (95% HPD 1979-1987), while the lineage 2 MRCA dated to around 1991 (95% HPD 1986-1995). These confidence intervals overlap with the confidence intervals reported for the emergence of the two lineages by Okoro *et al*., 2012 (Okoro et al., 2012). The two African lineages do not form a monophyletic group, and share an MRCA which is very close to that of ST313 as a whole. Both lineages 1 and 2 are separated from others by long branches, indicating a distant MRCA with other isolates and suggesting that a tight population bottleneck has occurred relatively recently in evolutionary history.

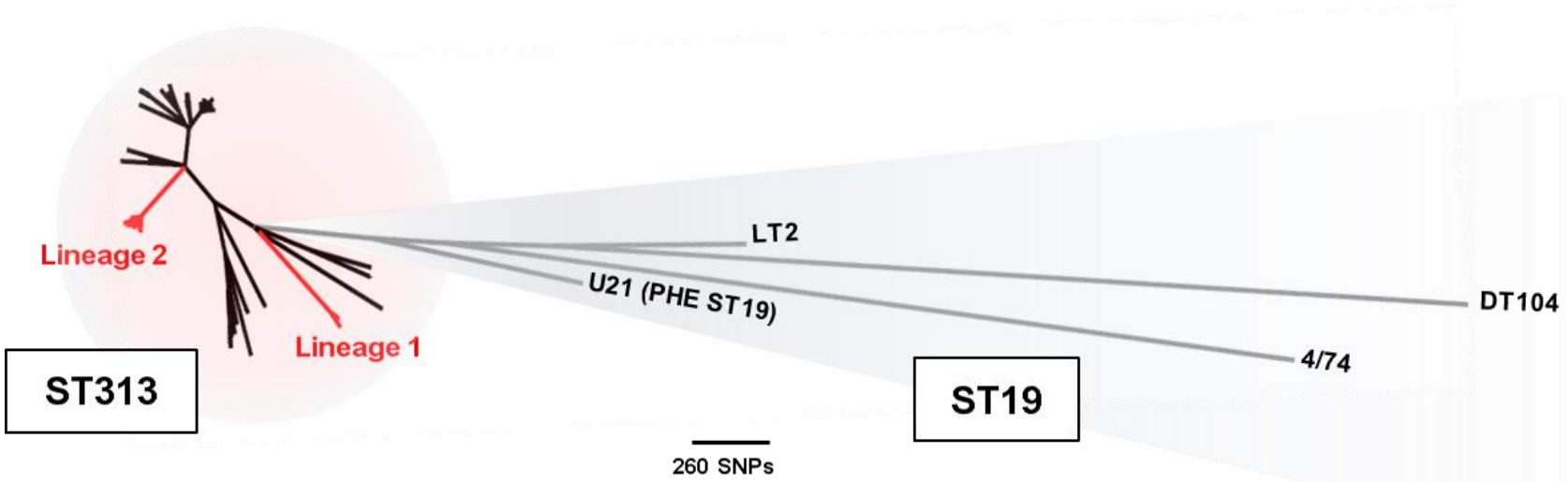



**Figure 4.1 UK-ST313 isolates are phylogenetically distinct from African lineages.** Unrooted maximum likelihood phylogeny of ST313 in the context of reference *Salmonella* Typhimurium ST19 isolates (2.12.3). Isolate U21 was an ST19 isolate, closely related to ST313, that was used as an outgroup in further ST313 analyses. The African epidemic ST313 lineages 1 and 2 are labelled. Branch colour indicates phylogenetic groupings; ST19 stains are shown with grey branches, African lineage 1 or 2 are shown with red branches, and the novel diversity represented by the UK-ST313 isolates are shown with black branches. The scale bar indicates the SNP-distance represented by branch lengths.

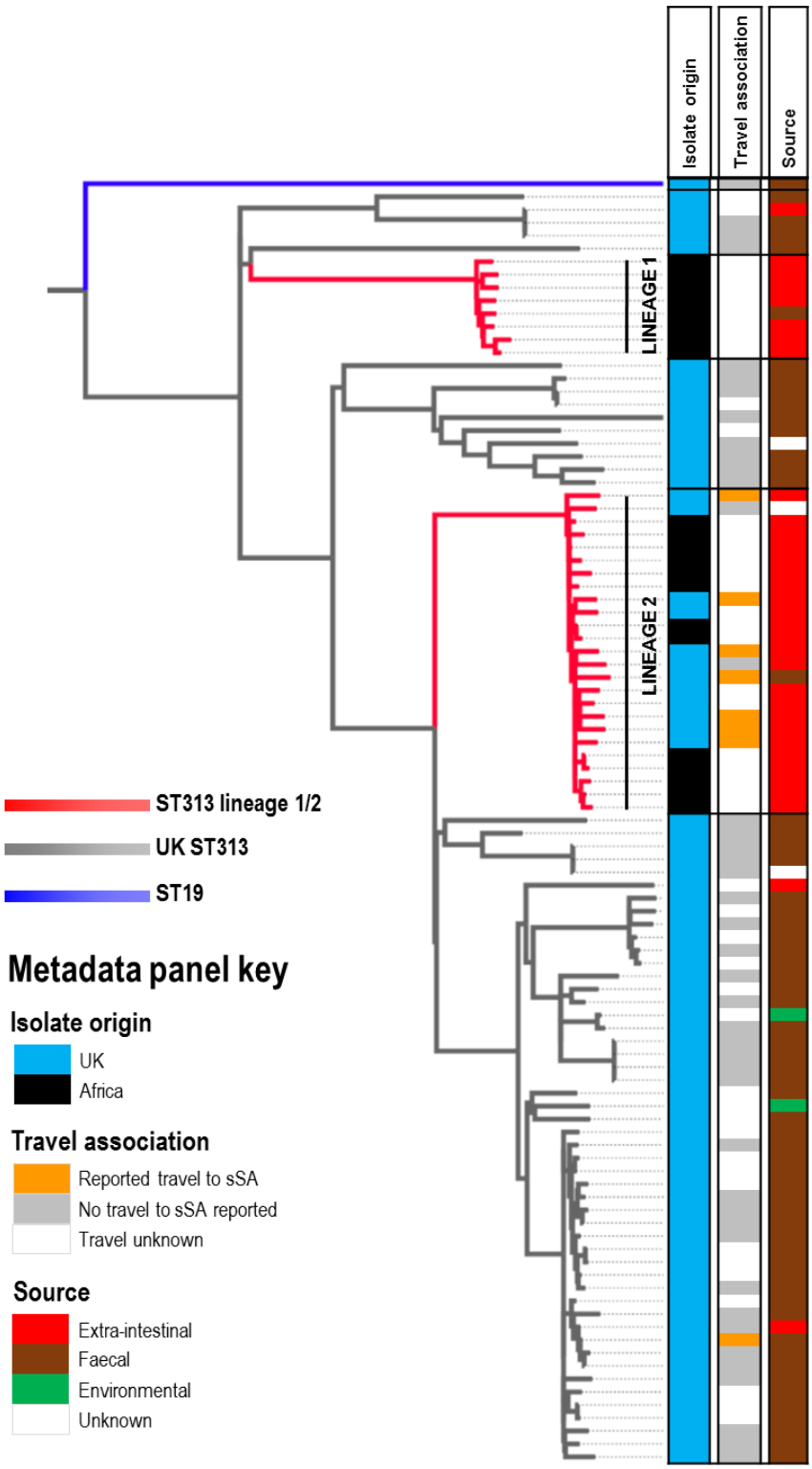



**Figure 4.2 UK-ST313 are associated with gastrointestinal infection and not associated with travel to Africa.** A Maximum likelihood phylogeny of 77 UK-isolated ST313 strains received by PHE in the context of 24 African ST313 sequenced by Okoro *et al.* (2012). Red branches indicate ST313 lineage 1 and 2. Adjacent metadata panel showing: 1. Country isolate was associated with, Africa- black, not Africa- blue; 2. Travel association, reported travel to Africa- orange, no reported travel to Africa-grey, travel history unknown- white; 3. Source, extra-intestinal- red, faecal- brown, environmental- green, unknown- grey.

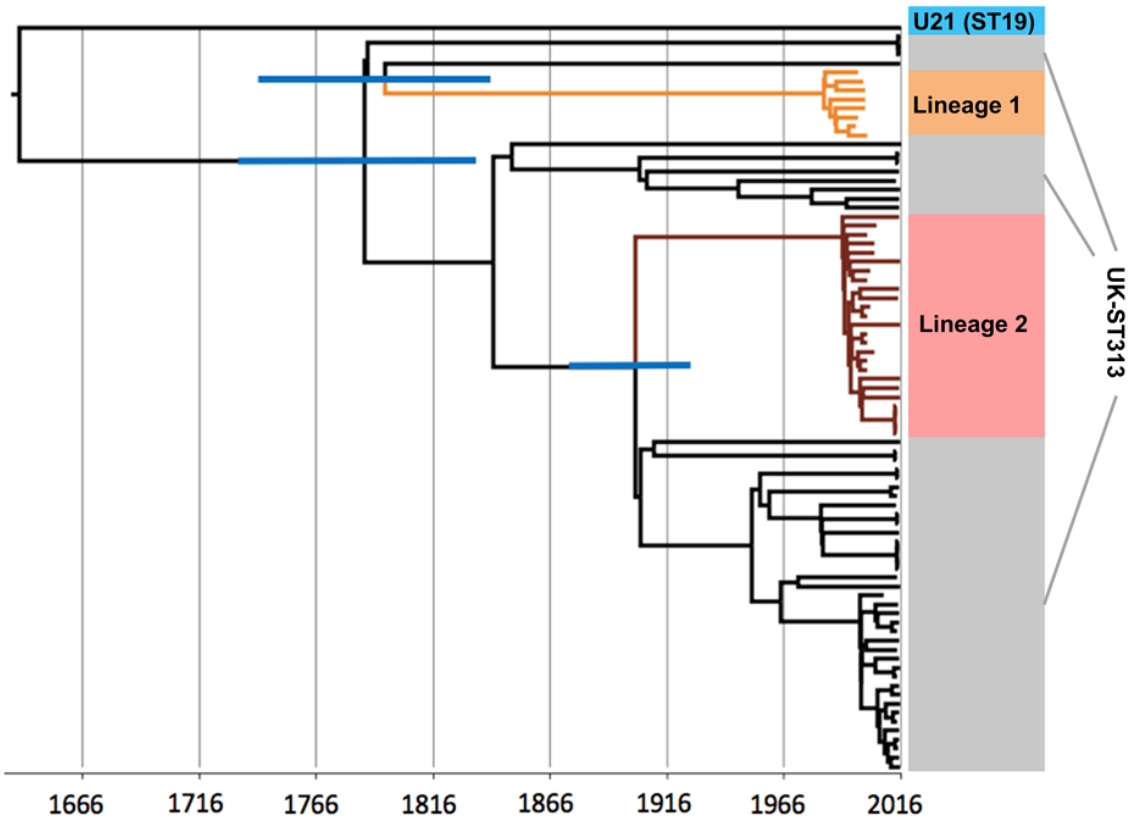



**Figure 4.3 The timed phylogeny of all UK-isolated ST313 strains from this study and a representative sub-sample of ST313 genomes from Okoro et al., 2012.** BEAST analysis was conducted by P. Ashton, Public Health England. Figure shows the maximum clade credibility tree from BEAST. Branches 95% highest posterior density (HPD) are displayed in blue for key nodes defining lineage 1 and lineage 2. Branches belonging to lineage 1 are coloured orange and branches belonging to lineage 2 are coloured brown.

## 4.4 The association between phylogenetic context and travel to Africa

The association between reported travel to sSA and infection with the lineage 2 isolates was investigated with assistance from epidemiologists from Public Health England. Of the 8 UK patients reporting travel to sSA during the seven days before disease onset, 7 were infected with a lineage 2 isolate. In contrast, of the 43 patients who reported no travel to sSA, 2 were infected with lineage 2 (Table 4.1B). This shows that travel to sSA was significantly associated with infection by ST313 lineage 2 (OR 143.5 [95% CI: 11.4,1802], p-value = 0.0001).

We investigated whether the ST313 lineage 2 isolates were more frequently associated with extra-intestinal or gastro-intestinal infection. Of the 11 patients infected with lineage 2 (and for which source of isolation data was available), 10 patients had isolates that originated from extra-intestinal sites. In contrast, for the patients infected with UK-ST313 isolates, 3 of 60 isolates were from extra-intestinal sites (Table 4.1C). These data show that in the UK, infection with ST313 lineage 2 is significantly associated with invasive disease (OR 190.0 [95% CI: 17.9, 2014.0], p-value < 0.0001).

## 4.5 Accessory genome signatures of ST313

Multi-drug resistance is a key phenotypic feature associated with both the African ST313 lineages and is encoded by Tn*21*-like integron insertions on the pSLT virulence plasmid. Analysis of the genome sequences indicated that all 76 ST313 isolates in this study carried the pSLT plasmid. However, the majority of the UK-ST313 isolates were antibiotic-sensitive (59/64 were sensitive to all antimicrobials tested) and no consistent AMR gene profile was identified. In contrast, 10 of 12 UK-isolated lineage 2 isolates contained the same pSLT-associated MDR locus as the African lineage 2 reference strain D23580 (Kingsley et al., 2009). Four UK-isolated lineage 2 strains exhibited an atypical AMR gene profile; one isolate (U45) lacked the chloramphenicol resistance *catA* gene and another (U73) was not MDR and only carried the beta-lactamase gene $bla_{TEM1}$. A third isolate, U1, was inferred to have acquired resistance to fluoroquinolones via a mutation in the DNA gyrase subunit A gene, *gyrA*.

The fourth atypical UK-isolated lineage 2 isolate, U60, carried additional antibiotic resistance genes including extended-spectrum beta-lactamases (ESBL) $bla_{CTX\text{-}M\text{-}15}$ and $bla_{OXA\text{-}1}$, and genes conferring resistance to aminoglycosides, trimethoprim and tetracycline; *aac(6')-Ib-cr*, *dfrA-14*, *tet(A)-1* (Figure 4.4). Isolate U60 also encoded the

tellurium heavy metal resistance operon (*terBCDEF*). Comparison to lineage 2 reference strain D23580 identified a 29kb deletion in the pSLT-BT virulence plasmid (Kingsley et al., 2009), which corresponded to the conjugal transfer region. Additionally, sequence reads mapped to 97% of the IncHI2 pKST313 plasmid, a novel plasmid which has recently been reported in lineage 2 isolates from Kenya and is known to encode ESBL resistance loci (Kariuki et al., 2015). Interestingly, enhanced epidemiological data for isolate U60 showed recent travel to Kenya by the patient.

**A**

| Antimicrobial resistance type | Ampicillin 8 mg/L | Chloramphenicol 8 mg/L | Chloramphenicol 16 mg/L | Colomycin 2 mg/L | Sulphonamide 256 mg/L | Gentamicin 2 mg/L | Tobramycin 2 mg/L | Amikicin 8 mg/L | Streptomycin 16 mg/L | Tetracycline 8 mg/L | Tryimethoprim 2 mg/L | Nalidixic acid 16 mg/L | Ciprofloxacin 0.64 mg/L | Ciprofloxacin 0.5 mg/L | Ceftazidime 1 mg/L | Ceftazidime 2 mg/L | Cefotaxime 0.5 mg/L | Cefotaxime 1 mg/L | Cefoxitin 8 mg/L | Cefpirome 8 mg/L | Ertapenem 0.064 mg/L | Ertapenem 0.5 mg/L | Temocillin 128 mg/L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Typical UK ST313 isolate e.g. U2 | | | | | | | | | | | | | | | | | | | | | | | |
| Typical lineage II isolate e.g. D23580 | ■ | ■ | ■ | | ■ | | | | ■ | | ■ | | | | | | | | | | | | |
| Isolate U60 | ■ | ■ | ■ | | ■ | | ■ | | | | | ■ | ■ | ■ | | | ■ | | | | | | |

| AMR loci detected in genome sequence | |
|---|---|
| Typical UK ST313 isolate | *FII-S1; FIB-17, aac(6')-Iaa* |
| Typical lineage II isolate | *FII-S1; FIB-17, aac(6')-Iaa, TEM-1, aph(6)-Id, strB, strA, dfrA-1, sul-2, sul-1, catA-1* |
| Isolate U60 | *FII-S1; FIB-17, aac(6')-Iaa, TEM-1, aph(6)-Id, strB, strA, dfrA-1, sul-2, sul-1, catA-1, HI2 CTX-M-15, OXA-1, aac(6')-Ib-cr, dfrA-14, tet(A)-1, terF, terE, terD, terC, terB* |

**B**

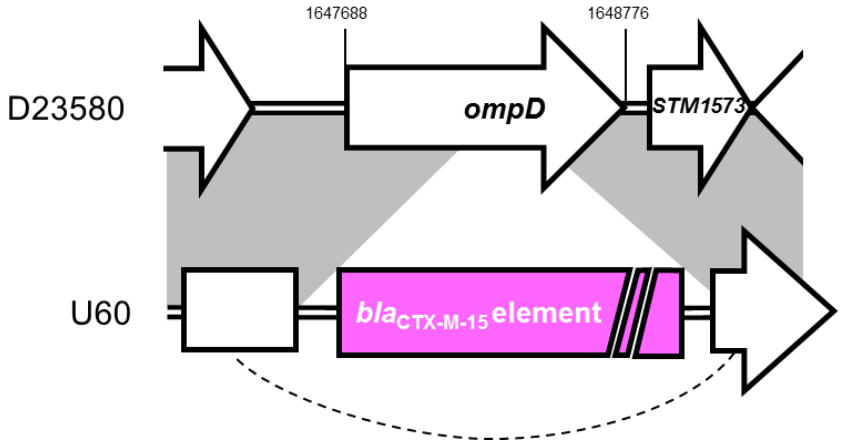



**Figure 4.4 Isolate U60 contains additional resistance genes including a *bla*$_{CTX-M-15}$ locus inserted into the chromosomal *ompD* locus.** A. Antimicrobial resistance typing data (2.3.5) and resistance genes (2.12.3) detected in genome sequences for isolate U60, compared to data for typical lineage 2 and UK-ST313 isolates. Heat map indicates resistance to the specified antibiotic concentration (black), or sensitivity (white). The resistance gene profiles are shown below. B. Schematic illustrating the insertion of *bla*$_{CTX-M-15}$ element into the chromosomal *ompD* locus in isolate U60.

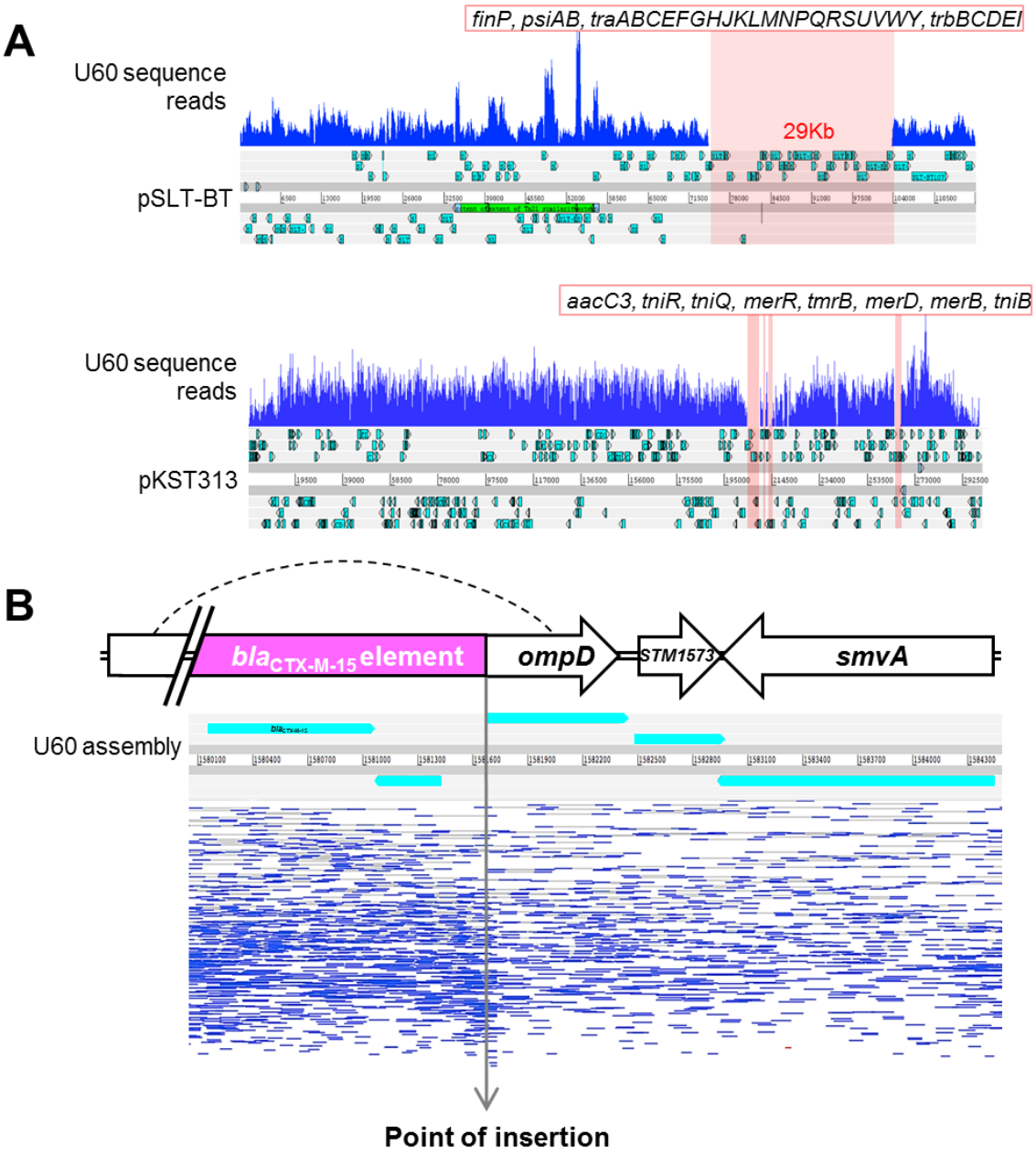



**Figure 4.5 Sequence-based validation of the insertion of a *bla*$_{CTX-M-15}$ locus into the chromosomal *ompD* locus of UK-isolates lineage 2 isolate U60.** A. Mapping coverage plots of U60 sequence reads on the pSLT-BT plasmid of D23580 (accession: LN794248 ) and the incH12 plasmid pKST313 (accession: FN432031). Read boxes indicate areas of zero coverage, suggesting the loss of these regions in isolate U60. The genes in the putative absent regions are listed above. B. Paired end read mapping of U60 sequence reads to the U60 genome assembly. An artemis screen shot shows part of the contig containing the disrupted ompD sequence and the *bla*$_{CTX-M-15}$ insertion. Sequence reads are represented as blue lines with grey lines joining each pair. Paired reads span the point of insertion, confirming the location of the *bla*$_{CTX-M-15}$ locus on the chromosome.

A more detailed analysis of the genome of isolate U60 identified a copy of the *bla*$_{CTX\text{-}M\text{-}15}$ gene inserted into the chromosome (location 1648104-1648109 on the D23580 reference genome), disrupting the *ompD* locus (Figure 4.4B, Figure 4.5). ESBL resistance genes have previously been reported in African ST313 isolates carrying plasmids such as pKST313 (Feasey et al., 2014; Kariuki et al., 2015), but this is the first report of a chromosomally-encoded ESBL resistance gene in *S.* Typhimurium ST313.

In order to assess the chromosomal similarity of UK-ST313 genomes to African ST313 genomes, the assembled UK-ST313 genome were compared to the African ST313 reference strain D23580 using BLAST (Figure 4.6). In agreement with published data (Kingsley et al., 2009), the majority of the core genome, including the *Salmonella* Pathogenicity Island (SPI) repertoire was conserved in the ST313 isolates in this study and in three ST19 gastroenteritis isolates (Figure 4.6). The African ST313 lineages carry two prophages, BTP1 and BTP5, that are absent from ST19 strains (Owen et al., 2017; Chapter 2). The entire BTP1 and BTP5 prophages were found in most ST313 isolates that belonged to African lineage 2 (12/13), but one UK-isolated lineage 2 strain, U68, lacked both prophages. The complete BTP1 and BTP5 prophages were not identified in any of the UK-ST313 isolates (Figure 4.6), though some isolates contained partial and fragmented identity to BTP1 and BTP5, indicating the presence of related prophages which may not occupy the same attachment site (Casjens and Thuman-Commike, 2011; Hatfull, 2008). As expected, the *st313-td* gene (Herrero-Fresno et al., 2014) was carried by all twelve lineage 2 strains isolated from the UK that contained prophage BTP1. Only 1/51 UK-ST313 isolates contained the *st313-td* gene (isolate U76), where it was located on a prophage with 90% identity to BTP1.

To confirm the conservation of chromosomal organization in the UK-ST313 isolates, a representative isolate U2 was re-sequenced by PacBio long-read sequencing (Chapter 2.11.2)**.** The assembly produced two closed circular contigs representing the chromosome and the pSLT virulence plasmid (Figure 4.7). Comparison with the ST313 lineage 2 reference genome D23580 identified no large chromosomal re-arrangements, deletions or duplications, and confirmed that the BTP1 and BTP5 attachment sites were unoccupied and did not contain additional prophages. No additional plasmids larger than the detection limit of the PacBio sequencing (~10kb in the case of the library preparation procedure used here- Chapter 2.11.2) were detected in isolate U2.

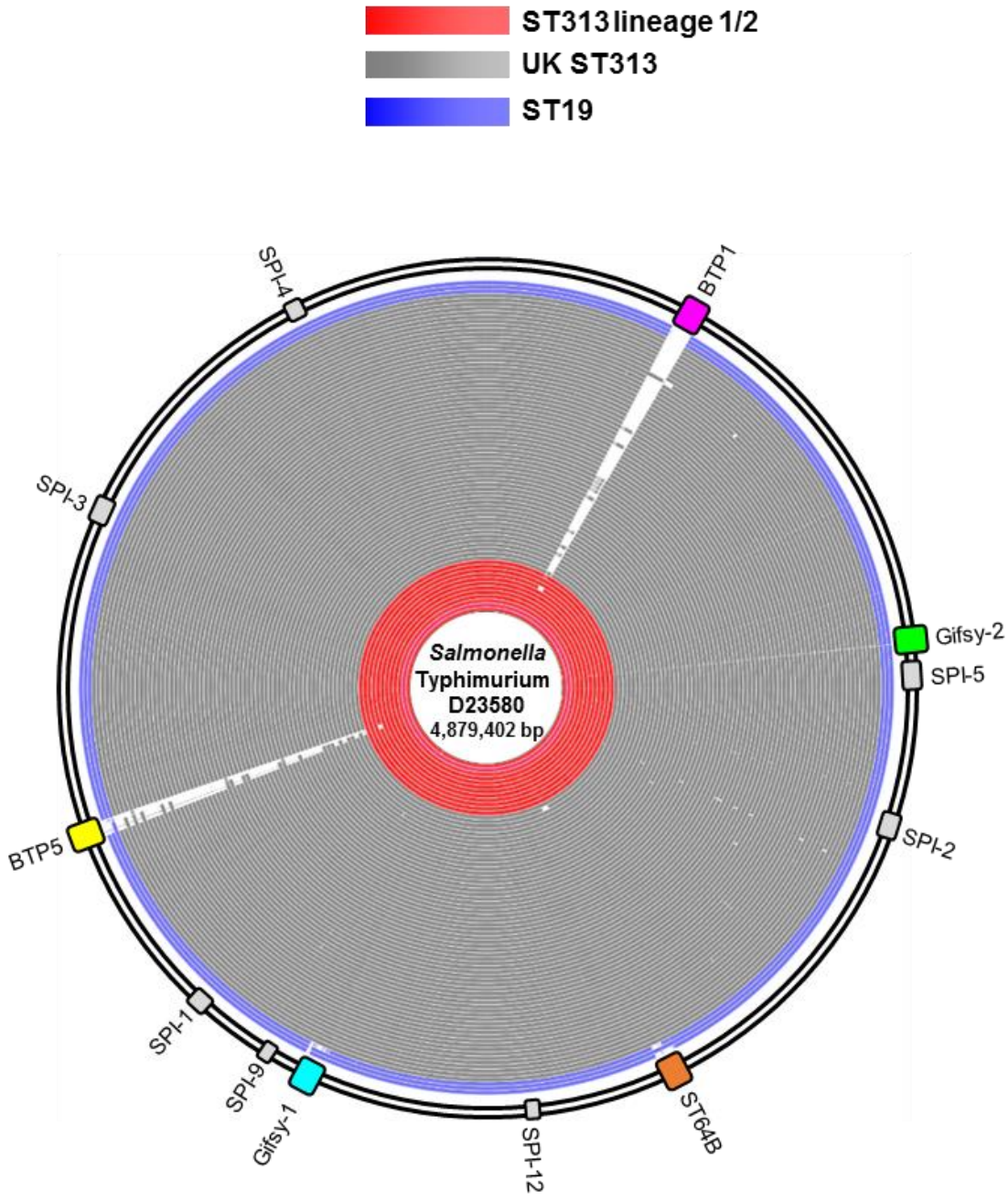



**Figure 4.6 UK-ST313 lack the BTP1 and BTP5 prophages that are characteristic of African ST313 lineages.** BLAST ring image showing BLAST comparison of all 76 UK-isolated ST313 genomes (red and grey rings) along with 3 reference ST19 strains (blue rings) 4/74, LT2 and DT104 against lineage 2 representative strain D23580. The position of the prophages (coloured blocks) and *Salmonella* pathogenicity islands (grey blocks) in lineage 2 strain D23580 are shown around the outside of the ring. This analysis was gererated using the BLAST ring image generator (BRIG) and methods are descibred in Chatpter 2.12.3.

## 4.5 Genome degradation and pseudogenes in UK and African ST313

The ST313 lineages 1 and 2 responsible for iNTS disease in Africa have undergone genome degradation (Kingsley et al., 2009; Okoro et al., 2015). The pseudogenes identified in lineage 2 representative strain D23580 were put into the context of the high-quality finished genome of UK-ST313 isolate U2 (Figure 4.7). The majority (34/44) of pseudogene mutations were conserved between strains U2 and D23580. The only pseudogenes associated with characterized genes present in lineage 2 that were functional in UK-ST313 strain U2 were *macB*, *sseI* and *lpxO* (Figure 4.8). Of the 19 isolates (originating from 12 patients) of lineage 2, all had disruptions to the *lpxO*, *ratB*, *allP*, *allB*, *pagO* and *pipD* genes, confirming that the genome degradation reported in African lineage 2 is conserved in lineage 2 strains isolated in the UK.

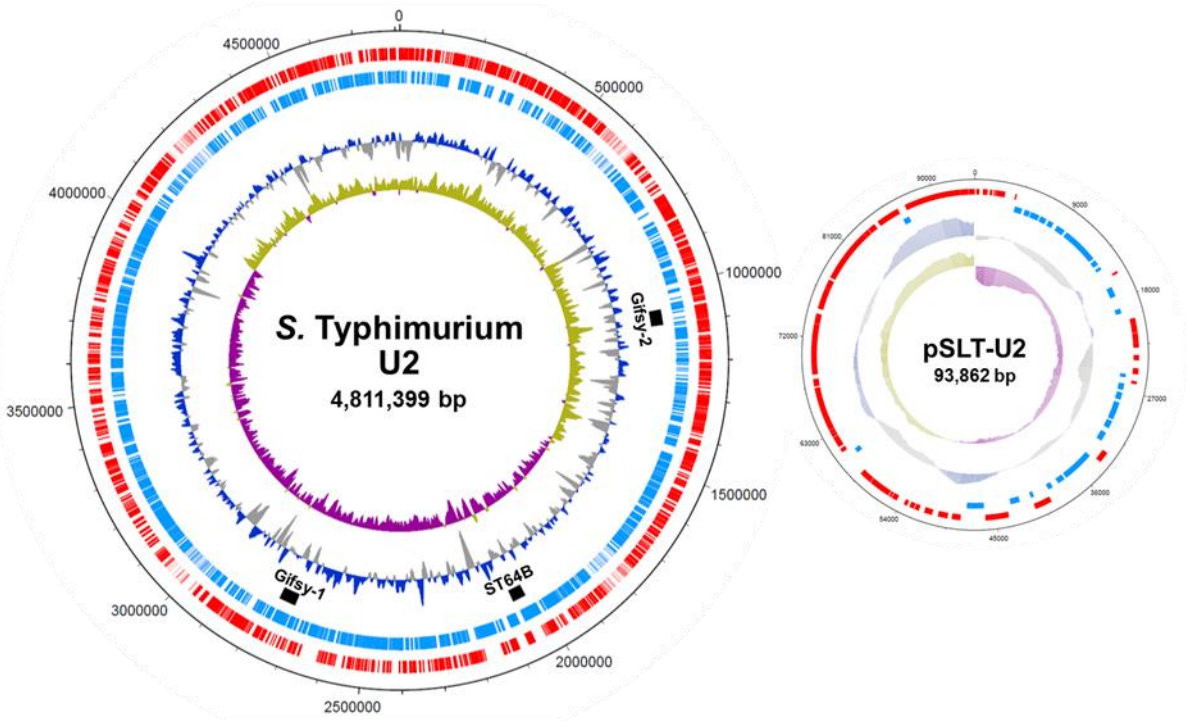



**Figure 4.7 Circular representation of the finished genome of UK-ST313 representative strain U2, showing the chromosome and the pSLT-U2 virulence plasmid.** The two outer most circles show the position of annotated CDS on the sense (red) and antisense (blue) strands respectively. For the chromosome, the location of the 3 prophages; Gifsy-1, ST64B and Gifsy-2 are shown as black bars. Finally, the G+C content (by 10kb windows) and the GC skew ([G-C]/[G+C] by 10 kb windows) is shown in the two center-most circles. The Genbank accession number for the U2 genome is LT855376 and U2 virulence plasmid accession number is LT855377.

| STM | STMMW | Gene name | Description | D23580 | U2 |
|---|---|---|---|---|---|
| 0157 | 01631 | yacH | putative outer membrane protein | | |
| 0522 | - | allP | allantoin transport protein | | |
| 0523 | - | allB | allantoinase | | |
| 0834 | 08851 | ybiP | putative Integral membrane protein | | |
| 0942 | 09551 | macB | putative ABC superfamily transport protein | | |
| 1014 | 10251 | | Probable regulation protein | | |
| 1023 | 10351 | | hypothetical on gifsy-2 prophage | | |
| 1051 | 10631 | sseI | type III secretion effector protein (SPI-2) | | |
| 1092 | - | orfX | putative cytoplasmic protein | | |
| 1093 | - | | putative cytoplasmic protein | | |
| 1094 | 11041 | pipD | similar to dipeptidase A | | |
| 1228 | 12371 | | putative periplasmic protein | | |
| 1516 | 15161 | ydeE | putative MFS family transport protein | | |
| 1548 | 15471 | | putative ribosyltransferase-isomerase | | |
| 1549 | - | | putative translation initiation inhibitor | | |
| 1550 | - | | putative cytoplasmic protein | | |
| 1551 | - | | putative cytoplasmic protein | | |
| 1551' | - | | hypothetical protein | | |
| 1552 | - | | putative cytoplasmic protein | | |
| 1637 | 16321 | | putative inner membrane protein | | |
| 1862 | - | pagO | PhoP activated gene | | |
| 1863 | - | | putative inner membrane protein | | |
| 1864 | - | | putative inner membrane protein | | |
| 1865 | - | | Putative DNA invertase | | |
| 1866 | - | | hypothetical | | |
| 1868 | - | mig-13 | phage tail assembly protein | | |
| 1868' | - | | lytic enzyme | | |
| 1869' | - | | hypothetical protein | | |
| 1870 | - | | RecE-like protein | | |
| 1896 | 18781 | | putative cytoplasmic protein | | |
| 1940 | 19221 | | putative cell wall associated hydrolase | | |
| 2238 | 22681 | | putative phage protein | | |
| 2514 | 25311 | ratB | Secreted protein | | |
| 2589 | 26091 | | hypothetical in Gifsy-1 prophage | | |
| 2680 | 26941 | | putative cytoplasmic protein | | |
| 2932 | 28951 | ygbE | putative inner membrane protein | | |
| 3012 | 29741 | | putative transcriptional regulator | | |
| 3075 | 30361 | | putative ABC-type cobalt transport system | | |
| 3355 | 33531 | | tartrate dehydratase | | |
| 3624 | 36131 | yhjU | putative inner membrane protein | | |
| 3745 | 37341 | | putative cytoplasmic protein | | |
| 3768 | 37571 | | putative selenocysteine synthase | | |
| 4196 | 41451 | | putative cytoplasmic protein | | |
| 4286 | 42371 | lpxO | putative dioxygenase | | |

**Figure 4.8 The majority of pseudogenes identified in lineage 2 strain D23580 are conserved in UK-ST313 representative strain U2.** Heatmap adapted from Kingsley et al (2009) showing genome degradation in ST313 strain D23580 (first heatmap column) in the context of strain U2 (final column). Grey indicates pseudogenes conserved in both strains, whilst red indicates genes which are not degraded and likely to be functional in strain U2.

## 4.6 UK and African ST313 strains share key phenotypes

Several studies have associated the ability of ST313 to cause iNTS disease with particular phenotypic characteristics, such as the lack of RDAR morphotype formation, reduced swimming motility and the inability to produce catalase at stationary phase (Carden et al., 2015; Ramachandran et al., 2015; Singletary et al., 2016). Logically if the reported phenotypes are important for, and have been selected for, in the sub-Saharan African niche, the phenotypes should be absent from non-African, UK ST313 isolates. I investigated whether the phenotypic characteristics were associated with UK-ST313 strains using a subset of 16 UK-isolated ST313 that included 13 UK-ST313 isolates and 3 lineage 2 isolates. The phylogenetic context of these 16 isolates is shown in Figure 4.9 Lineage 2 representative strain D23580 and ST19 representative strain 4/74 were included as positive and negative controls.

The swimming motility of UK-isolated ST313 was highly variable between isolates (Figure 4.10A). One lineage 2 strain, U1, appeared to show low levels of motility but this strain was observed to have a significant growth defect (data not shown). The ST313 lineage 2 representative strain D23580 was less motile than ST19 strain 4/74, consistent with previous reports (Ramachandran et al., 2015). However, there was no apparent correlation between motility (as measured by migration diameter) and the phylogenetic context of the lineage 2 strains and the UK-ST313 strains.

The SNP mutation in the *katE gene* was reported to be responsible for the lack of catalase activity in ST313 lineage 2 (Singletary et al., 2016). All 16 UK-isolated strains were negative for stationary phase catalase activity, as was the lineage 2 representative strain D23580 (Figure 4.10B). In contrast, ST19 strain 4/74 showed high levels of stationary phase catalase activity, consistent with previous findings (Singletary et al., 2016).

The RDAR morphotype of *Salmonella enterica* is linked to resistance to desiccation and exogenous stresses (White et al., 2006). African lineage 2 ST313 are reported to be defective in RDAR morphotype formation due to a truncated BcsG protein generated by a premature stop codon (Singletary et al., 2016). All the UK-isolated strains and the African lineage 2 reference strain D23580 did not exhibit the RDAR morphotype. Only the ST19 strain 4/74 exhibited the RDAR morphotype (Figure 4.10C).

These experiments did not identify any phenotypic differences that distinguished the UK-ST313 isolates from ST313 lineage 2. Future work is needed to identify African

ST313-specific phenotypes, which may provide clues to the evolutionary selection pressures that contributed to the niche adaptation of *S.* Typhimurium ST313 in Africa.

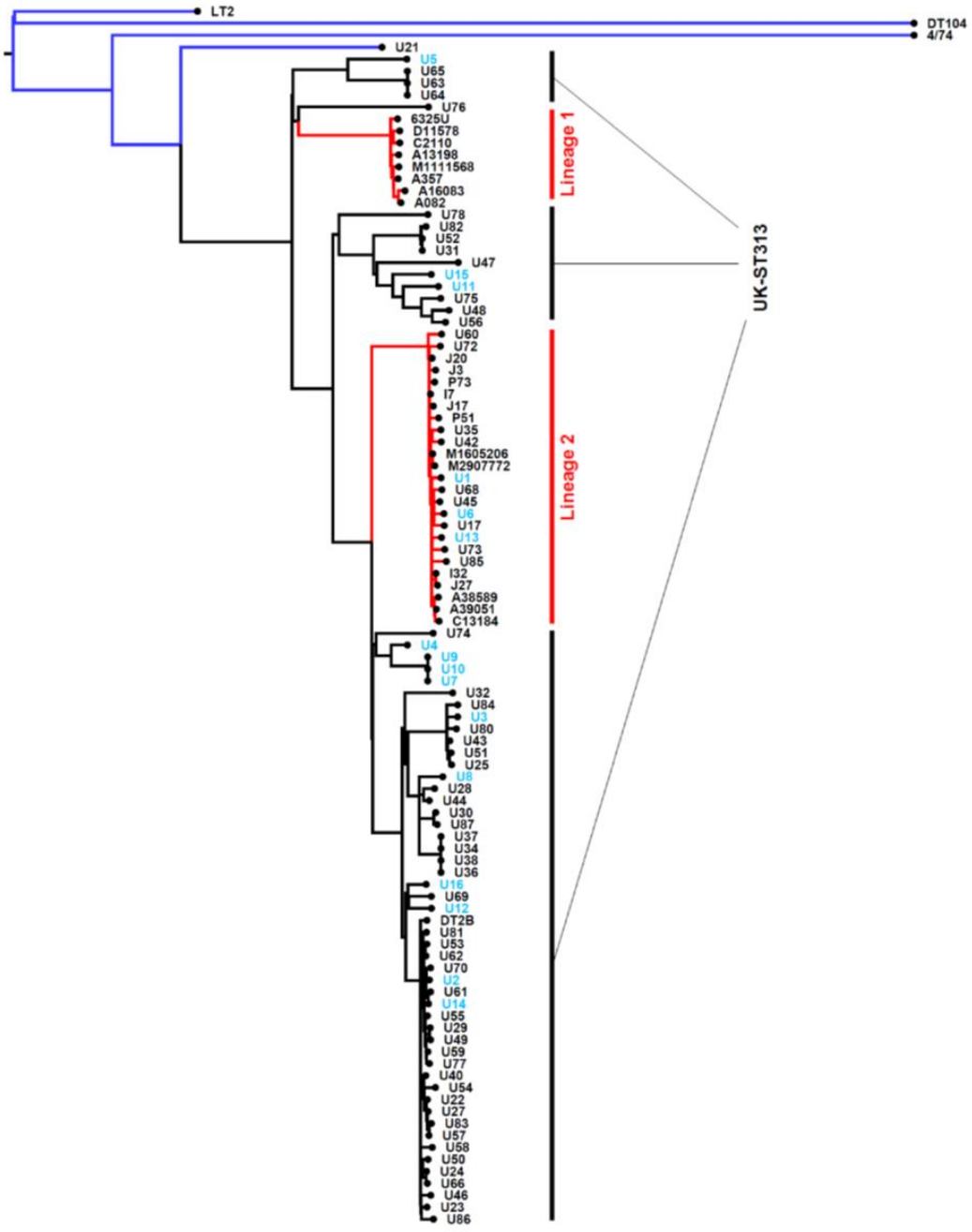



**Figure 4.9 Dendrogram representation of the phylogenetic tree shown in Figure 4.1 but with fully labelled tips to highlight the 16 UK-isolated ST313 strains included in phenotypic testing (tips labelled in light blue text).** Lineage 1 and 2 are shown with red branches, ST19 strains with blue branches and UK-ST313 with black branches.

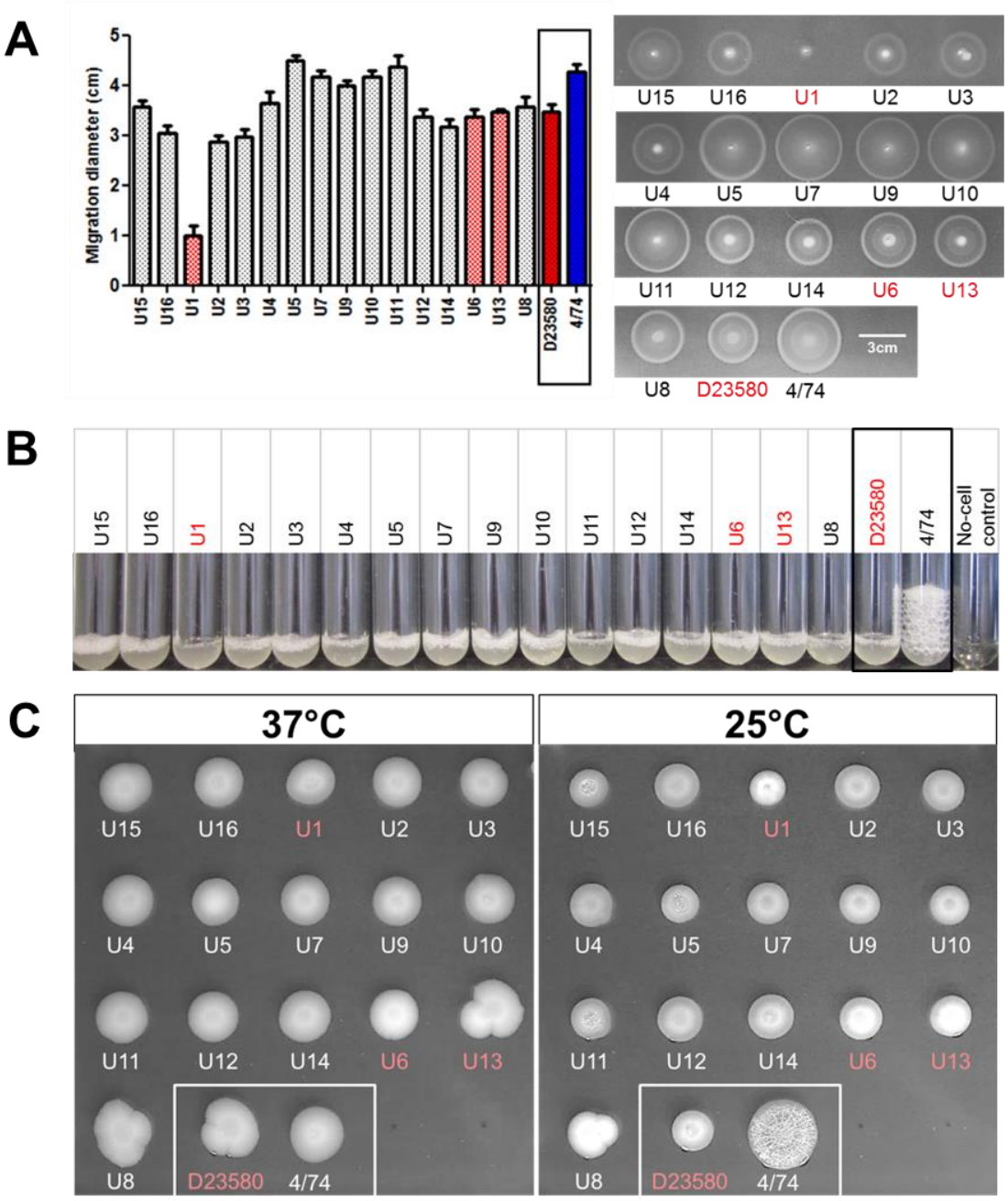



**Figure 4.10 In vitro phenotypes of a subset of UK-isolated ST313 strains in the context of the representative ST313 lineage 2 and ST19 strains D23580 and 4/74.** UK-isolated strains that belong to African lineage 2 (U1, U6 and U13) are highlighted in red throughout. A. Motility as represented by migration diameter after 5 hours (average of 3 replicates is shown together with error bars representing standard deviation) (2.3.2). A representative plate is shown, right. B. Stationary phase catalase activity represented by bubble column height after 5 minutes exposure to 20µl 20% $H_2O_2$ (2.3.4). C. RDAR morphology assay. RDAR phenotype forms after prolonged incubation at 25°C but not at 37°C (2.3.3).

## 4.7 UK-ST313 strains carry BTP1-like and BTP5-like prophages

Alignment of the UK-ST313 genomes to the lineage 2 representative strain D23580 showed that some UK-ST313 strains carry prophages with mosaic homology to BTP1 and BTP5 (Figure 4.6). To reveal the true extent of homology between the BTP1 and BTP5-like prophages detected, prophages were manually extracted from a subset of 12 UK-ST313 draft genome assemblies that were representative of the phylogenetic diversity of all UK-ST313s. Draft genome assemblies consist of a number of short assembled fragments (contigs), which increase in length and frequency depending on the quality of the genome assembly. To extract prophages sequences, the contigs containing the phage *attR* and *attL* sites (proximal sites to either end of the prophage) were identified, and the sequence between these sites was determined to be a putative prophage. In some cases, it was not possible to extract prophage sequences, as the BTP1 or BTP5 attachment sites were separated across contigs. However, in most cases, the BTP1 or BTP5 attachment site was either intact on a single contig (indicating no prophage), or separated by a length corresponding to a putative prophage on a single contig (indicating the presence of a prophage occupying the attachment site). Additionally, to control for the possibility that the BTP1 and BTP5-like prophages in the UK-ST313 strains might have distinct attachment sites to BTP1 and BTP5, large contigs (>50 kb) were extracted which contained homology to BTP1 or BTP5, and contained core genome sequence (non-phage). Using this approach, 3 BTP1 or BTP5 like prophages were identified in a total of 10 of the 12 UK-ST313 genomes.

Two UK-ST313 strains, U15 and U8 contain a prophage with 71% identity to BTP1 at the nucleotide level (Figure 4.11A). This prophage, denoted BTP1$^{UK-1}$, contains almost identical structural genes to BTP1, but differs in the immunity, tail and *gtr* operon regions. The attachment site of BTP1$^{UK-1}$ was the same as that of BTP1 (data not shown).

Two UK-ST313 strains, U12 and U5 contain a prophage with 48.5% identity to BTP1 at the nucleotide level (Figure 4.11B). This prophage, denoted BTP1$^{UK-2}$, contains many of the same structural genes as BTP1, but has no similarity to the tailspike and *gtr* operon of BTP1. Interestingly, prophage BTP$^{UK-2}$ carries a repressor region with considerable identity to the repressor region of BTP1 including the *ST313-td* gene, but contains a different *cI* repressor and *cro* allele to BTP1. Prophage BTP1$^{UK-2}$ was found to be > 99% identical to a temperature phage called SE1. Phage SE1 was

originally isolated from a strain of *S.* Enteritidis, and mediates generalised transduction in *S.* Typhimurium (Llagostera et al., 1986). The attachment site of the prophage BTP1$^{UK-2}$ (SE1) was the same as BTP1, as the core chromosomal sequence bordering the prophage on this contig corresponds to the attachment site of the BTP1 prophage in D23580.

Finally, a prophage with 57% identity to BTP5 at the nucleotide level was found in 6 UK-ST313 strains (U11, U16, U7, U9, U10, U4) (Figure 4.12). This prophage, denoted BTP5$^{UK-1}$, encoded identical capsid, tail and lysis genes to BTP5, but did not encode genes homologous to the replication, immunity or integrase genes of BTP5. The BTP5 attachment site was intact in strains containing prophage BTP5$^{UK-1}$. The attachment site of the prophage was found to be between the CpxQ sRNA and the *fieF* gene (STMMW_40261), though neither element was interrupted by the integration of the prophage (data not shown).

The presence of the novel BTP1-like and BTP5-like prophages discovered in the UK-ST313 genomes did not correlate with the phylogeny (Figure 4.13) i.e. prophages were not associated with particular phylogenetic clusters, showing that the presence of these prophage is not under high selection pressure in the UK-ST313 niche. For example, BTP5$^{UK-1}$ was present in UK-ST313 isolates spanning the diversity of the tree, which may suggest that, though there is considerable genetic diversity amongst the UK-ST313 isolates in this study, they exist within the same ecological niche, and are therefore exposed to the same phage population.

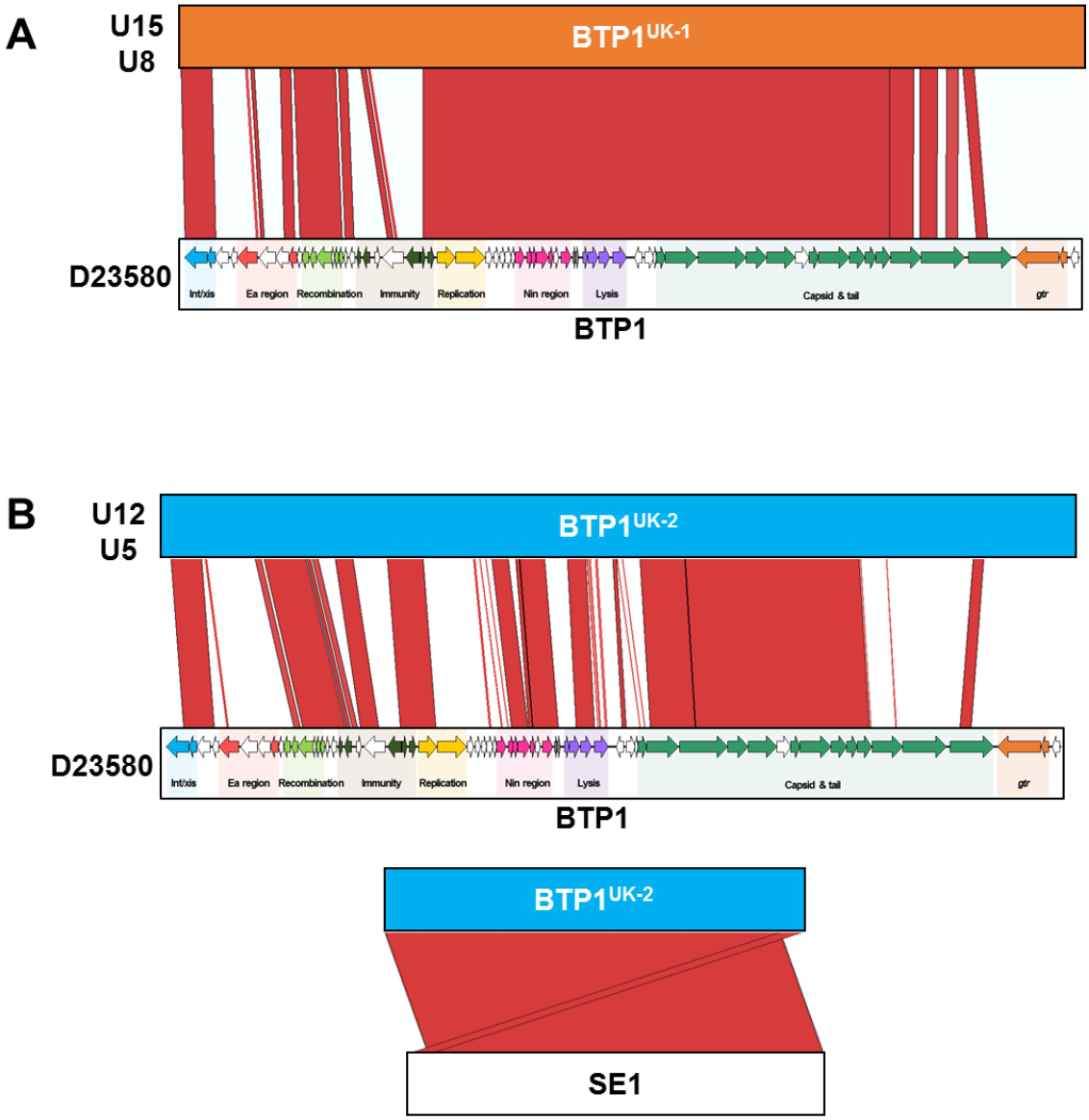



**Figure 4.11 Novel BTP1-like prophages identified in the genomes of UK-ST313 isolates.** A. Prophage BTP1$^{UK-1}$ was identified in UK-ST313 strains U15 and U8 and shares 71% identity to prophage BTP1. B. Prophage BTP1$^{UK-2}$ was identified in UK-ST313 strains U12 and U5 and shared 48.5% identity with BTP1. Prophage BTP1$^{UK-2}$ was found to be more than 99% identical to phage SE1 derived from *Salmonella* Enteritidis (inset). Comparative genomic analysis was conducted as described in 2.12.5.

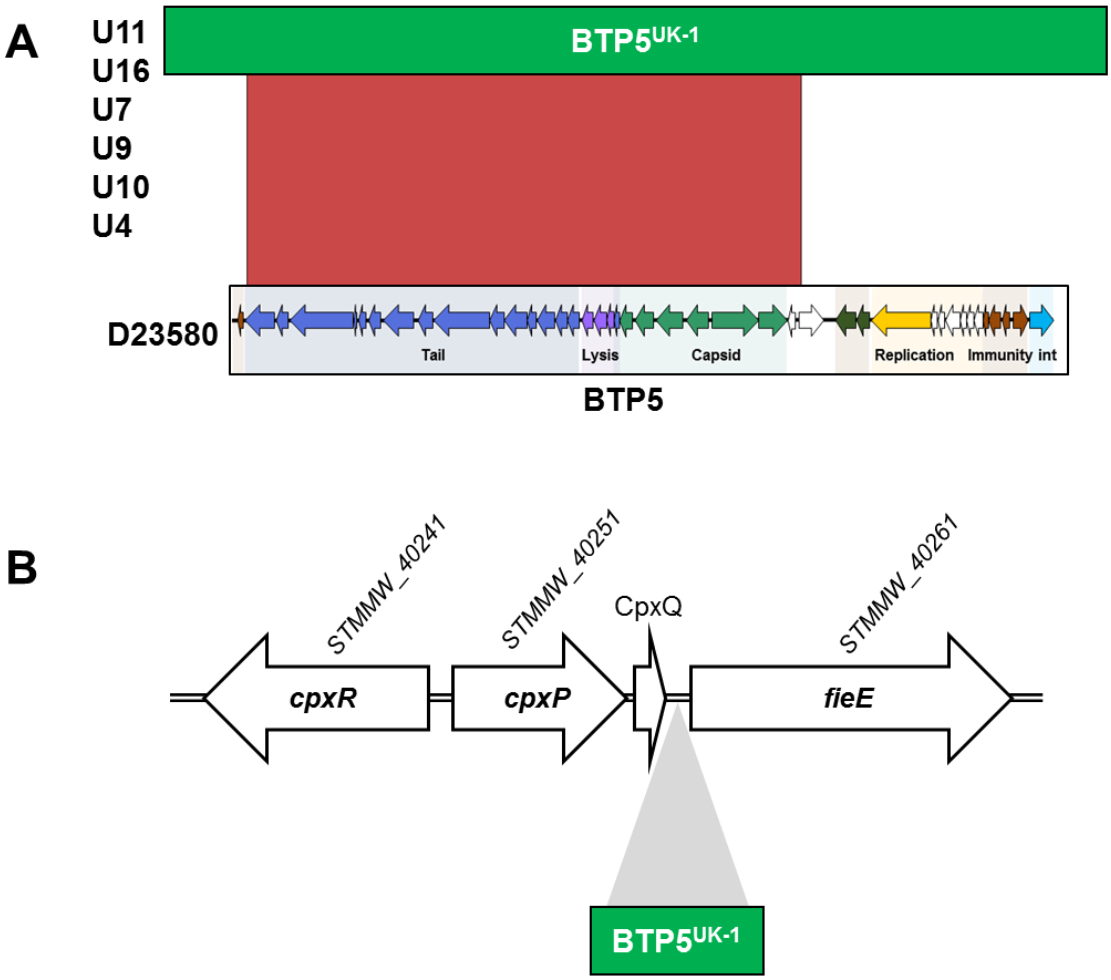



**Figure 4.12 A novel BTP5-like prophage identified in the genomes of UK-ST313 isolates.** A. Prophage BTP5$^{UK-1}$ was identified in UK-ST313 strains U11, U16, U7, U9, U10 and U4. BTP5$^{UK-1}$ was 57% identical to BTP5. B. Prophage BTP5$^{UK-1}$ occupies a distinct attachment site to BTP5, between the CpxQ sRNA and the *fieE* gene. Comparative genomic analysis was conducted as described in 2.12.5.

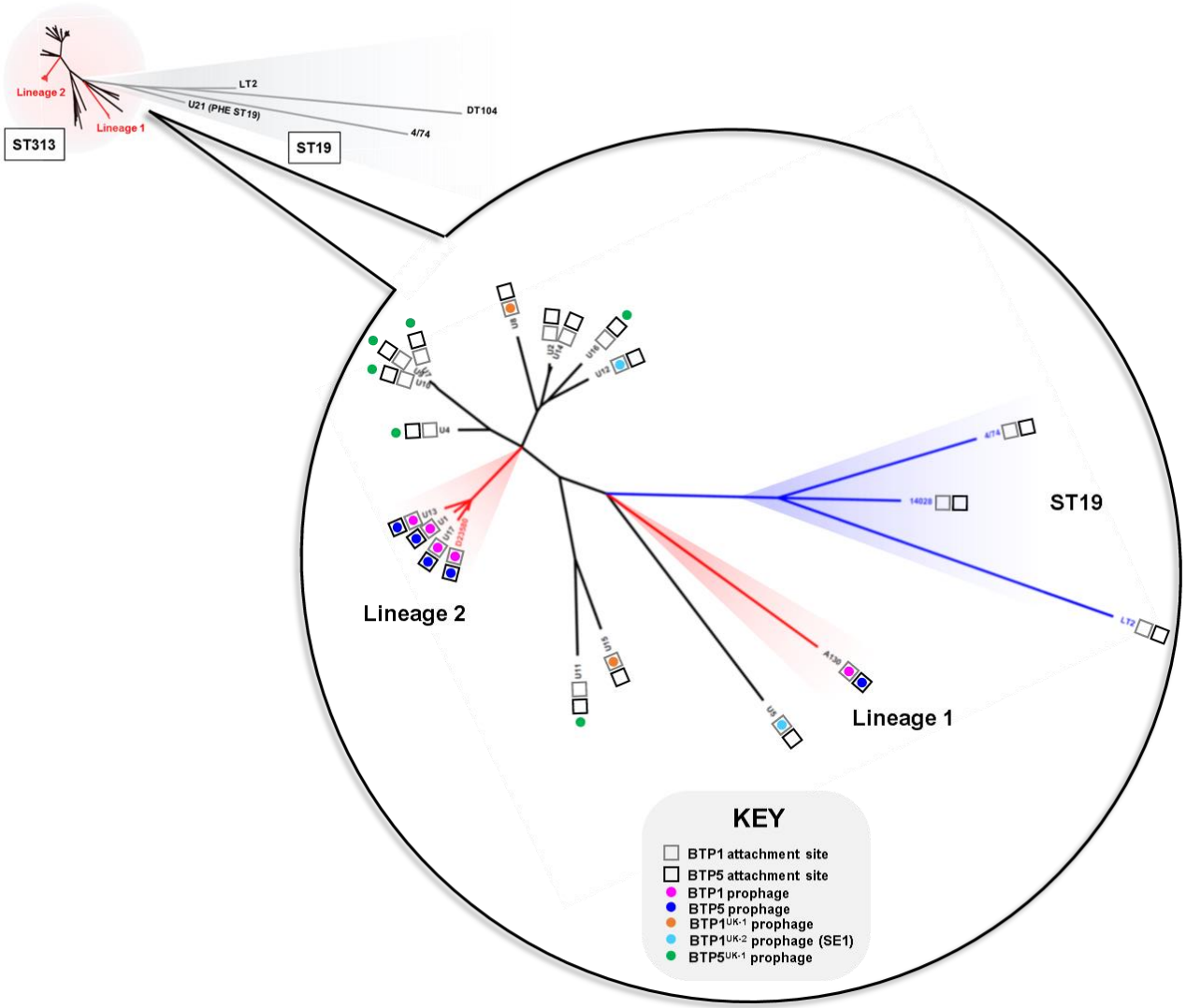



**Figure 4.12 The phylogenetic distribution of the BTP1-like and BTP-5 like prophages identified in the UK-ST313 genomes.** The BTP1 & BTP5 attachment site are represented as grey and black un-filled squares. Presence of a prophage (indicated by a filled coloured circle), in an attachment site is indicated by the position of the circle inside the square. Circles outside of the attachment site boxes represent prophages occupying attachment sites distinct from that of BTP1 and BTP5. Prophages were identified by comparative genomic analysis as described in 2.12.5.

## 4.8 Discussion

Recent reports of iNTS disease have been associated with *Salmonella* Typhimurium ST313 in sSA (Kingsley et al., 2009; Okoro et al., 2015) and suggested that ST313 was geographically restricted to sSA. This prompted investigation into the presence of ST313 amongst *S.* Typhimurium isolates from the UK.

It was discovered that 2.7% of *Salmonella* Typhimurium isolates referred to Public Health England are of MLST type ST313 and that this sequence type is heterogeneous in terms of clinical presentation, genomic characteristics and epidemiology. The UK-isolated ST313 strains are predominantly fully antimicrobial-susceptible and cause gastroenteritis. This was also the case with the recently described ST313 strains from Brazil (Almeida et al., 2017), raising the possibility that the Brazilian isolates are more closely related to the UK-ST313 reported here than to African lineage 2.

Significant correlation was identified between travel to Africa and infection with the previously described African-associated, ST313 lineage 2. The amount of diversity between the lineage 2 ST313 isolates from the UK (Figure 4.4) was not consistent with an immediate, common source of exposure. Instead, it is more probable that these isolates were linked to travel to Africa, a hypothesis that was supported by epidemiological data that linked the isolates to African travel (Table 4.1).

This study revealed novel diversity within ST313, which was previously restricted to two African lineages that had exhibited recent clonal expansion (Okoro et al., 2012). The African lineages are placed into an evolutionary context by showing that lineage 1 and 2 do not form a monophyletic group within ST313, which is suggestive of two separate introductions of ST313 into sSA. African lineages 1 and 2 diverged from their MRCA with UK-lineages around 1796 and 1903 respectively (4.3). These findings reflect the limitations of classifying bacterial pathogens simply on the basis of sequence type and show that in the post-genomic era, the resolution offered by MLST may not be sufficient to describe epidemiologically relevant population structures of *Salmonella*.

It has been estimated that 9.2% of cases of Salmonellosis in the EU can be attributed to international travel and therefore sequencing *Salmonella* isolated in Europe can provide valuable information regarding the global diversity of *Salmonella* associated with human disease (Guerin et al., 2007; Pires et al., 2014). The genome of one UK-isolated lineage 2 isolate associated with travel to Kenya, U60, contained sequences

with high nucleotide similarity to a recently described IncHI2 plasmid, pKST313, that was carried by ceftriaxone-resistant ST313 isolates from Kenya (Kariuki et al., 2015). Until now the *bla*$_{CTX-M-15}$ gene has only been found to be plasmid-associated in *Salmonella*. A *bla*$_{CTX-M-15}$ gene carried on the chromosome was discovered in UK-isolated lineage 2 strain U60, causing disruption of the *ompD* locus which has two implications. Firstly, chromosomal integration ensures stability of ESBL-resistance even in the absence of the plasmid. Secondly, *ompD* encodes an outer membrane porin of *S.* Typhimurium that is highly immunogenic (Gil-Cruz et al., 2009) and absent from *S.* Typhi. Accordingly, the disruption of *ompD* could enhance the reported "stealth" phenotype of ST313 lineage 2 infection (Carden et al., 2015). Notably, OmpD has been proposed as a vaccine target for iNTS (Ferreira et al., 2015) and the absence of OmpD from African ST313 populations could have implications for future iNTS vaccine development.

People infected with ST313 lineage 2 in the UK were significantly more likely to suffer from invasive disease than patients infected with UK-ST313 isolates, though the HIV / immunosuppression-status was not known for these patients. This observation provided an excellent opportunity to use comparative genomics to put genetic findings that have been linked the pathology of lineage 2 ST313 into the context of closely related, gastrointestinal-associated strains. The only genetic characteristics found to be common to both lineages 1 and 2 and absent from the UK-ST313 genomes were the BTP1 and BTP5 prophages and the plasmid-associated MDR loci. The two African lineages do not share a common ancestor that carried either prophage, suggesting independent acquisition of BTP1 and BTP5 by ST313 lineage 1 and 2. Whilst the MDR loci of lineage 1 and 2 confer similar patterns of AMR, they are genetically distinct indicating independent origins. The maintenance of the prophage and plasmid-encoded accessory genome in two distinct ST313 lineages in Africa implies a strong selection pressure that caused convergent evolution of the two African lineages. In contrast, there was evidence for an assortment of distinct prophage repertoires in the UK-ST313 isolates, indicating an absence of selection for specific mobile elements.

Aside from the addition of mobile genetic elements and virulence factors, genome degradation by the accumulation of pseudogenes and deletion events accompanies adaption to a more invasive lifestyle (Georgiades and Raoult, 2011; Nuccio and Bäumler, 2014). Initial analysis of the African ST313 representative strain D23580 genome identified 23 pseudogenes compared to the 6 present in ST19 strain SL1344

(Kingsley et al., 2013). However, the majority of the genome degradation found in lineage 2 strain D23580 was conserved in UK representative strain U2. The only pseudogenes associated with characterized genes that were found to be specific to African lineage 2 ST313 were the SPI2-secreted effector gene *sseI,* lipid A modification gene *lpxO* and macrolide efflux pump gene *macB,* each of which could potentially play a role in infection dynamics (Andersen et al., 2015; Carden et al., 2017; Gibbons et al., 2005).

A number of the *in vitro* phenotypes that have been reported for lineage 2 ST313, which could contribute to a host-adapted lifestyle (Carden et al., 2015; Ramachandran et al., 2015; Singletary et al., 2016), were examined in the UK-ST313 to look for an African lineage-specific phenotype. Swimming motility was highly variable amongst the strains tested, and UK-ST313 isolates behaved identically to African lineage 2 isolates in the catalase and RDAR morphotype assays. No African lineage-specific phenotypic characteristics were detected, suggesting that reduced motility, defective catalase activity and loss of RDAR formation are not directly linked to iNTS disease.

A key contributing factor to iNTS disease is host immunosuppression (Feasey et al., 2012), and one limitation of this retrospective study was that the underlying health status of the patients was unknown. This study does highlight the extraordinary epidemiological insights that routine genomic surveillance of pathogens by public health agencies can offer and the ability to understand the pathogenesis of novel pathovars.

This chapter has described previously unknown diversity in the ST313 sequence type that gives an insight into the convergent evolution towards niche specialization that has occurred in African ST313 lineages 1 and 2. The routine genomic surveillance of pathogens described here is now being adopted internationally and will bring an unprecedented ability to monitor emerging threats, such as the arisal of extended-spectrum beta-lactamase resistance. More generally, whole-genome sequencing of clinical isolates represents a new window with which to view the epidemiology and microbiology of infectious diseases.

# Chapter 5

## Investigating novel prophages BTP1 & BTP5 at the transcriptional level

## 5.1 Introduction

The BTP1 and BTP5 prophages are a defining feature of African *S.* Typhimurium ST313 isolates belonging to both lineages, and ST313 genomes lacking either or both of the prophages are rare (Okoro et al., 2015). Findings described in Chapter 4 suggested that the BTP1 and BTP5 prophages were acquired independently by the two African lineages of *S.* Typhimurium ST313, suggesting a strong selection pressure for this particular prophage complement in the sub-Saharan African niche, and representing convergent evolution of the two lineages in Africa (Chapter 4.5). What is the selection pressure that drove the acquisition and is driving the conservation of these prophages? The aim of this chapter is to put the prophages characterised in Chapter 3 into context within the genome of African *S.* Typhimurium ST313, and to explore how the novel prophages may contribute to the biology of the lysogen, at the transcriptomic level.

### 5.1.1 Acknowledgement of the specific contribution of collaborators to the results described in Chapter 5.

The work described in this chapter is unpublished. I acknowledge the following contribution of collaborators to the results described in this chapter. Unless specified below, all work was completed by the Author.

| | |
|---|---|
| **Rocío Canals & Disa Hammarlöf**<br>University of Liverpool, UK | Collection of all D23580 WT RNA-seq & dRNA-seq samples<br>(Canals et al., 2018; Hammarlöf et al., 2017) |
| **Will Rowe**<br>University of Liverpool, UK | Generation of differential gene expression values (CPMs) for prophage mutant RNA-seq experiments (described in 2.12.4) |

## 5.2 The D23580 prophage transcriptome

As in previous analysis, strain D23580 was used as a representative of African *S.* Typhimurium as it was the first strain for which a completed genome sequence was made publically available. RNA-seq data from Canals et al. (2018) was used to analyse the transcriptomes of the five D23580 prophages in 16 infection-relevant conditions, following a previously described approach (Kröger et al., 2013). The suite of conditions in which the bacteria were grown before RNA extraction were designed to represent the different environments that pathogens experience during infection of a mammalian host. The data also include RNA-seq experiments from strain D23580 during infection of murine macrophages, following previously described methods

(Srikumar et al., 2015). The transcriptomes reveal the expression patterns of prophage genes whilst the prophages are in a lysogenic state within the cell.

As well as looking at gene expression profiles, the transcription start sites (TSS) located in the prophage regions were also analysed. The approach pioneered by Kröger et al., 2012 to define TSS known as differential RNA-seq (dRNA-seq) was used to locate the TSS in the prophage regions (Hammarlöf et al., 2017). TSS are important as they indicate the activity of promoters. Together, these data reveal the transcriptional landscape of five distinct prophages during lysogeny. Prophage genes which are expressed during lysogeny are frequently associated with lysogenic conversion of the host, such as genes encoding virulence factors or superinfection immunity factors (Casjens and Hendrix, 2015).

### 5.2.1 BTP1 prophage transcriptome

The transcriptome map (Figure 5.1) of BTP1 in a lysogenic state showed relatively little transcription, interspersed with a small number of highly transcribed regions. Five genes (excluding the two central tRNAs) showed particularly high expression (Figure 5.2); *ST313-td* (STMMW_03531), *cI* (*STMMW_03541*), *pid* (*STMMW_03751*), *gtrC* (*STMMW_03911*) and *gtrA* (*STMMW_03921*). The *cI* and the *ST313-td* genes represent an operon, as only one TSS was associated with the two genes, and RNA-seq reads mapped across the 88 bp intergenic region. The *cI-ST313-td* operon showed a high level of expression in most conditions, consistent with the putative role of the CI protein as the prophage repressor. The *pid* gene, linked to maintenance of the pseudolysogenic state in phage P22 (Cenens et al., 2013) also showed a high level of transcription in the majority of conditions (TPM >10 in all conditions) (Figure 5.2). Lastly, a two-gene operon encoding *gtrA* and *gtrC* showed high expression in the majority of conditions (TPM >10 in 16/17 conditions). *gtrA* encodes a bactoprenol-linked glucosyltranslocase (also known as the 'flippase') and *gtrC* encodes a putative acetyltransferase that mediates the addition of an acetyl group to the rhamnose subunit of the O-antigen (Kintz et al., 2015). The *gtr* operon is commonly found in P22-like phages and functions to modify the O-antigen component of the lipopolysaccharide to inhibit superinfection of the lysogen (Davies et al., 2013). The high expression of the BTP1 *gtr* operon in the majority of conditions tested suggests that unlike the gtr operon of phage P22 (Wallecha et al., 2002) the acetylation of the rhamnose LPS moiety is not a phase-variable phenotype and occurs in all growth conditions (Kintz et al., 2015).

The structural (capsid and tail) and lysis genes (late lytic genes) showed some, albeit low level, expression in the majority of conditions tested (Figure 5.1, Figure 5.2). Work described in Chapter 3 showed that the BTP1 prophage exhibits an extremely high level of spontaneous induction within the cellular population, and BTP1 is estimated to undergo spontaneous induction in 0.2% of bacterial cells. As the transcriptome generated by RNA-seq represents the average gene expression across an entire bacterial population, I hypothesise that the apparent low-level late lytic gene transcription is derived from extremely high level transcription of these genes within the fraction of the population in which the BTP1 prophage is spontaneously induced. This hypothesis is discussed in the context of the other prophages of D23580 in section 5.4. Late lytic-gene expression was highest in the ESP and oxygen shock conditions, which may suggest that proportionately greater spontaneous induction occurs in these growth conditions, though this possibility has not been tested experimentally.

The prophage late lytic genes were transcribed in one long operon that begins upstream of the lysis gene cluster, directly after the central tRNA genes. Due to position of the tRNA genes upstream of this transcript, the exact position of the TSS for this transcript is not clear. As well as primary TSS, operons may also exhibit secondary TSS, which are located downstream of the initial TSS and produce shorter transcripts (Kröger et al., 2012). Within the BTP1 prophage capsid gene cluster there were three putative secondary TSSs on the coding strand. The BTP1 tailspike gene contained a number of secondary TSS, and showed a different transcriptional pattern to the rest of the late lytic gene operon (Figure 5.1). Antisense transcription of the tailspike gene is investigated in Chapter 6.4. A total of 10 putative non-coding RNAs (ncRNAs) were identified in the BTP1 transcriptome (Figure 5.1) (Hammarlöf et al., 2017). One of these, designated $OOP^{BTP1}$, occupies the same position as the OOP sRNA of phage Lambda, antisense to and overlapping, the 3' end of the *cII* gene (Krinke and Wulff, 1987). In phage lambda, OOP is thought to regulate the expression of the *cII* gene, thereby modulating the switch between lysogenic and lytic infection. The remaining 9 putative ncRNAs have not been identified in other lambdoid phages and prophages and therefore their function is unknown.

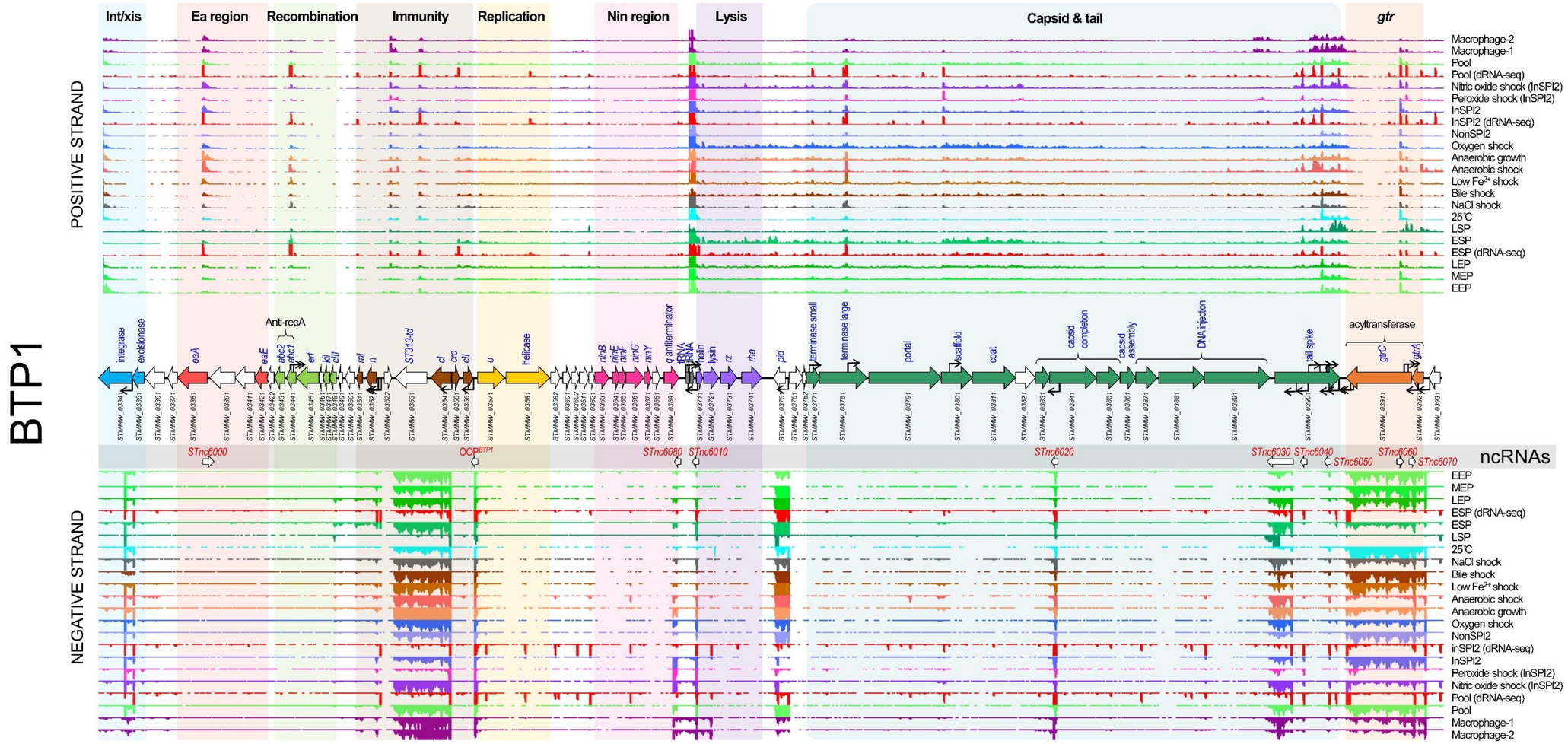



**Figure 5.1 Transcriptional map of prophage BTP1 of D23580 in 17 infection-relevant conditions.** RNA-seq and dRNA-seq data from Canals et al. (2018) and Hammarlöf et al. (2017). Each track represents a different condition, and upper panel shows sequence reads mapped to the positive strand and lower panel, negative strand. TSS were identified by dRNA-seq and are indicated by curved black arrows on the annotation track. Methods associated with this analysis can be found in Chapter 2.12.4.

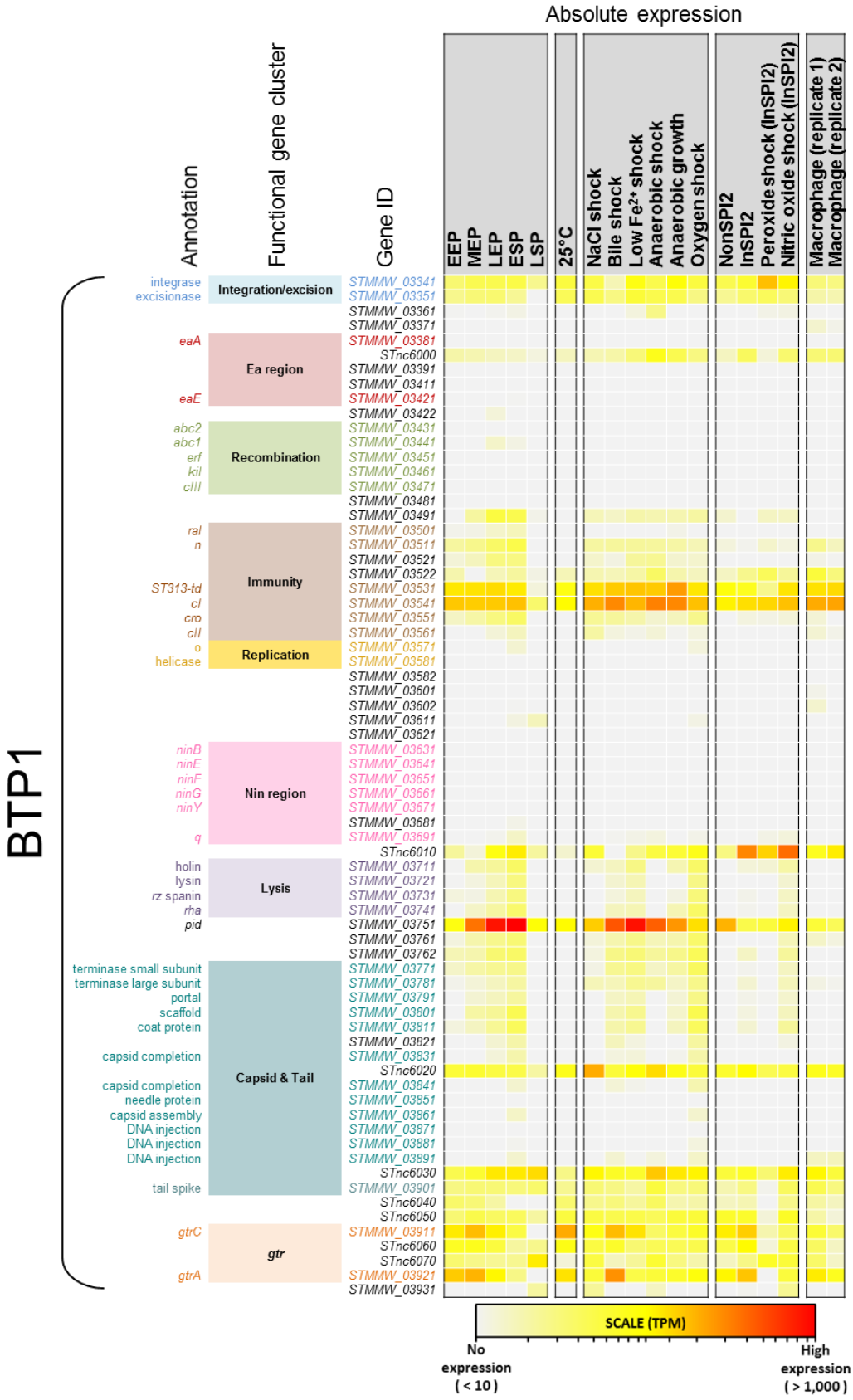



**Figure 5.2 Heat map showing the absolute expression of BTP1 genes in 17 infection-relevant conditions.** RNA-seq data from Canals et al. (2018) were used to generate transcript per million (TPM) absolute expression values for each gene and annotated ncRNA of the BTP1 prophage. Methods associated with this analysis can be found in Chapter 2.12.4.

### 5.2.2 The Gifsy-2 prophage transcriptome in D23580 & 4/74

Unlike the BTP1 prophage transcriptome, where expression of the lytic genes can be seen in most conditions tested, the late-lytic genes of Gfisy-2 were not expressed (TPM <10) (Figure 5.3, 5.4). The highly expressed genes of Gifsy-2 were restricted to known prophage accessory genes, such as the virulence-associated genes *sodCI*, *gtgE* and *sseI* (known to be a pseudogene in D23580).

To examine the conservation of gene expression patterns between prophages in different host backgrounds, the expression of Gifsy-2 genes in D23580 was compared to the expression in ST19 strain 4/74 (Kröger et al., 2013) (Figure 5.4). The most obvious difference between the expression patterns is that Gifsy-2 sRNA *IsrB-1* appeared to be expressed in strain 4/74 but not in D23580. However, subsequent investigation showed that *IsrB-1* is duplicated in the D23850 genome (it is present on both Gifsy-$2^{D23580}$ and Gifsy-$1^{D23580}$), meaning that *IsrB-1* RNA-seq reads cannot be uniquely mapped to the D23580 genome and the absence of *IsrB-1* expression in D23580 is likely to be an artefact. Aside from this discrepancy, there was remarkable consistency in the gene expression of the two Gifsy-2 prophages, showing that prophage gene expression is independent of host background, and that the RNA-seq conditions were highly reproducible between the two *Salmonella* strains.

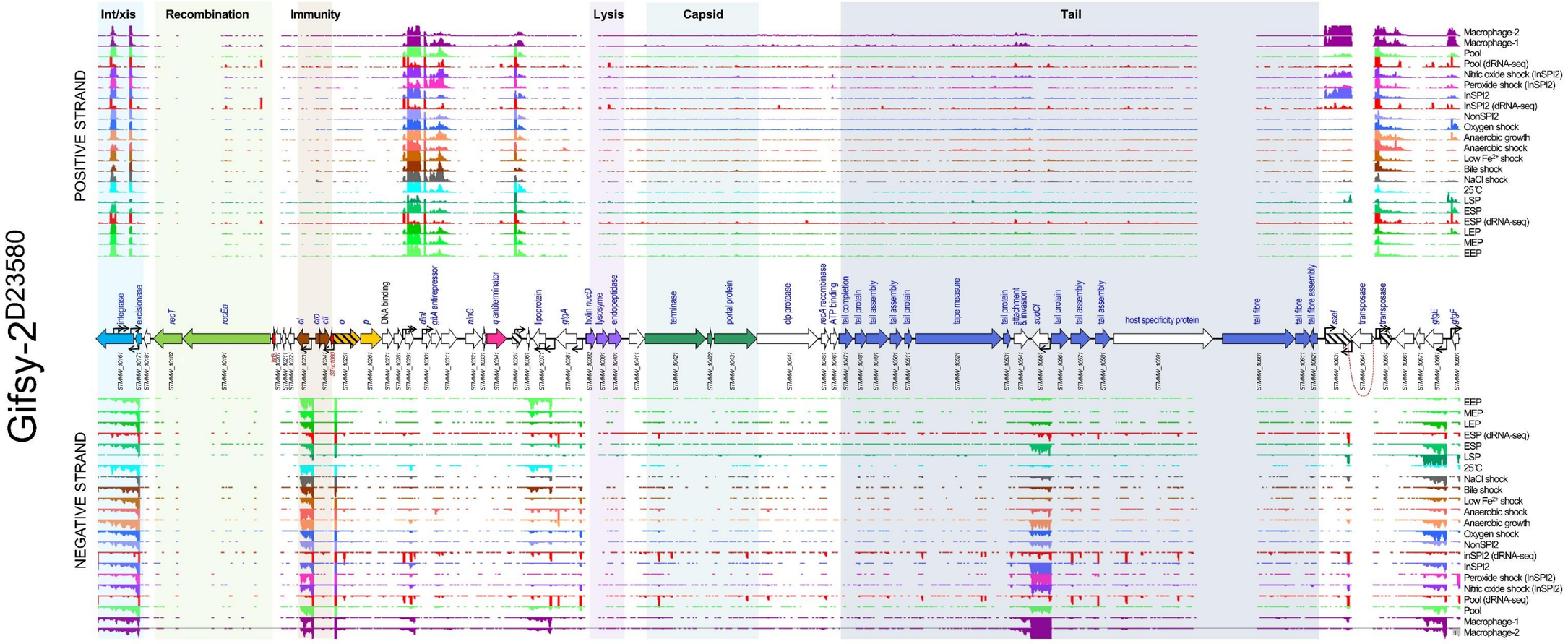



**Figure 5.3 Transcriptional map of prophage Gifsy-2 of D23580 in 17 infection-relevant conditions.** RNA-seq and dRNA-seq data from Canals et al. (2018) and Hammarlöf et al. (2017). Each track represents a different condition, and upper panel shows sequence reads mapped to the positive strand and lower panel, negative strand. TSS were identified by dRNA-seq and are indicated by curved black arrows on the annotation track. Arrows filled with slanted black stripes represent pseudogenes. Red dotted lines connect the separated parts of pseudogenised genes which have been disrupted by early truncation or insertion. Methods associated with this analysis can be found in Chapter 2.12.4.

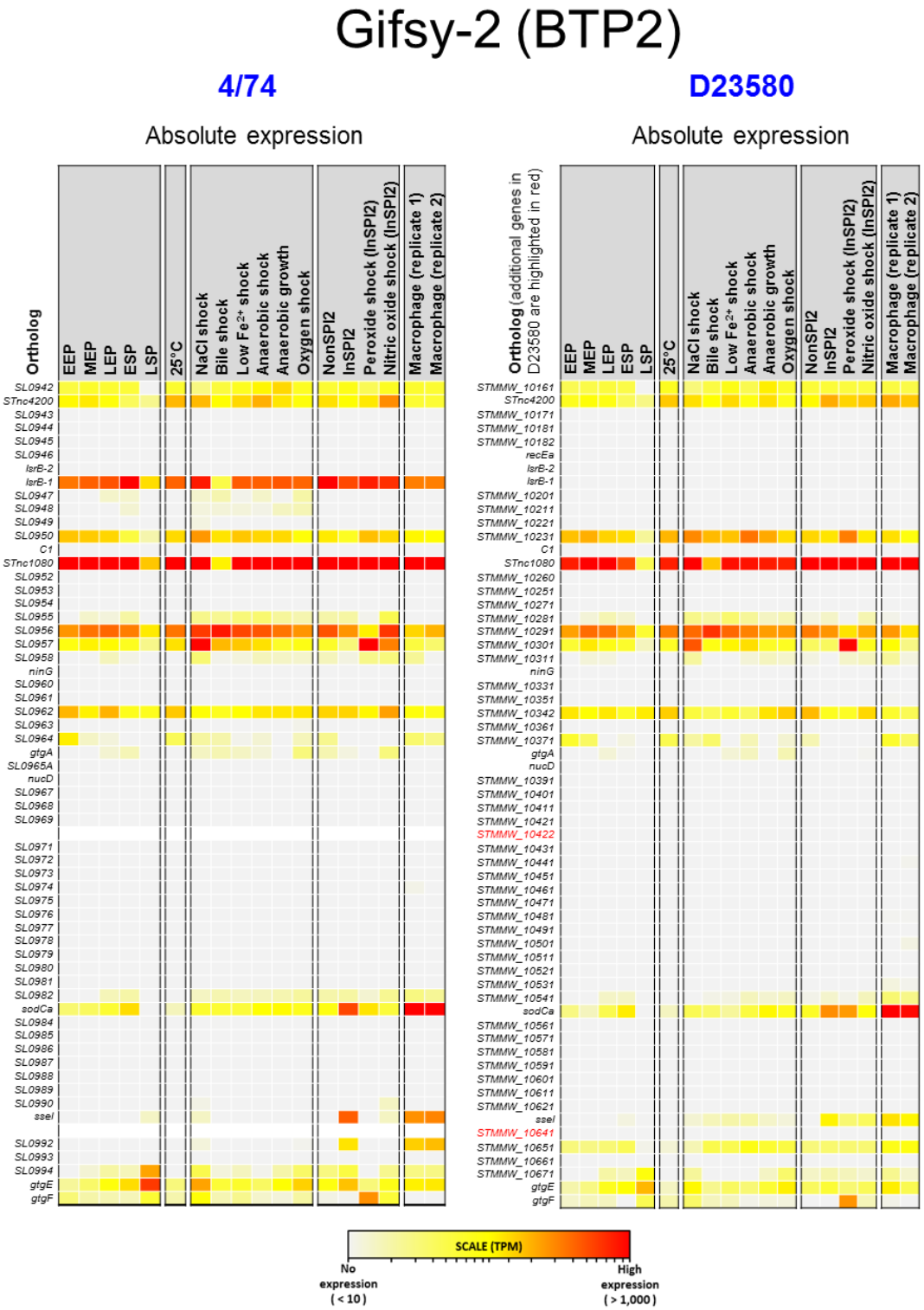



**Figure 5.4 Heat map showing the absolute expression of the Gifsy-2 prophage of D23580 and 4/74 strains in 17 infection-relevant conditions.** RNA-seq data from Canals et al. (2018) were used to generate transcript per million (TPM) absolute expression values for each gene and annotated RNA of the Gifsy-2 prophage. Ortholog IDs in red text indicate genes unique to Gifsy-$2^{D23580}$ or Gifsy-$2^{4/74}$. Methods associated with this analysis can be found in Chapter 2.12.4.

### 5.2.3 The ST64B prophage transcriptome in D23580 & 4/74

The ST64B prophage transcriptome only showed significant late lytic-gene transcription in 2 of the RNA-seq conditions; peroxide shock and nitric oxide shock (Figure 5.5 and Figure 5.6). As hydrogen peroxide and nitric oxide cause oxidative stress and putative DNA damage, transcription of the lytic genes in these conditions is likely to represent induction of the ST64B prophage. The lack of evidence for increased induction of the remaining D23580 prophages in these conditions suggests that the ST64B prophage in D23580 is specifically sensitive to peroxide and nitric oxide stress. The prophage contains a gene that encodes the secreted effector protein SseK3. The *sseK3* gene was only expressed in conditions linked to intracellular infection, such as InSPI2 (a medium designed to mimic intracellular conditions), peroxide shock, nitric oxide shock and the intra-macrophage environment. A similar expression profile has been observed for genes that encode other SPI2-translocated effector proteins in *S.* Typhimurium strain 4/74 (Srikumar et al., 2015). The patterns of gene expression of $ST64B^{D23580}$ and $ST64B^{4/74}$ showed many similarities, though in the 4/74 data the prophage did not show late lytic gene expression in the peroxide shock condition suggesting that the prophage may exhibit different induction behaviour in strain 4/74.

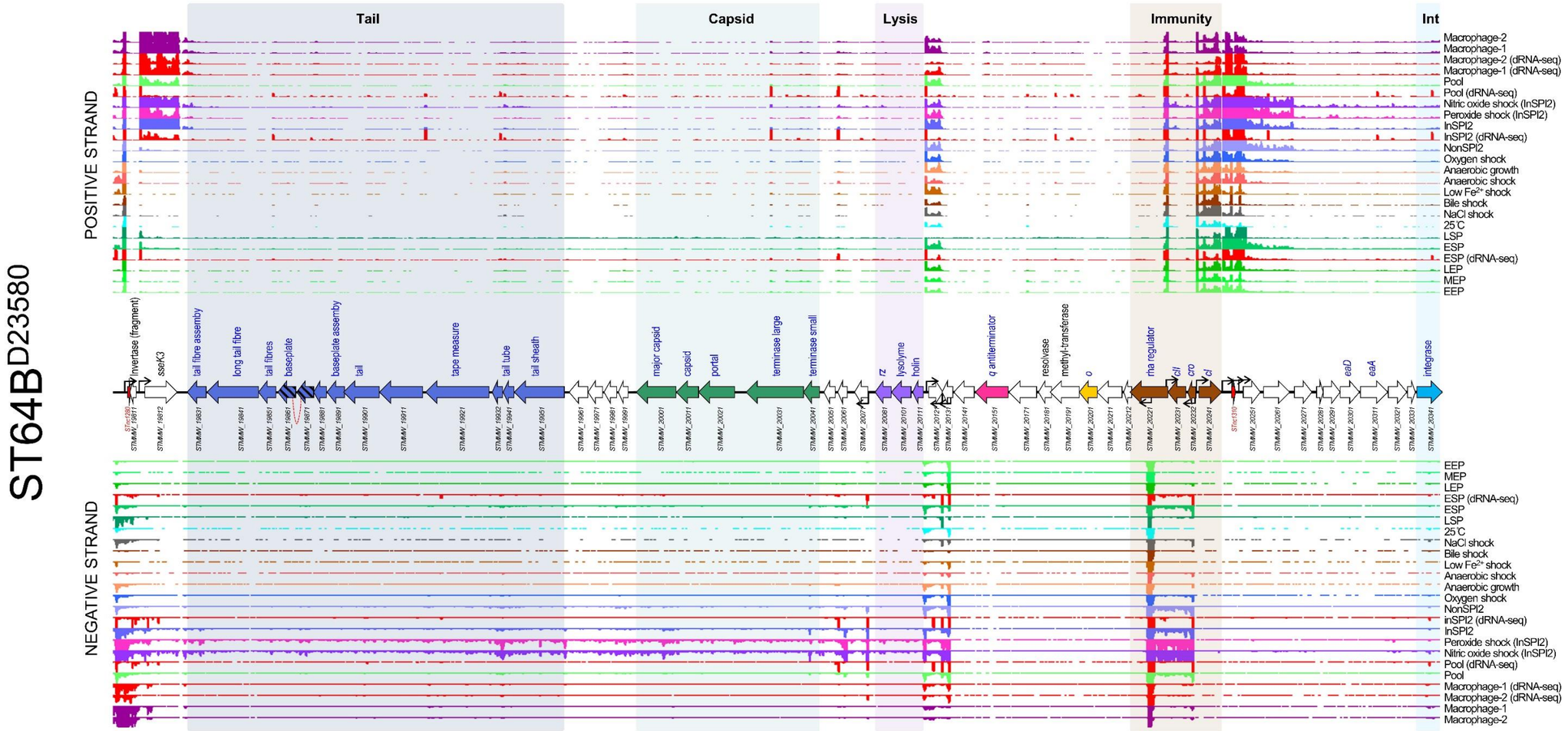



**Figure 5.5 Transcriptional map of prophage ST64B of D23580 in 17 infection-relevant conditions.** RNA-seq and dRNA-seq data from Canals et al. (2018) and Hammarlöf et al. (2017). Each track represents a different condition, and upper panel shows sequence reads mapped to the positive strand and lower panel, negative strand. TSS were identified by dRNA-seq and are indicated by curved black arrows on the annotation track. Arrows filled with slanted black stripes represent pseudogenes. Red dotted lines connect the separated parts of pseudogenised genes which have been disrupted by early truncation or insertion. Methods associated with this analysis can be found in Chapter 2.12.4.

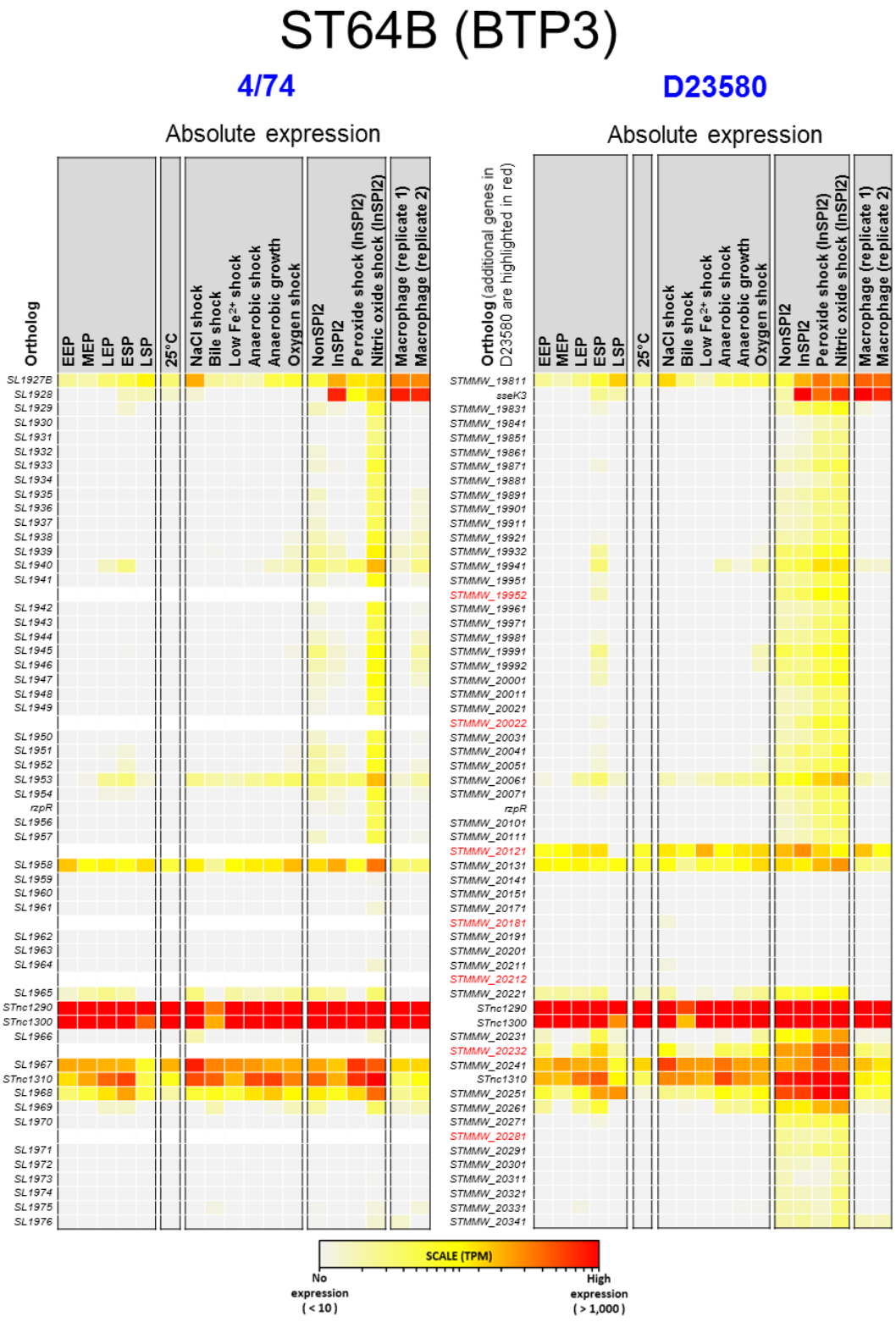



**Figure 5.6 Heat map showing the absolute expression of the ST64B prophage of D23580 and 4/74 strains in 17 infection-relevant conditions.** RNA-seq from Canals et al. (2018) was used to generate transcript per million (TPM) absolute expression values for each gene and annotated RNA of the ST64B prophage. Ortholog IDs in red text indicate genes unique to ST64B$^{D23580}$ or ST64B$^{4/74}$. White colouring of the heatpmaps indicates that there is no ortholog in the opposing strain. Methods associated with this analysis can be found in Chapter 2.12.4.

### 5.2.4 The Gifsy-1 prophage transcriptome in D23580 & 4/74

Like the Gifsy-2 prophage, Gifsy-1 showed no evidence of lytic-gene transcription in any of the 17 environmental conditions. This is consistent with the very low level of spontaneous induction that was observed for the resuscitated Gifsy-1 prophage in Chapter 3.5. However, Gifsy-1 contained a number of ncRNAs which were highly transcribed in all conditions tested (Figure 5.7). The virulence-associated genes *gogB* (*STMMW_26001*), *pagJ* (*STMMW_26011*) and *pagK* (*STMMW_26041*) were only expressed in intracellular infection-linked conditions. Contrastingly, the virulence-associated gene *gipA* (*STMMW_26191*) was transcribed in all conditions tested, despite being previously reported to be specifically induced during colonisation of the small intestine (Stanley et al., 2000). Another virulence-associated gene, *gtgA* (*STMMW_26331*) showed very little transcription in any of the conditions tested. The functionally uncharacterised gene *STMMW_26411* was specifically induced in the anaerobic shock and anaerobic growth conditions, suggesting that this gene could be important during anaerobic conditions.

The Gifsy-1 prophage is not identical between strains D23580 and 4/74, and shows considerable difference in gene content particularly at the terminal end. Therefore, comparison of the Gifsy-1 gene expression profiles between the two strains is difficult to interpret (Figure 5.8). However, the expression patterns of orthologous genes shared by the two prophages showed many similarities across the conditions.

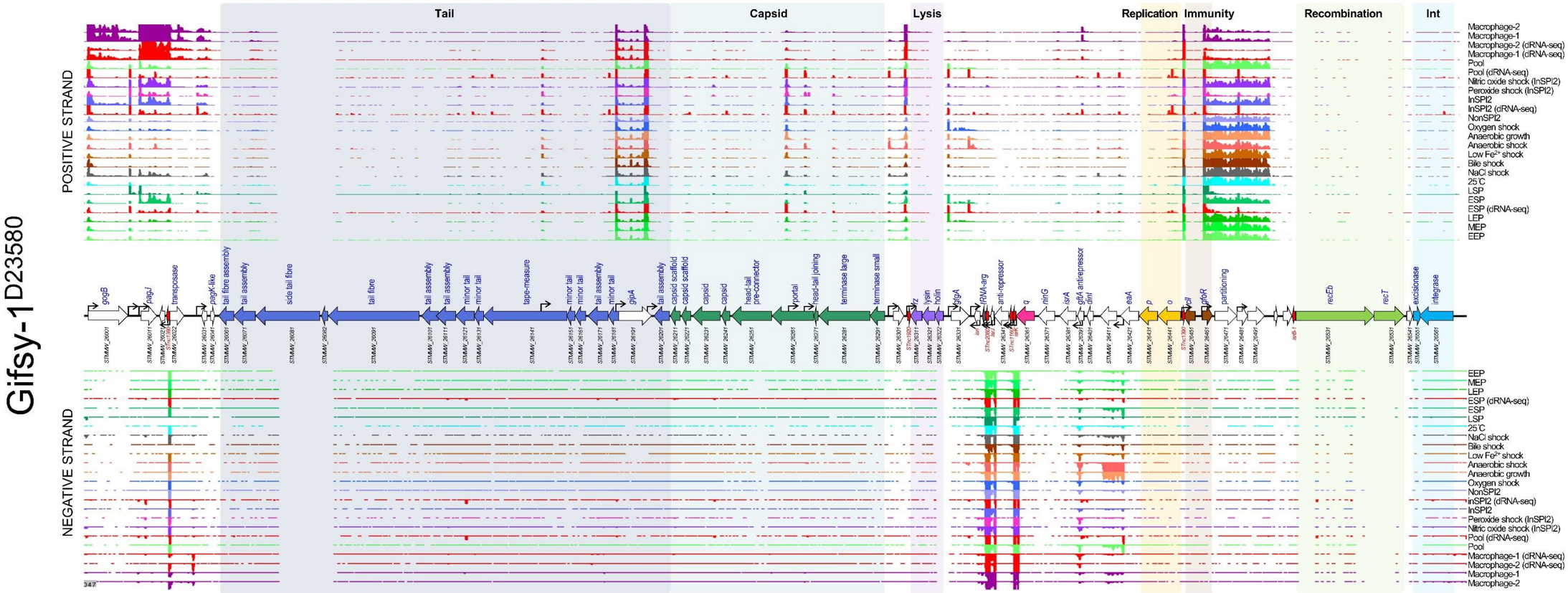



**Figure 5.7 Transcriptional map of prophage Gifsy-1 of D23580 in 17 infection-relevant conditions.** RNA-seq and dRNA-seq data from Canals et al. (2018) and Hammarlöf et al. (2017). Each track represents a different condition, and upper panel shows sequence reads mapped to the positive strand and lower panel, negative strand. TSS were identified by dRNA-seq and are indicated by curved black arrows on the annotation track. Methods associated with this analysis can be found in Chapter 2.12.4.


**Figure 5.8 Heat map showing the absolute expression of the Gifsy-1 prophage of D23580 and 4/74 strains in 17 infection-relevant conditions.** RNA-seq data from Canals et al. (2018) were used to generate transcript per million (TPM) absolute expression values for each gene and annotated RNA of the Gifsy-1 prophage. Ortholog IDs in red text indicate genes unique to Gifsy-1$^{D23580}$ or Gifsy-1$^{4/74}$. Methods associated with this analysis can be found in Chapter 2.12.4.

### 5.2.5 The BTP5 prophage transcriptome

The BTP5 prophage showed the least transcriptional activity of all the D23580 prophages (Figure 5.9, Figure 5.10). However, the expression of BTP5 genes presents a confusing insight into the biology of the prophage. The most highly expressed transcript in the prophage belonged to the *tum* gene (*STMMW_32041*), a homolog of the *tum* antirepressor of coliphage 186 (Shearwin et al., 1998). The antirepressor was expressed particularly highly in the nitric oxide shock condition, raising the possibility that nitric oxide could stimulate induction of BTP5. However, this observation does not fit the inability to identify spontaneously induced BTP5 phages in Chapter 3.6, even with chemical induction. Additionally, transcription was observed from the TSS upstream of the *apI* gene (*STMMW_32112*) through to the uncharacterised gene *STMMW_32601* in certain conditions, including early stationary phase (ESP), bile shock and anaerobic shock. The corresponding genes in coliphage 186 have been defined as the early lytic operon (Shearwin and Egan, 2000) and represent the genes initially expressed during lytic phage replication. Expression of a 3-gene operon consisting of tail structural genes (*STMMW_31821*, *STMMW_31811* and *STMMW_31801*) was also observed in a number of conditions, particularly ESP, bile shock and nitric oxide shock. However, the tail structure of P2-like phages (Myoviruses) is complex (Nilsson and Haggård-Ljungquist, 2006), and expression of these three genes alone would not produce functional phage tail particles and therefore the functional relevance of this putative transcript is unclear. Additionally the *ogr* gene (*STMMW_31741*), reported to be involved in control of late gene expression in phage P2 (Christie et al., 1986), is expressed in all conditions tested.

Despite the evidence for transcription of late lytic genes in the BTP5 prophage, the repressor and integrase genes were expressed in all conditions tested, albeit at low levels relative to the level of tail gene transcription. In summary, the BTP5 prophage transcriptome does not help to ascertain the functionality of the prophage, and as in Chapter 3, the lysogeny and lysis behaviour of the BTP1 prophage remains mysterious. There may be further control of prophage BTP5 at the post-transcriptional level, or alternatively the transcriptome may reflect heterogeneity in the activity of the BTP5 prophage across the cellular population, suggested for the BTP1 prophage transcriptome (Chapter 5.2.1).

BTP5

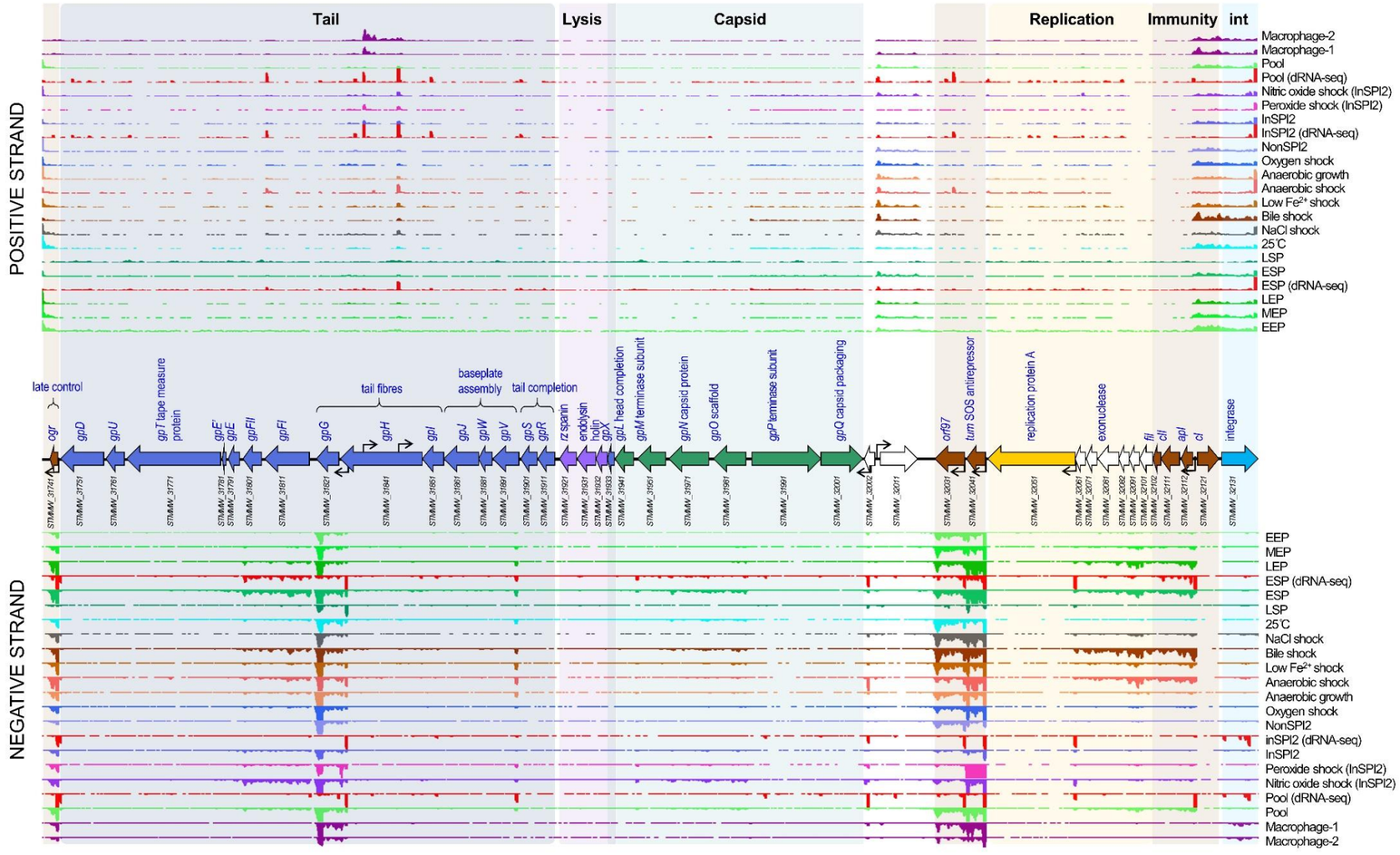



**Figure 5.9 Transcriptional map of prophage BTP5 of D23580 in 17 infection-relevant conditions.** RNA-seq and dRNA-seq data from Canals et al. (2018) and Hammarlöf et al. (2017). Each track represents a different condition, and upper panel shows sequence reads mapped to the positive strand and lower panel, negative strand. TSS were identified by dRNA-seq and are indicated by curved black arrows on the annotation track. Methods associated with this analysis can be found in Chapter 2.12.4.

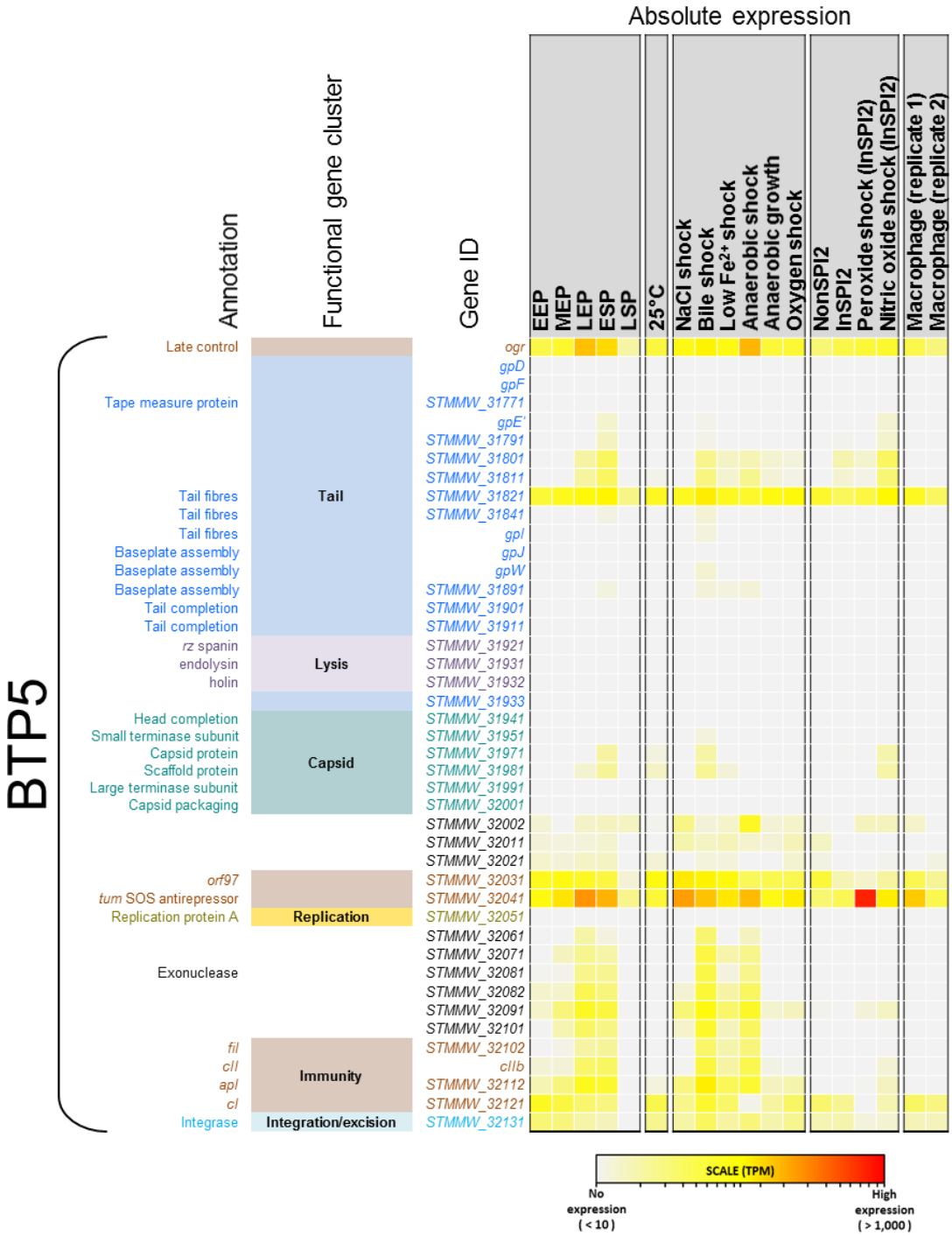



**Figure 5.10 Heat map showing the absolute expression of BTP5 genes in 17 infection-relevant conditions.** RNA-seq data from Canals et al. (2018) were used to generate transcript per million (TPM) absolute expression values for each gene and annotated ncRNA of the BTP5 prophage. Methods associated with this analysis can be found in Chapter 2.12.4.

## 5.3 Investigating the regulatory contribution of BTP1 & BTP5 at the transcriptional level

The function of temperate phages in bacterial pathogens is exemplified by the presence of phage-encoded toxins, for example the Shiga toxin of Stx phages in Shiga-toxinigenic *E. coli* (STEC), or the cholera toxin carried by the CTXΦ phage of *Vibrio cholera* (Brüssow et al., 2004). The acquisition of a temperate phage carrying a toxin or virulence gene by a bacterial pathogen may facilitate rapid adaption to a new niche, for example by altering the disease tropism of the pathogen, known as 'lysogenic conversion'. However, the mechanism of lysogenic conversion of a bacterial pathogen by a prophage may not always be as overt as the addition of a toxin gene. For example, whilst Stx phages encode the Stx enterotoxin, the *cII* regulatory gene of Stx phage $\Phi 24_B$ has also been shown to up-regulate the glutamic acid decarboxylase operon, conferring acid resistance to the lysogen which may increase bacterial survival in the intestinal tract (Veses-Garcia et al., 2015). Additionally, the A118 prophage of *Listeria monocytogenes* dynamically integrates and excises from its attachment site within the *comK* gene, encoding the master activator of the competence system. This mechanism serves as a genetic switch that modulates the intracellular lifestyle of the pathogen (Rabinovich et al., 2012). Evidently, a prophage could plausibly transcriptionally rewire a host bacterium for its own benefit (Feiner et al., 2015), perhaps akin to examples of protozoan parasites such as *Toxoplasma gondii,* for which infection of an intermediate rodent host appears to enhance transmission to the feline definitive host (Adamo and Webster, 2013). Could the disease tropism differences between African *S.* Typhimurium ST313 and more typical *S.* Typhimurium sequence types such as ST19, be due to transcriptional changes mediated by novel prophages BTP1 and BTP5?

### 5.3.1 Identification of D23580 genes differentially expressed by the presence of prophages by RNA-seq

To investigate the transcriptional contribution of prophages BTP1 and BTP5, RNA-seq experiments were conducted on isogenic mutants lacking either one, or both of the prophages (the strains used in these experiments are represented diagrammatically in Figure 5.11, and strain construction is detailed in chapter 2.8.2). Gene expression was investigated in 3 different environmental conditions (Kröger et al., 2013); early stationary phase growth in L-broth (ESP), anaerobic growth in L-broth ($noO_2$), and a *Salmonella* Pathogenicity Island 2-inducing condition in PCN media (inSPI2) (see chapter 2.9.1 for full description of growth conditions). All experiments were conducted in triplicate, and an overview of the experiment is shown in Table 5.1.

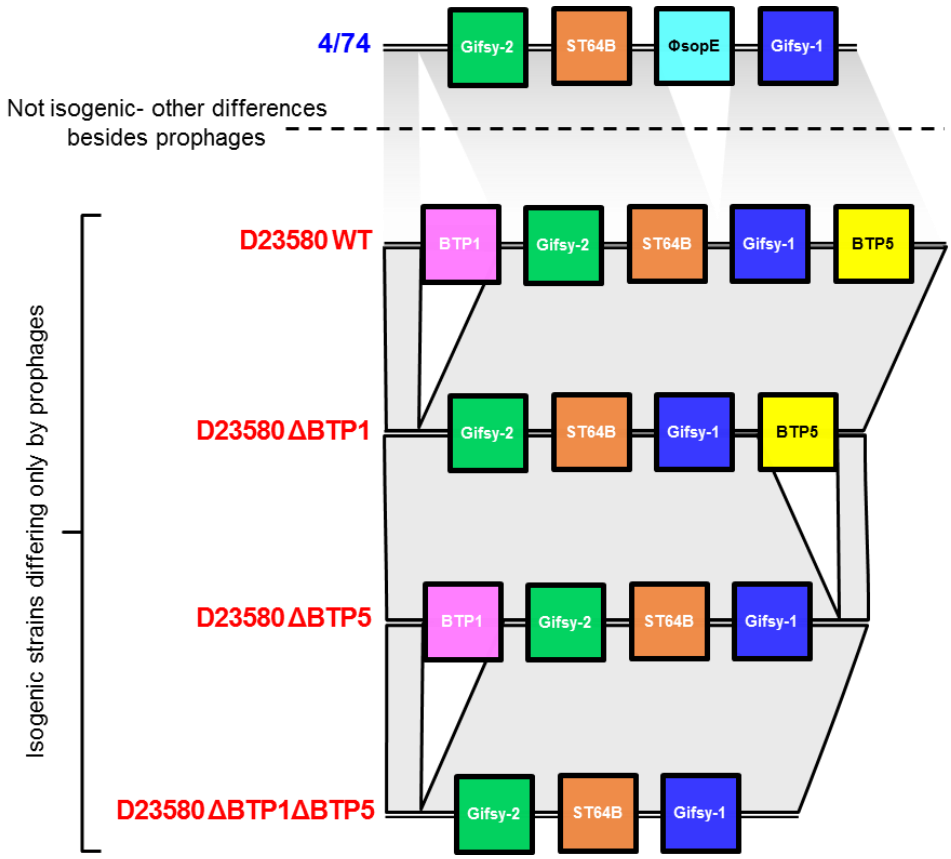



**Figure 5.11 Schematic representation of the strains used in Chapter 5.6.** Grey regions indicate shared sequence identity. White regions indicate missing/deleted regions.

**Table 5.1 Experimental design of the RNA-seq experiments described in Chapter 5.6.** Final column indicates who the RNA samples were collected by.

| Sample | Condition | Biological Replicates | RNA collected by± |
|---|---|---|---|
| D23580 wt | ESP | 3 | RC |
| D23580 wt | InSPI2 | 3 | RC |
| D23580 wt | $NoO_2$ | 3 | RC |
| D23580 ΔBTP1 | ESP | 3 | SO |
| D23580 ΔBTP1 | InSPI2 | 3 | SO |
| D23580 ΔBTP1 | $NoO_2$ | 3 | SO |
| D23580 ΔBTP5 | ESP | 3 | SO |
| D23580 ΔBTP5 | InSPI2 | 3 | SO |
| D23580 ΔBTP5 | $NoO_2$ | 3 | SO |
| D23580 ΔBTP1 ΔBTP5 | ESP | 3 | SO |
| D23580 ΔBTP1 ΔBTP5 | InSPI2 | 3 | SO |
| D23580 ΔBTP1 ΔBTP5 | $NoO_2$ | 3 | SO |

±RC= Rocío Canals SO= Siân Owen

All experiments were conducted by the Author, except the comparator D23580 WT samples, which were generated by Dr Rocio Canals for logistical reasons.

How the differential gene expression analysis of the RNA-seq data was conducted is described in chapter 2.12.4. Using D23580 WT as a comparator, up to 1100 gene genes were differentially expressed by more than 2-fold (FDR ≤0.01) in the prophage mutants in the ESP and anaerobic growth conditions (Figure 5.12). In the SPI-2 inducing condition (inSPI2), very few genes were differentially expressed more than 2-fold (only 3 were up-regulated in D23580 ΔBTP5 and D23580 ΔBTP1ΔBTP5 vs. D23580 WT).

Recent experiments in the Hinton Lab have identified a high level of variability between RNA-seq replicates collected by different lab workers (R. Canals, Pers. Comm.), leading to the conclusion that the very high numbers of differentially expressed genes seen in the ESP and anaerobic growth conditions are likely to represent experimental noise caused by comparing the D23580 prophage mutants (RNA-seq samples collected by the Author), with D23580 WT (RNA-seq samples collected by Rocio Canals).

To reduce any experimental noise that may result from such experimental variation, a second comparison was analysed that only involved RNA-seq samples generated by the Author. It was reasoned that, assuming the effects of the BTP1 and BTP5 prophages are independent from each other, the impact of the BTP1 prophage upon global gene expression should be evident by comparing D23580 WT to D23580 ΔBTP1 (as described above), or by comparing D23580 ΔBTP5 to D23580 ΔBTP1ΔBTP5. This is because, in both cases, the difference between the strains is the presence of BTP1 (illustrated in Figure 5.13, comparison A vs. B). Equally, the effect of the BTP5 prophage should be revealed by comparing D23580 WT to D23580 ΔBTP5, or by comparing D23580 ΔBTP1 to D23580 ΔBTP1ΔBTP5 (Figure 5.13, comparison C vs. D). Using this single vs. double prophage mutants comparator strategy, the number of genes differentially expressed >2-fold (FDR ≤0.01) in the presence of prophage BTP1 and BTP5 was significantly reduced (Figure 5.14). The comparison showed that the greatest impact upon global gene expression was mediated by the presence of the BTP5 prophage in the ESP condition, with a total of 91 genes differentially expressed >2-fold (FDR ≤0.01). In general, more genes were up-regulated in the presence of the prophages than down-regulated, suggesting the prophages are more likely to activate gene expression than to repress gene expression.

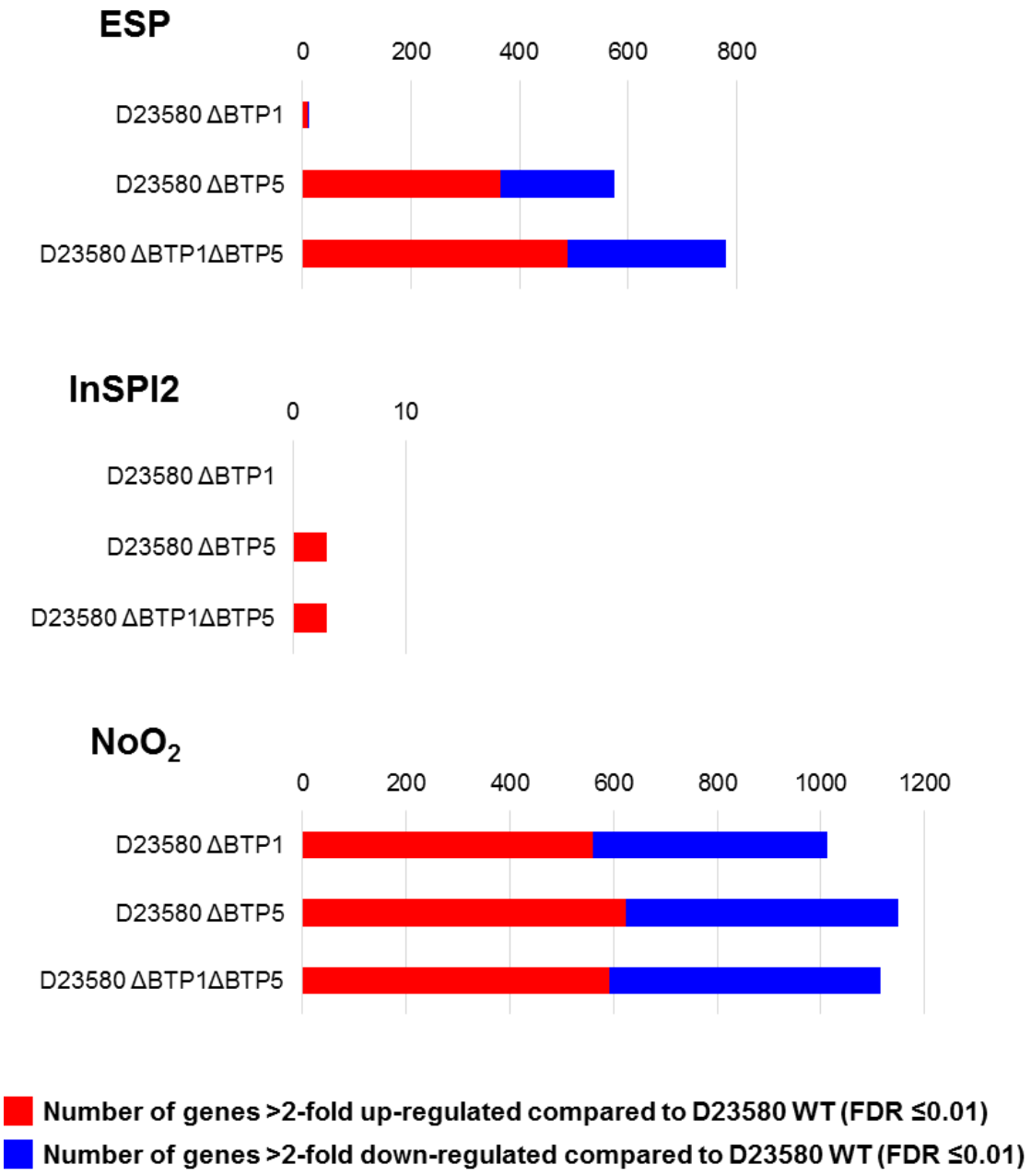



**Figure 5.12 Number of genes differentially expressed more than 2-fold in the D23580 prophage mutants compared to D23580 WT in three growth conditions, ESP, InSPI2 and NoO2.** Details of growth conditions are given in Table 2.8.

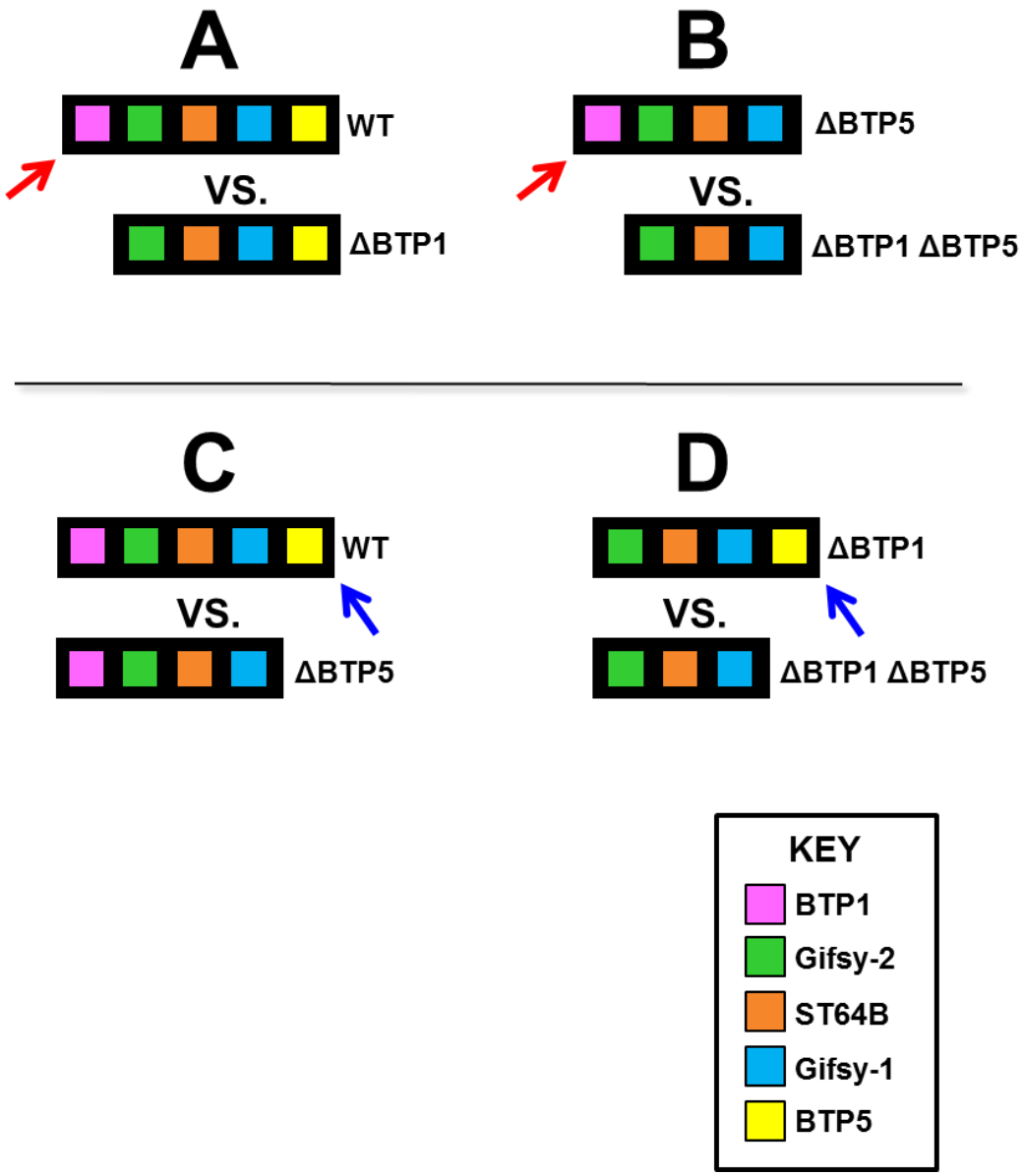



**Figure 5.13 Figurative representation of the transcriptomic comparisons used to identify the differential gene expression mediated by the BTP1 and BTP5 prophages.** Prophages are represented by coloured blocks (see key). Comparisons A and B reveal the transcriptomic impact of the BTP1 prophage (indicated by red arrow) and comparisons C and D reveal the effects of the BTP5 prophage (indicated with blue arrow).

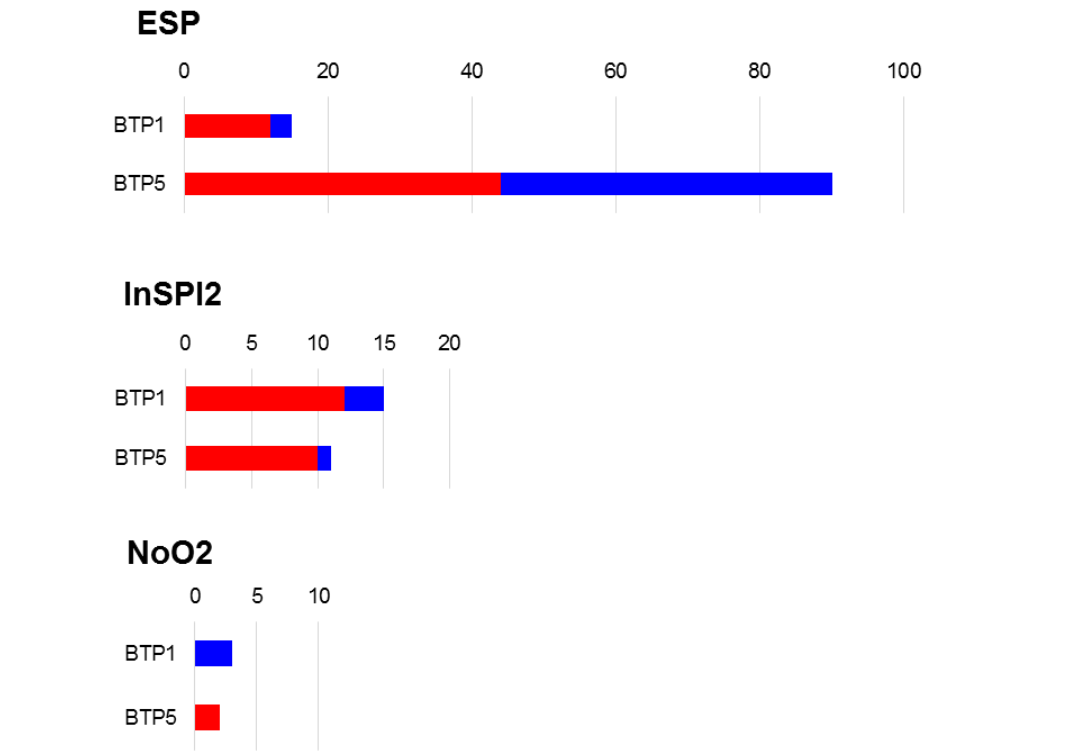



**Figure 5.14 Number of genes differentially expressed more than 2-fold in the presence of the BTP1 or BTP5 prophage, using the D23580 ΔBTP1 ΔBTP5 mutant as a comparator in the three growth conditions, ESP, InSPI2 and NoO2.** Growth conditions are described in Table 2.8.

### 5.3.2 Impact of the BTP1 prophage upon global gene expression of D23580

To further minimise experimental noise and identify only the most significant transcriptomic changes, the data were further scrutinised. The comparisons which should show the effects of the BTP1 prophage were determined to be:

- D23580 WT vs D23580 ΔBTP1
- D23580 WT vs D23580 ΔBTP1ΔBTP5
- D23580 ΔBTP5 vs D23580 ΔBTP1ΔBTP5
- D23580 ΔBTP1 vs D23580 ΔBTP5

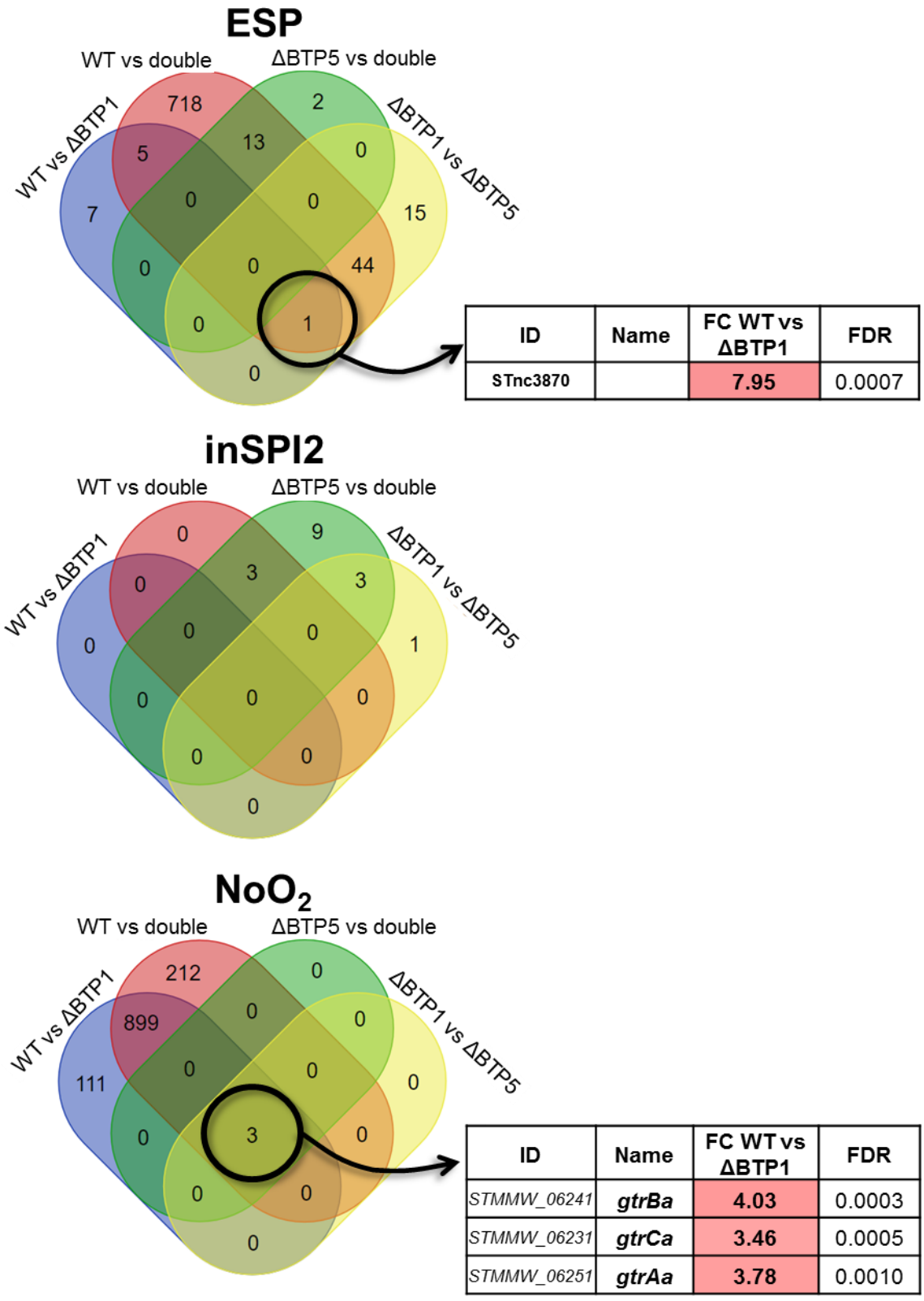
Genes differentially expressed >2-fold (either up-regulated or down-regulated) (FDR ≤0.01), in all four comparisons were cross-referenced to identify common signatures (Figure 5.15). Genes identified in at least 3 of the 4 comparisons were judged to be most significant.

In the ESP condition, the uncharacterised sRNA STnc3870 was putatively repressed by the BTP1 prophage, as the sRNA was up-regulated in D23580 ΔBTP1 nearly 8-fold compared to D23580 WT (Figure 5.16). No significant changes were identified in the inSPI2 condition. In the anaerobic growth condition, 3 genes showed BTP1-associated differential expression; *gtrBa* (*STMMW_06241*), *gtrCa* (*STMMW_06231*) and *gtrAa* (*STMMW_06251*). The *gtrABCa* operon encodes a glycosyltransferase enzyme system that mediates the addition of a glucose moiety to the galactose subunit of the O-antigen. The operon is encoded on the Def-1 remnant prophage element (Chapter 3.7). The *gtrABCa* was putatively repressed by the BTP1 prophage in the anaerobic growth condition, as the genes were up-regulated approximately 4-fold in D23580 ΔBTP1 vs. D23580 WT.

Figure 5.16 shows the expression values for STnc3870 and *gtrABCa* for all strains in ESP and anaerobic growth respectively, as well as a normalised read-coverage plot for the *gtrABCa* operon. In both cases, though the difference between D23580 WT and D23580 ΔBTP1 was significant, a much greater differential expression effect was seen in D23580 ΔBTP5. The STnc3870 sRNA appeared to be moderately down-regulated by the presence of the BTP1 prophage, and highly down-regulated by the presence of the BTP5 prophage, leading to high repression in the double prophage mutant as anticipated. On the other hand, the *gtrABCa* operon was moderately down-regulated in D23580 ΔBTP1, highly down-regulated in D23580 ΔBTP5, but only moderately down-regulated in the double prophage mutant. There are two possible explanations for this result. It is possible that the BTP1 prophage mildly represses the *gtrABCa* operon, whilst the BTP5 prophage highly represses the operon, but the mild

repression mediated by BTP1 is dominant when both prophages are present. Alternatively, the strains can be considered in terms of which prophages they possess, rather than which have been deleted. If D23580 ΔBTP1ΔBTP5 can be considered to be analogous to a version of D23580 before it acquired the two prophages, then D23580 ΔBTP5 is analogous to the ancestor after acquisition of BTP1. In this scenario, the BTP1 prophage highly activates expression of *gtrABCa* by 26-fold. Likewise, D23580 ΔBTP1 could represent the phage-less ancestor after acquisition of BTP5, which results in no change in *gtrABCa* expression. However, following acquisition of both prophages (D23580 WT), there is an overall repression of *gtrABC* by 3.7-fold.

Experimental validation of these data by methods such as quantitative PCR, or analysis of the LPS to study the GtrABCa-modification, are necessary to verify these results.




| ID | Name | FC WT vs ΔBTP1 | FDR |
|---|---|---|---|
| STnc3870 | | 7.95 | 0.0007 |

| ID | Name | FC WT vs ΔBTP1 | FDR |
|---|---|---|---|
| *STMMW_06241* | ***gtrBa*** | **4.03** | 0.0003 |
| *STMMW_06231* | ***gtrCa*** | **3.46** | 0.0005 |
| *STMMW_06251* | ***gtrAa*** | **3.78** | 0.0010 |

**Figure 5.15 Four-way Venn diagrams showing the shared BTP1-associated differential gene expression in all three conditions tested.** Numbers represent all differentially expressed genes (both up and down-regulated). 'double' refers to the D23580 ΔBTP1 ΔBTP5 mutant.

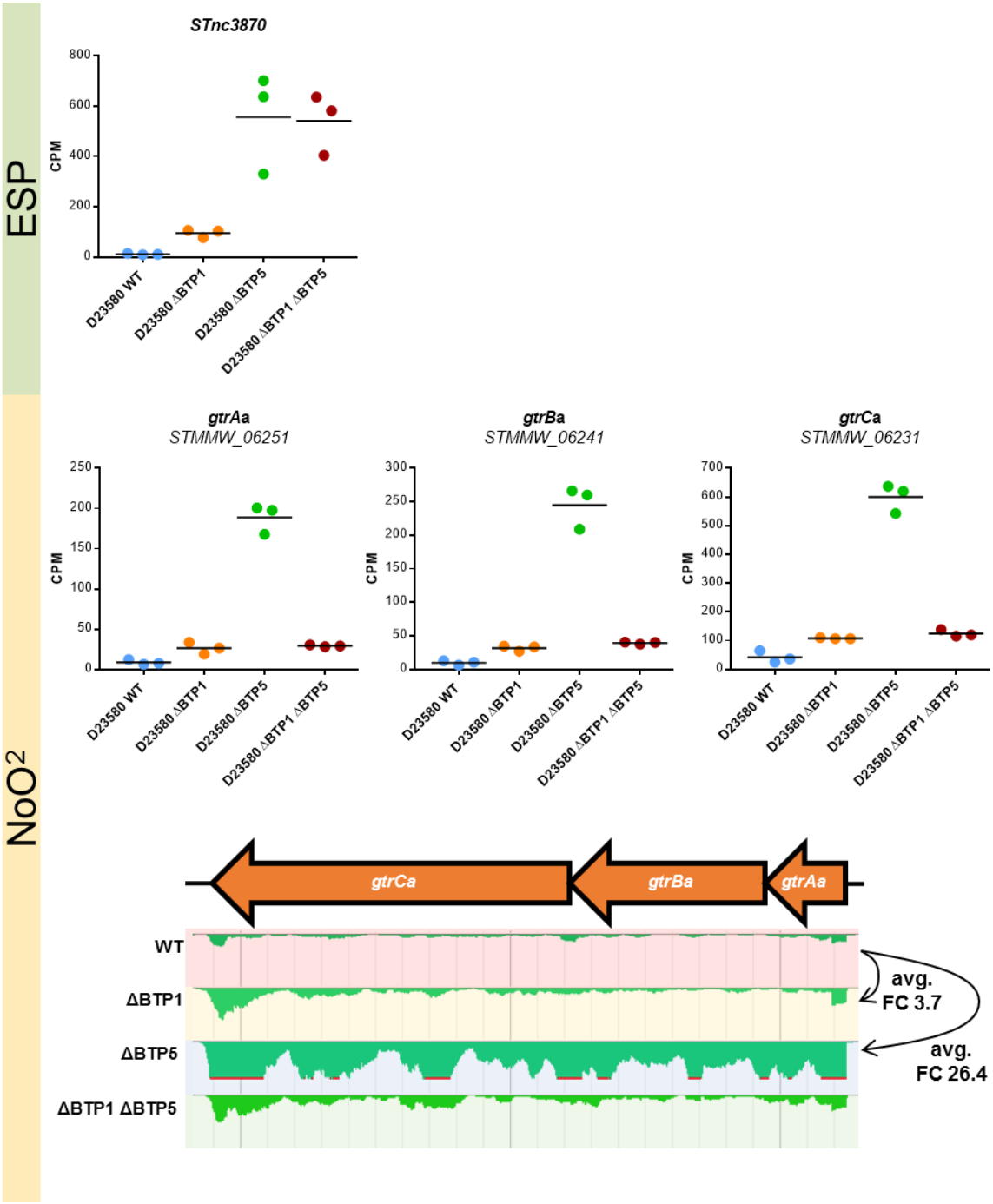



**Figure 5.16 The STnc3870 sRNA and gtrABCa operon show prophage-associated expression level changes.** Counts per million (CPM) and normalised RNA-seq coverage plots for the STnc3870 sRNA in ESP and the *gtrABCa* genes in anaerobic growth. Data are presented for all mutants. For the *gtrABCa* gene operon, transcriptomic data is also represented as mapped read coverage plot. The average fold-change in expression level (CPM) between the mutants is indicated.

### 5.3.3 Impact of the BTP5 prophage upon global gene expression of D23580

The same approach described in chapter 5.3.2 was used to identify significant transcriptional changes mediated by the BTP5 prophage. The comparisons which should show the effects of the BTP5 phage were determined to be:

- D23580 WT vs D23580 ΔBTP5
- D23580 WT vs D23580 ΔBTP1ΔBTP5
- D23580 ΔBTP1 vs D23580 ΔBTP1ΔBTP5
- D23580 ΔBTP1 vs D23580 ΔBTP5

A total of 30 genes were found to be >2-fold (FDR ≤0.01) differentially expressed in the ESP condition in all four comparisons associated with the presence of the BTP5 prophage (Figure 5.17). No genes showed significant BTP5-associated differential expression in the inSPI2 condition, and only the *gtrABCa* genes showed differential expression in the anaerobic growth condition (discussed in 5.3.2). The 30 genes that showed BTP5-associated differential expression in the ESP condition are listed in Table 5.2. 11 out of 30 genes were up-regulated in D23580 ΔBTP5 vs. D23580 WT, and 19 out of 30 genes were down-regulated, suggesting that BTP5 is responsible for both repression and activation of transcription. The expression values for the genes putatively repressed by BTP5 (up-regulated in D23580 ΔBTP5 vs. D23580 WT) are shown in Figure 5.18. To visualise the variability associated with this experiment, expression values for each strain from three biological replicates are shown. Amongst the genes putatively down-regulated by the presence of BTP5 are the co-transcribed genes *ripA* and *lgl,* which encode enzymes associated with itaconate metabolism and methylglyoxal detoxification respectively (Chakraborty et al., 2014; Sasikaran et al., 2014), and are down-regulated between 19 and 42-fold in D23580 WT compared to D23580 ΔBTP5. A normalised read coverage plot for the operon that encodes the *ripABC* and *lgl* genes as well as the downstream STnc3870 sRNA is shown in Figure 5.19. Additionally, the type III secreted effector gene, *steB*, is putatively down-regulated 3-fold in the presence of BTP5.

Among the genes putatively activated by BTP5 (those down-regulated in D23580 ΔBTP5 vs D23580 WT) in ESP are *katE* and *putA* (Figure 5.20). Interestingly, the *putA* gene and the protein encoded by the *katE* gene have been found to be differentially expressed at both the transcriptomic and proteomic levels in ESP, comparing strains D23850 vs. 4/74 (R. Canals, Pers. comm), consistent with the activation of *katE* and *putA* by BTP5 (the BTP5 prophage is not present in strain 4/74). The *putA* gene encodes a membrane-bound dehydrogenase enzyme that

oxidises proline to glutamate when proline is transported into the cell via the PutP permease (Surber and Maloy, 1998). This result suggests that D23580 WT may have an enhanced ability to metabolise proline, a phenotype which may be able to be verified by growth experiments using proline as a carbon source.

The *katE* gene encodes a stationary-phase catalase enzyme which is required for oxidative stress-resistance during multi-cellular colonial growth (Ma and Eaton, 1992). However, the *katE* gene of *S.* Typhimurium ST313 strain D23580 encodes a non-functional catalase protein due to a non-synonymous SNP mutation (Singletary et al., 2016). Consequently, the up-regulation of *katE* by prophage BTP5 is unlikely to confer a biological advantage. Detection of the *katE* mRNA or the KatE protein by quantitative PCR or western blot respectively would allow the confirmation of these findings.

Furthermore, many of the genes in the ATP synthase operon were found to be up-regulated by the BTP5 prophage including *atpA, atpC, atpD, atpG* and *atpH,* as well as the Cytochrome O ubiquinol oxidase operon *cyoAB*.

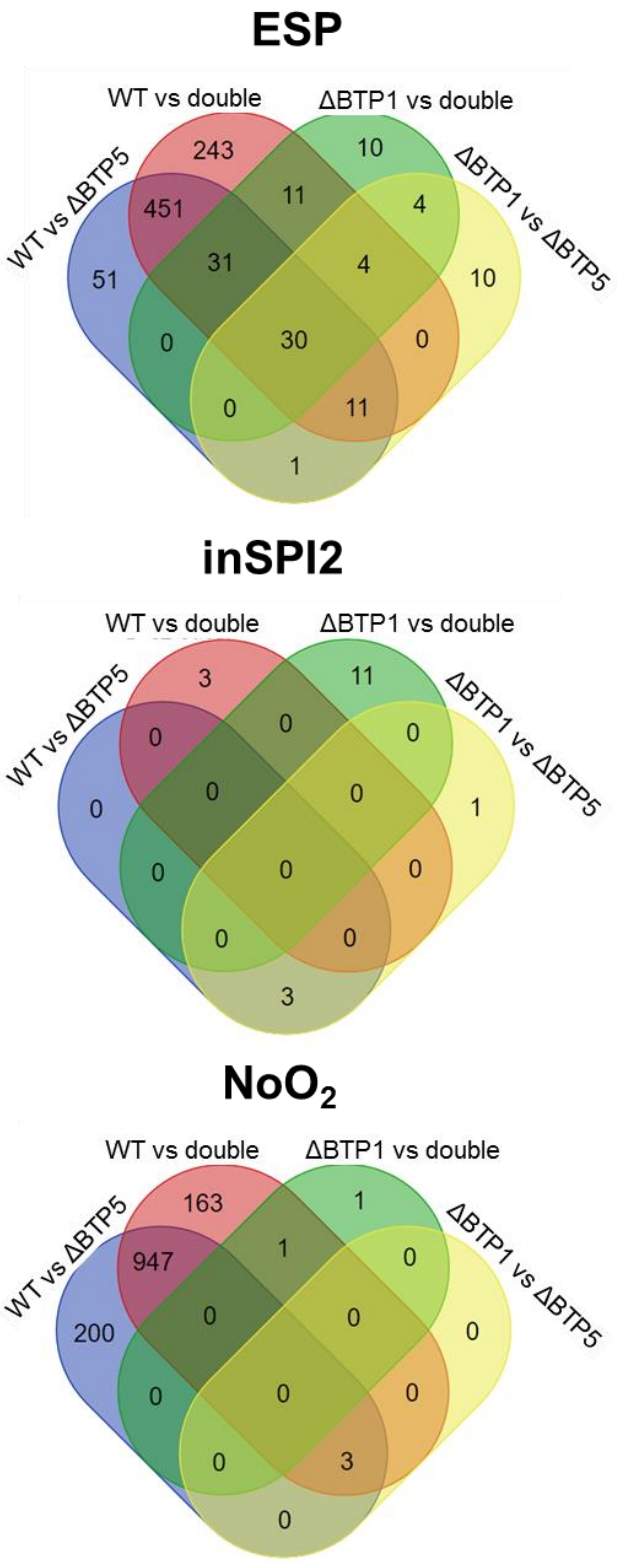



**Figure 5.17 Four-way Venn diagrams showing the shared BTP5-associated differential gene expression in all three conditions tested.** Numbers represent all differentially expressed genes (both up and down-regulated). 'double' refers to the D23580 ΔBTP1 ΔBTP5 mutant.

**Table 5.2 Genes that showed BTP5-associated differential expression.** Genes differentially expressed more than 2-fold in D23580 ΔBTP5 vs D23580 WT and also differentially expressed in the three other comparisons (Figure 5.17).

| ID | Name | FC WT vs ΔBTP5 | FDR | Description |
|---|---|---|---|---|
| *STnc3870* | | 42.66 | 2.17E-08 | Non-coding RNA |
| *STMMW_30781* | ***ripA*** | 42.58 | 8.81E-12 | Putative hydrolase |
| *STMMW_30771* | ***lgl*** | 19.02 | 5.51E-10 | Methylglyoxal detoxification protein |
| *STMMW_23811* | | 7.08 | 4.82E-08 | Putative amino acid transporter |
| *STMMW_44111* | | 5.94 | 3.01E-05 | Carbamate kinase |
| *STMMW_44271* | ***idnR*** | 4.52 | 1.68E-07 | Positive regulator of L-idonate catabolism |
| *STMMW_29071* | ***ygcB*** | 3.39 | 8.64E-09 | CRISPR-associated helicase Cas3, protein |
| *STMMW_16241* | ***steB*** | 3.20 | 4.03E-09 | Type III secretion system effector protein |
| *STMMW_18251* | | 2.82 | 0.000107 | Putative membrane protein |
| *STMMW_01851* | ***yadE*** | 2.37 | 2.5E-06 | Putative polysaccharide deacetylase |
| *STMMW_23801* | | 2.11 | 6.15E-06 | Putative cytoplasmic protein |
| *STMMW_12481* | ***icdA*** | -2.05 | 4.36E-07 | Isocitrate dehydrogenase |
| *STMMW_11331* | ***putA*** | -2.10 | 0.003465 | Proline dehydrogenase (proline oxidase) |
| *STMMW_31171* | ***yghA*** | -2.11 | 2.61E-07 | Putative oxidoreductase |
| *STMMW_44231* | ***pepA*** | -2.16 | 2.28E-07 | Cytosol aminopeptidase |
| *STMMW_34371* | ***tufA*** | -2.25 | 9.88E-08 | Elongation factor Tu |
| *STMMW_15051* | ***rspA*** | -2.33 | 0.000457 | Starvation sensing protein |
| *STMMW_13251* | ***katE*** | -2.43 | 1.55E-07 | Catalase |
| *STMMW_21991* | ***dld*** | -2.54 | 3.40E-07 | D-lactate dehydrogenase |
| *STMMW_05131* | ***cyoB*** | -2.56 | 2.82E-08 | Cytochrome O ubiquinol oxidase subunit I |
| *STMMW_40991* | ***tufB*** | -2.57 | 1.53E-08 | Elongation factor Tu |
| *STMMW_09711* | ***lrp*** | -2.65 | 1.44E-08 | Leucine-responsive regulatory protein |
| *STMMW_38501* | ***atpD*** | -2.71 | 4.45E-08 | ATP synthase beta subunit |
| *STMMW_05141* | ***cyoA*** | -2.91 | 6.82E-09 | Cytochrome O ubiquinol oxidase subunit II |
| *STMMW_38491* | ***atpC*** | -3.09 | 2.92E-08 | ATP synthase epsilon subunit |
| *STMMW_38511* | ***atpG*** | -3.11 | 9.95E-09 | ATP synthase gamma subunit |
| *STMMW_02231* | ***tsf*** | -3.21 | 4.41E-09 | Translation elongation factor Ts |
| *STMMW_38521* | ***atpA*** | -3.23 | 5.74E-09 | ATP synthase alpha subunit |
| *STMMW_38531* | ***atpH*** | -3.31 | 1.41E-08 | ATP synthase delta subunit |
| *STMMW_10101* | ***ompF*** | -4.62 | 9.84E-10 | Outer membrane protein F precursor |

*FC= Fold change. Genes listed in this table were identified by the four-way Venn diagram shown in Figure 5.17*

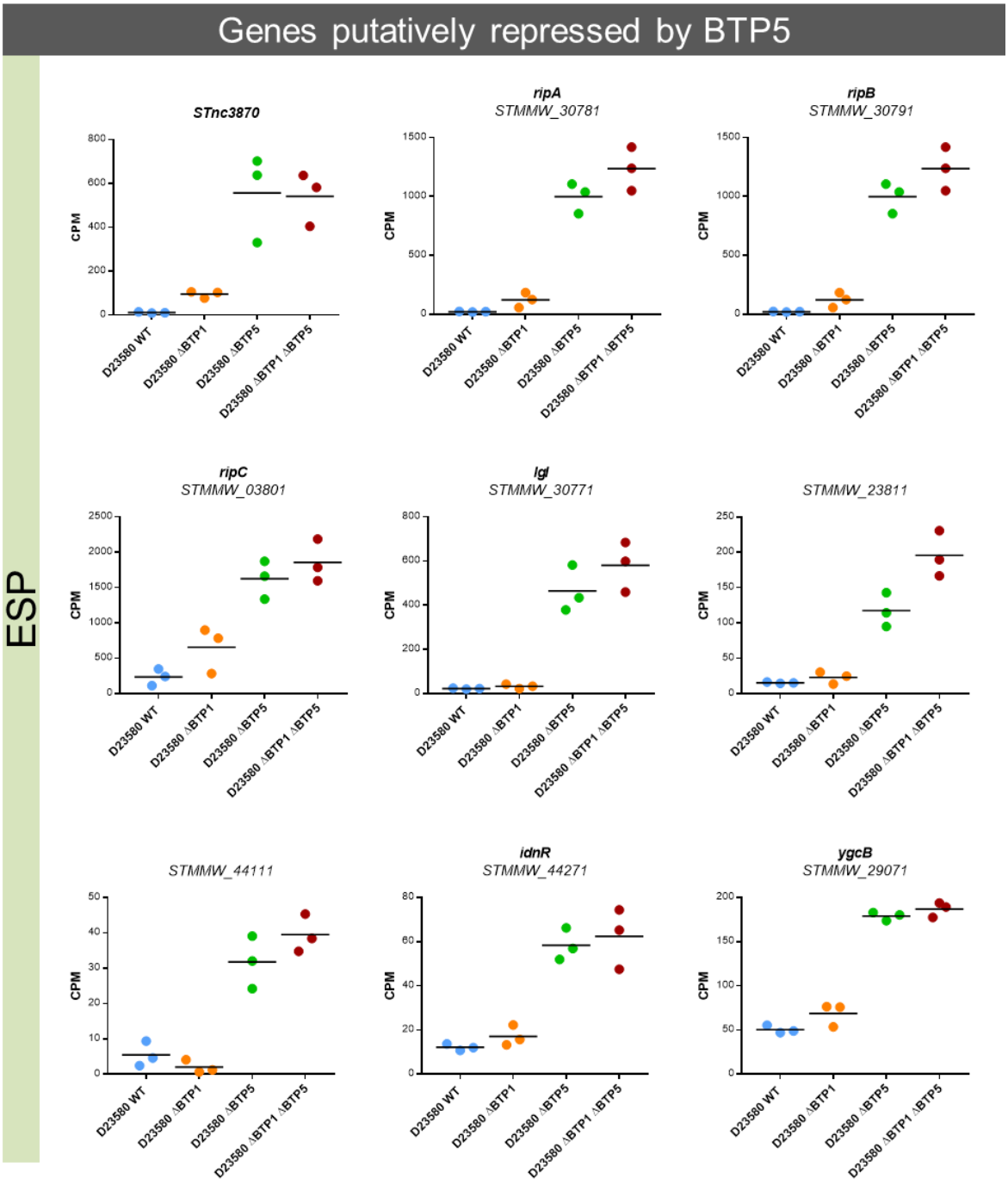



**Figure 5.18 Genes putatively repressed by BTP5 in ESP.** Counts per million (CPM) data for all mutants with three biological replicates of RNA-seq in the ESP growth condition are presented

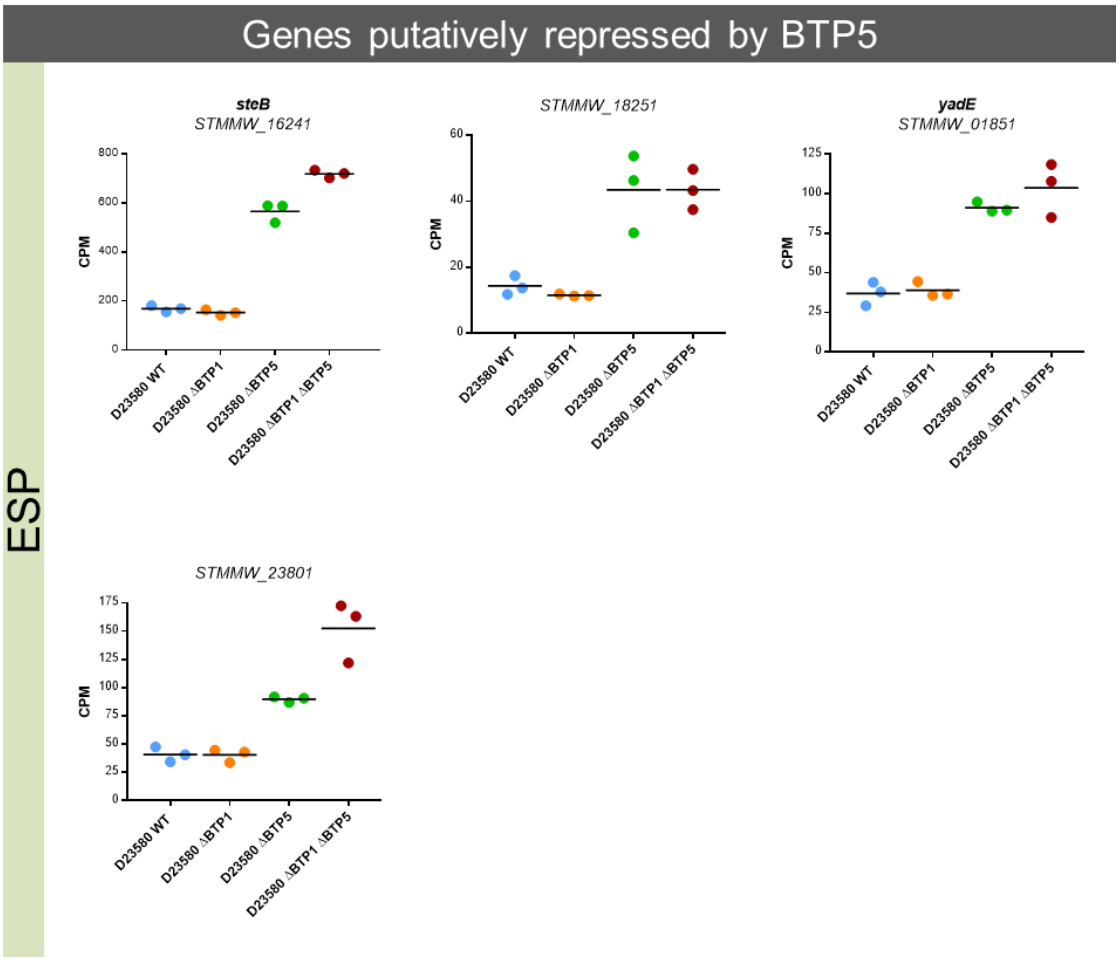



**Figure 5.18 (continued) Genes putatively repressed by BTP5 in ESP.** Counts per million (CPM) data for all mutants with three biological replicates of RNA-seq in the ESP growth condition are presented.

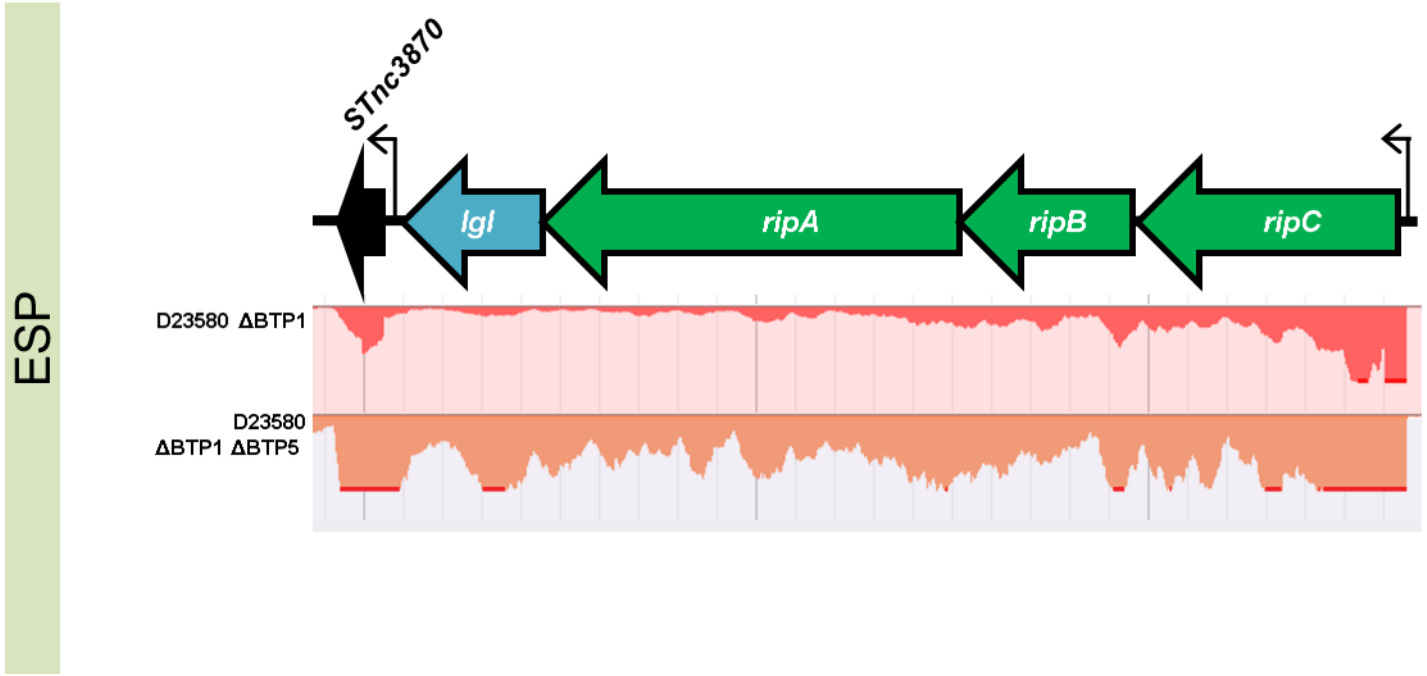



**Figure 5.19 Normalised RNA-seq read coverage plot of the ripABC, lgl and STnc3870 region of D23580 in ESP.** Mapped sequence reads are shown for D23580 ΔBTP1 and D23580 ΔBTP1 ΔBTP5.

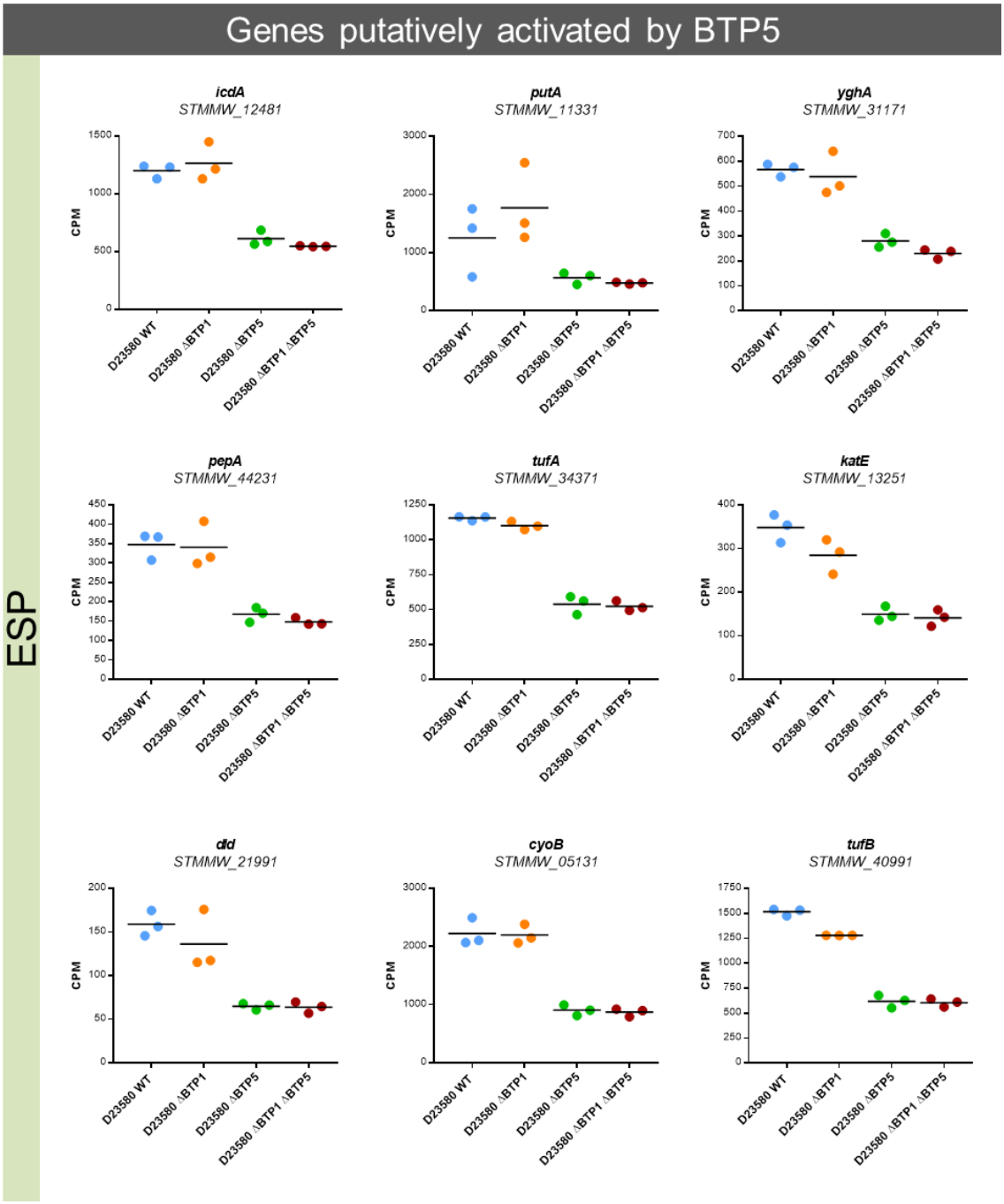



**Figure 5.20 Genes putatively activated by BTP5 in ESP.** Counts per million (CPM) data for all mutants with three biological replicates of RNA-seq in the ESP growth condition are presented.

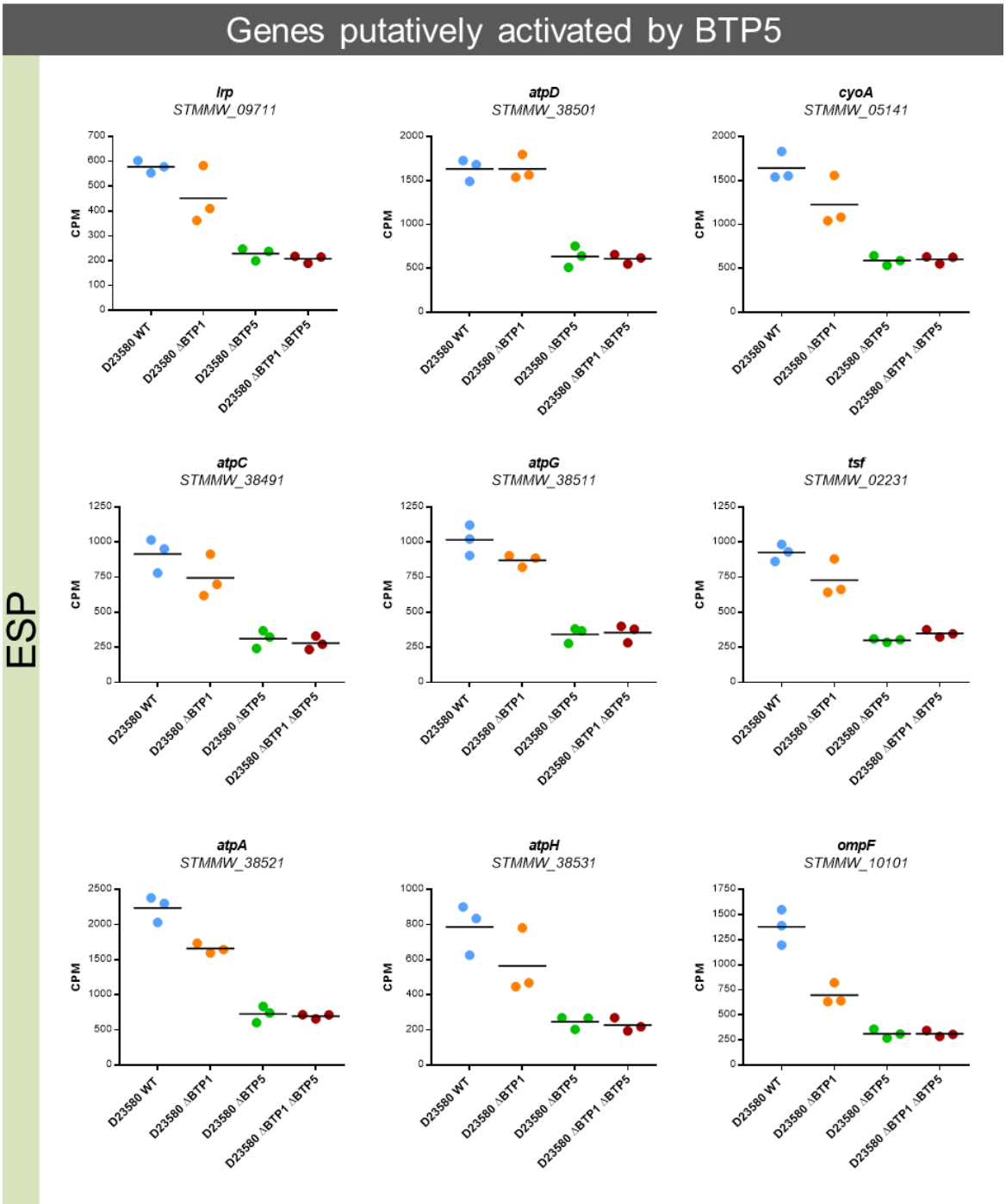



**Figure 5.20 (continued) Genes putatively activated by BTP5 in ESP.** Counts per million (CPM) data for all mutants with three biological replicates of RNA-seq in the ESP growth condition are presented.

### 5.3.4 An overall picture of the transcriptomic contribution of BTP1 & BTP5

A summary of the putative impact of the BTP1 and BTP5 prophages of D23580 at the phenotypic level is shown in Figure 5.21. Overall, the BTP5 prophage had the largest effect on global gene expression, and the BTP1 prophage did not mediate many transcriptomic changes. No significant effects of either prophage were observed in the inSPI2 condition. The only effect of the BTP1 prophage in anaerobic growth was a putative activation of the glycosyltransferase operon *gtrABCa*, whilst the presence of both prophages in anaerobic growth led to a putative repression of *gtrABCa.* In ESP, the BTP5 prophage putatively mediated a number of effects including activation of the proline dehydrogenase *putA,* the non-functional catalase gene *katE*, and many genes in the ATP synthase operon (*atpACDGH*). Putative repression effects of BTP5 in ESP include the down-regulation of the itaconate degradation genes *ripABC*, the uncharacterised sRNA STnc3870 and the outer membrane protein encoding gene *ompF*.

Confirmation of the transcriptomic changes mediated by the BTP1 and BTP5 prophages could not be pursued due to time constraints. Further experiments will prove critical to understand the contribution of the novel prophages to the biology of African *S.* Typhimurium ST313.

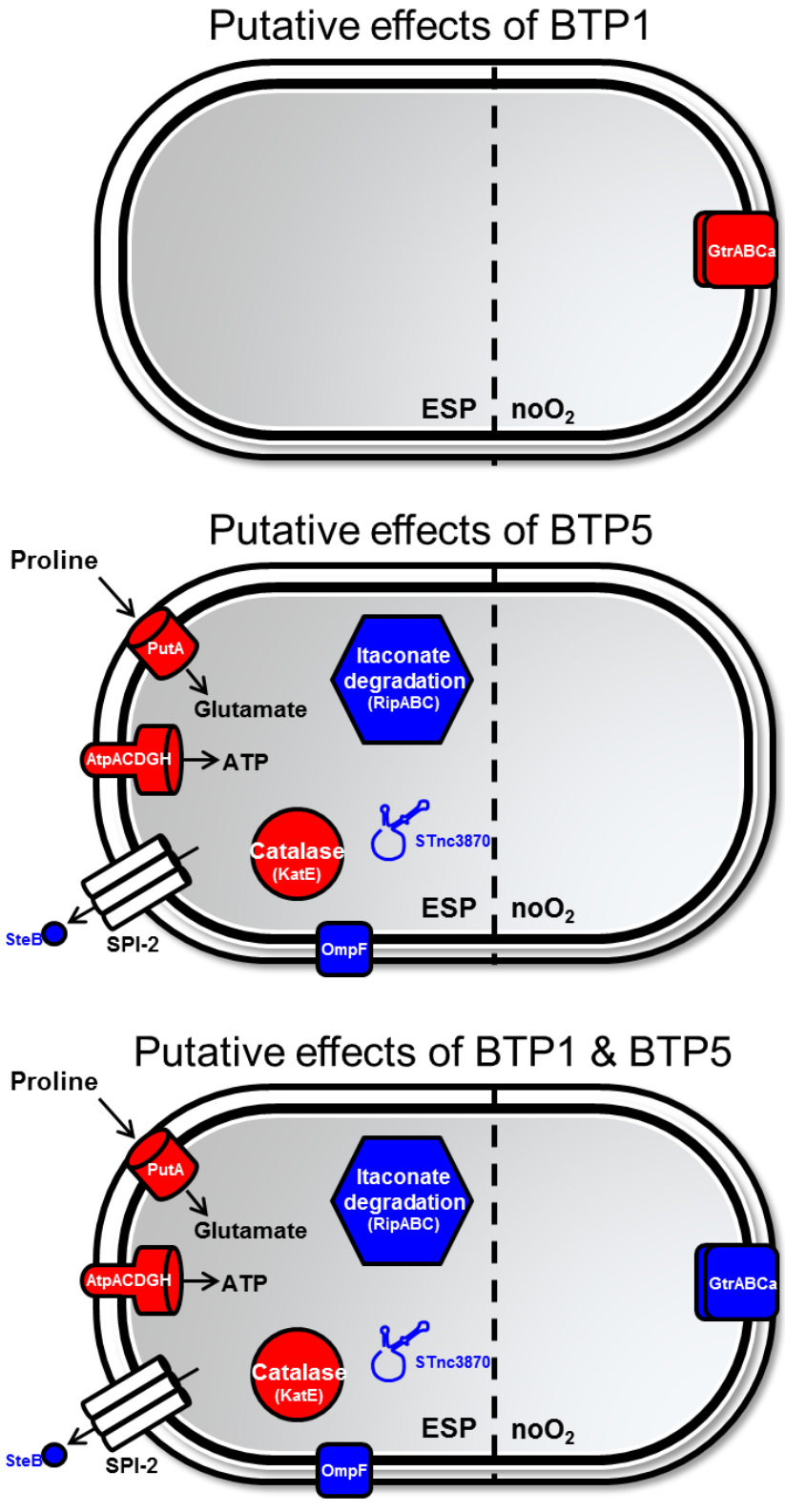



**Figure 5.21 Summary of the putative phenotypic effects of BTP1 and BTP5 based on prophage-associated transcriptomic changes of global expression levels identified.** Red indicates up-regulated functions whereas blue indicates down-regulated functions.

## 5.4 Discussion

In this chapter I have investigated the hypothesis that novel prophages associated with African ST313 lineages, BTP1 and BTP5, modulate the biology of their host bacterium at the transcriptional level. The rationale for this hypothesis was the evidence for co-evolution of the African ST313 lineages that cause iNTS disease with the BTP1 and BTP5 prophages. The results of Chapter 4 suggested that these prophages were acquired independently by the two African ST313 lineages in sub-Saharan African, as they were absent from the UK-ST313 strains which are more closely related to Lineage 1 or Lineage 2 than the African lineages are to each other.

The transcriptome of the D23580 prophage regions were generally consistent with the findings on the functionality of the prophages described in Chapter 3. BTP1, a prophage which exhibits an unusually high level of spontaneous induction (Chapter 3.3.4), also showed an unusually high level of transcriptional activity for a prophage region. Canonically, the prophage genes are transcriptionally repressed during lysogeny, apart from those genes whose products maintain prophage lysogeny, such as the gene encoding the CI repressor in lambdoid prophages (Ptashne, 2004) or genes with accessory functions that affect the lysogen. Studies investigating the gene expression of phage λ lysogens using microarray or ribosome profiling analysis have shown that most transcriptional activity arises from the *cI* gene encoding the repressor, the *rexA, rexB* encoding a superinfection immunity factor, and the *lom* and *bor* genes which encode virulence factors (Chen et al., 2005; Liu et al., 2013; Osterhout et al., 2007).

Unlike the other prophage regions of D23580, transcription of the BTP1 structural genes was observed in almost all growth conditions tested. It is impossible for a lysogen to constitutively express phage structural and lytic genes, because once the prophage molecular switch has moved to lytic from lysogenic replication, the terminal consequence is cell death (Ptashne, 2004). Therefore, if the lytic gene expression observed in the RNA-seq data occurred in the entire cellular population, we would expect to see a population collapse as ultimately sufficient lysis proteins were accumulated to initiate cell lysis. However, the BTP1 prophage lysogen (D23580) exhibits normal growth dynamics that are comparable to strains not lysogenised by the BTP1 prophage. I hypothesis that the lytic gene expression observed in the RNA-seq data is the consequence of averaging gene expression across a heterogeneous population which has very high lytic gene expression in the subset of the population undergoing spontaneous BTP1 prophage induction. This gene expression appears as low-level expression when it is averaged across the majority of the population in

which the BTP1 lysis genes are transcriptionally repressed. Consistent with this hypothesis, the remaining D23580 prophages; Gifsy-2, ST64B, Gifsy-1 do not exhibit significant spontaneous induction levels and show very little lytic gene expression in the majority of conditions tested. I propose that the absolute expression level of prophage structural genes could therefore be used as an *in silico* marker of the fraction of the population undergoing lytic prophage replication.

As well as providing insight into the replication state of the D23580 prophages, the transcriptome maps also allow the identification of putative accessory regions; genes (or transcripts) that are expressed during lysogenic replication and so may not be necessary for phage function. Such regions are of importance in regard to understanding pathogen disease tropism, as they represent 'smoking guns' that could function as virulence factors. The transcriptome maps of prophages BTP1 and BTP5 show several more regions of transcription than are theoretically necessary in a lambdoid prophage to maintain lysogeny (typically the *cI* repressor only) (Hendrix et al., 1983). A number of genes of the BTP1 prophage show an expression pattern consistent with an accessory function, including *ST313-td, pid, gtrA* and *gtrC*. In the BTP5 prophage, fewer 'smoking guns' were identified, as the majority of regions with expression patterns indicative of accessory function corresponded to genes with homology to genes with reported phage function, such as the two adjacent genes STMMW_32031 and STMMW_32041 which are homologous to the *tum* antirepressor or *orf97* of coliphage186 (Shearwin et al., 1998).

However, it cannot be ruled out that phage genes expressed during lysogeny may serve a dual purpose by additionally regulating the core genome. For example, recently the *cII* gene of Stx-phage $\Phi 24_B$ was shown to control acid response genes in the lysogen as well as performing the phage-related function of controlling expression of the *cI* repressor (Veses-Garcia et al., 2015). One confounding result was the high expression of the putative BTP5 antirepressor gene in the transcriptome data. Expression of the antirepressor gene should theoretically cause prophage induction, but figures 5.9 and 5.10 show the BTP5 structural genes are not wholly expressed. This contradiction casts doubt on the putative function of STMMW_32041 as an antirepressor, and further investigation is required to understand the molecular mechanisms governing the lysis-lysogeny control of the BTP5 prophage.

A number of putative RNA transcripts in the BTP1 prophage which did not correspond to protein coding sequences were observed to have an expression pattern consistent with an accessory function, including the 10 putative novel ncRNAs (Hammarlöf et al.,

2017). Several prophage-encoded ncRNAs have been implicated in bacterial virulence, for example, the Gifsy-1 prophage encoded sRNA IsrJ which is an ~74 nt transcript reported to be required for efficient invasion of *Salmonella* into non-phagocytic cells and effector translocation by the SPI-1 T3SS (Padalon-Brauch et al., 2008). Prophage-encoded ncRNAs are also known to mediate non-virulence accessory functions. For example, the *sas* asRNA of phage P22 induces a translational switch between distinct peptides encoded by the *sieB* gene, which is critical to the function of the SieB superinfection exclusion system (Ranade and Poteete, 1993). Lastly, the phage λ sRNA OOP inhibits CII protein synthesis thereby pushing the phage molecular decision towards lysis, rather than lysogeny (Krinke and Wulff, 1987).

The second approach taken in this chapter was a transcriptional 'target-hunt' experiment to identify genes regulated by the BTP1 & BTP5 prophages. RNA samples were collected from the D23580 WT, D23580 ΔBTP1, D23580 ΔBTP5 and D23580 ΔBTP1ΔBTP5 in three different growth conditions. A number of putatively prophage-regulated genes were identified, however, more transcriptional changes were found in the D23580 ΔBTP5 mutant compared to D23580 ΔBTP1. In order to increase the probability of finding significant transcriptional changes, genes changing in four different comparisons were cross-referenced to highlight the changes which made the most biological sense (e.g. genes that changed in both the single prophage mutant and the double prophage mutant). Using this approach, it was found that the expression of the core genome *gtrABCa* operon was affected differentially by the presence of the BTP1 and BTP5 prophages. The *gtrABC*a operon demonstrated a complex interaction with the BTP1 and BTP5 prophages, and appeared to be highly up-regulated in the presence of BTP1, not affected by the presence of BTP5 alone, but moderately down-regulated in the presence of both prophages together.

All the findings from the transcriptional target hunt approach require further validation in the form of qPCR or phenotypic experiments to confirm these findings, however the hypothesis that there may be crosstalk between the prophages and the core-genome encoded *gtr* system is compelling. The *gtr* operons are thought to be phage-derived systems, utilised by phages to modulate the receptor biology of the host to prevent the infection of self-lysogens or to prevent self-adsorption upon lysis (Davies et al., 2013). In *Salmonella* the phenomenon of polylysogeny is common, and the majority of *Salmonella* strains are lysogenised with multiple temperate phages (Thomson et al., 2004). Crosstalk between *Salmonella* prophages has been demonstrated by a study that showed the antirepressor of Gifsy phages can recognise non-cognate

repressors, allowing one Gifsy prophage to be induced by the natural induction of another in the same cell. It is therefore plausible that another element of prophage crosstalk could be the antagonism, or even cooperation of accessory functions that modulate the physiology of the host. It is worth noting that the BTP1 phage encodes its own O-antigen modification system, *gtrACc* (though the function of these enzymes is to acetylate, rather than glycosylate the O-antigen) (Kintz et al., 2015).

Amongst the other putatively prophage-regulated genes identified here were the *ripABC* operon which appeared to be down-regulated in the presence of the BTP5 prophage in early stationary phase. The *ripABC* operon is thought be involved in *Salmonella* infection as it's expression is highly activated inside macrophages (Srikumar et al., 2015). The RipABC proteins are associated with itaconate metabolism and Lgl is responsible for methylglyoxal detoxification (Chakraborty et al., 2014; Sasikaran et al., 2014). The modulation of expression of this operon by the BTP5 prophage therefore deserves further investigation.

Perhaps the most counter-intuitive result from this experiment was the putative up-regulation of the *katE* gene in the presence of BTP5 in early stationary phase. The KatE protein has been found to be highly abundant in D23580 WT cells in early stationary phase compared to ST19 strain 474 (which does not harbour the BTP5 prophage) (R. Canals- Pers. comm.). Though these two results are consistent, and strongly suggest the BTP5 prophage is activating expression of the *katE* gene, the finding is confounded by the fact that *katE* encodes a non-functional catalase enzyme in African ST313 strains due to a single non-synonymous SNP (Singletary et al., 2016). What could be the biological advantage of activating the expression of a non-functional protein? It is a possibility that the modified KatE protein of African *S.* Typhimurium ST313 performs a different function mediated by the substitution of a single amino acid.

Overall, the results of this chapter provide evidence that the prophages are not discrete transcriptional units, but undergo regulatory interactions with the core genome and each other. The presence of genes exhibiting expression patterns consistent with an accessory function, and the putative dysregulation of chromosomal genes upon removal of the BTP1 & BTP5 prophages represents a promising basis for further research into the impact of the prophages at the phenotypic, as well as transcriptional level.

# Chapter 6

## Evidence for lysogenic conversion of African *S.* Typhimurium ST313 by prophages BTP1 & BTP5

## 6.1 Introduction

The transcriptomic data presented in Chapter 5 identified a number of highly expressed regions of the BTP1 and BTP5 prophages that could encode accessory genes because they are expressed during lysogeny. Chapter 3 described the unusual spontaneous induction phenotype associated with the BTP1 prophage lysogen and speculated that the spontaneous induction may contribute to the infection biology of African *S.* Typhimurium ST313 isolates. An explanation for the putative evolutionary convergence of prophage profiles between the two African ST313 lineages detected in Chapter 4 could be that prophages BTP1 and BTP5 contribute to the niche adaptation of the iNTS-associated lineages in sub-Saharan Africa. Temperate phages are known to increase their own evolutionary fitness by encoding genetic elements that benefit the host in a phenomenon known as lysogenic conversion (Brüssow et al., 2004). In this chapter, I investigate the hypothesis that novel prophages BTP1 and BTP5 contribute to the biology of African ST313 lineage using a molecular genetic approach.

### 6.1.1 Acknowledgement of the specific contribution of collaborators to the results described in Chapter 6.

The work described in this chapter is unpublished. Unless specified below, all work was completed by the Author.

| | |
|---|---|
| **Rocío Canals**<br>University of Liverpool, UK | Completed macrophage infection experiments together with the Author |
| **Lizeth Lacharme-Lora & Paul Wigley**<br>University of Liverpool, UK | Conducted chicken infection experiments together with the Author |
| **Nico Wenner**<br>University of Liverpool | Constructed prophage deletion mutants in strain 4/74 (6.3.3) and STnc6030 expression plasmid (6.4.1) |

## 6.2 Assessment of virulence phenotypes associated with prophages BTP1 & BTP5

### 6.2.1 Motility phenotypes of prophage mutants

Motility is linked to the virulence of many bacterial pathogens including *Salmonella* (Ottemann and Miller, 1997; Wang et al., 2004). Additionally, motility and motility-related phenotypes have previously been shown to differentiate African *S.* Typhimurium ST313 strains from gastrointestinal-associated ST19 strains (Carden et al., 2015; Ramachandran et al., 2015). To determine whether the novel prophages BTP1 & BTP5 may contribute to motility differences in ST313, the swimming motility of prophage mutants was assayed.

Motility assays showed that ST19 strain 4/74 was significantly more motile than strain ST313 strain D23580 (P=0.03) (Figure 6.1A), as has been reported previously (Carden et al., 2015; Ramachandran et al., 2015). However, there was no significant difference between the motility of D23580 WT, D23580 ΔBTP1::FRT and D23580 ΔBTP5::FRT. These results show that the novel prophages do not affect motility.

### 6.2.2 Intra-macrophage replication of prophage mutants

Successful intracellular replication within phagocytic cells such as macrophages is a defining part of *Salmonella* infection in mammalian hosts that allows bloodstream-associated *Salmonella* to evade host immune defences and disseminate through the lymph and bloodstream (Fàbrega and Vila, 2013). Consequently, an enhanced ability to replicate within phagocytic cells may indicate an adapation to an invasive, extra-intestinal niche (though this may not always be the case- Chapter 1.4.1). Several studies have reported differential phenotypes for African *S.* Typhimurium ST313 strains during proliferation assays in cultured mammalian macrophages (Carden et al., 2015; Herrero-Fresno et al., 2014; Ramachandran et al., 2015). To test whether the BTP1 or BTP5 prophages modulate the replication of ST313 in macrophages, the proliferation of prophage mutants was assayed in cultured murine macrophages.

The strains D23580 ΔBTP1::FRT and D23580 ΔBTP5::FRT replicated well inside macrophages over a 15 h infection period, relative to the attenuated negative control strain D23580 Δ*rpoE*::FRT. No significant difference was observed between the level of intra-macrophage proliferation for the prophage mutants and D23580 WT (Figure 6.1B). These results indicate that novel prophages BTP1 and BTP5 do not affect the ability of their *Salmonella* host to replicate inside macrophage cells.

### 6.2.3 Phenotypes in a chicken infection model

Animal infection models are arguably the most relevant and most practical method to assess the virulence potential of a bacterial pathogen in a human or animal host (Tsolis et al., 2011). *S.* Typhimurium ST313 strains have been reported to show enhanced virulence in a number of animal infection models including the chicken (Parsons et al., 2013). Therefore, the chicken model was used to test whether the novel prophages of ST313 are required for successful animal infection. Due to resource constraints, only data for the BTP1 prophage mutant were obtained.

The numbers of bacteria in the spleen of 2-week-old chickens were enumerated 2 days after initial oral infection (Figure 6.1C). The burden of bacteria in the spleen of chickens infected with either D23580 WT or D23580 ΔBTP1::FRT was similar. The number of infected vs. uninfected spleens in the D23580 ΔBTP1::FRT group was reduced compared to D23580 WT (2 splenic infected chickens in the D23580 ΔBTP1::FRT group vs. 5 splenic infected chickens in the D23580 WT group). However, the difference between the two strains was not statistically significant (P=0.11). Further experiments are required to elucidate the role of prophages BTP1 & BTP5 in the chicken infection model.

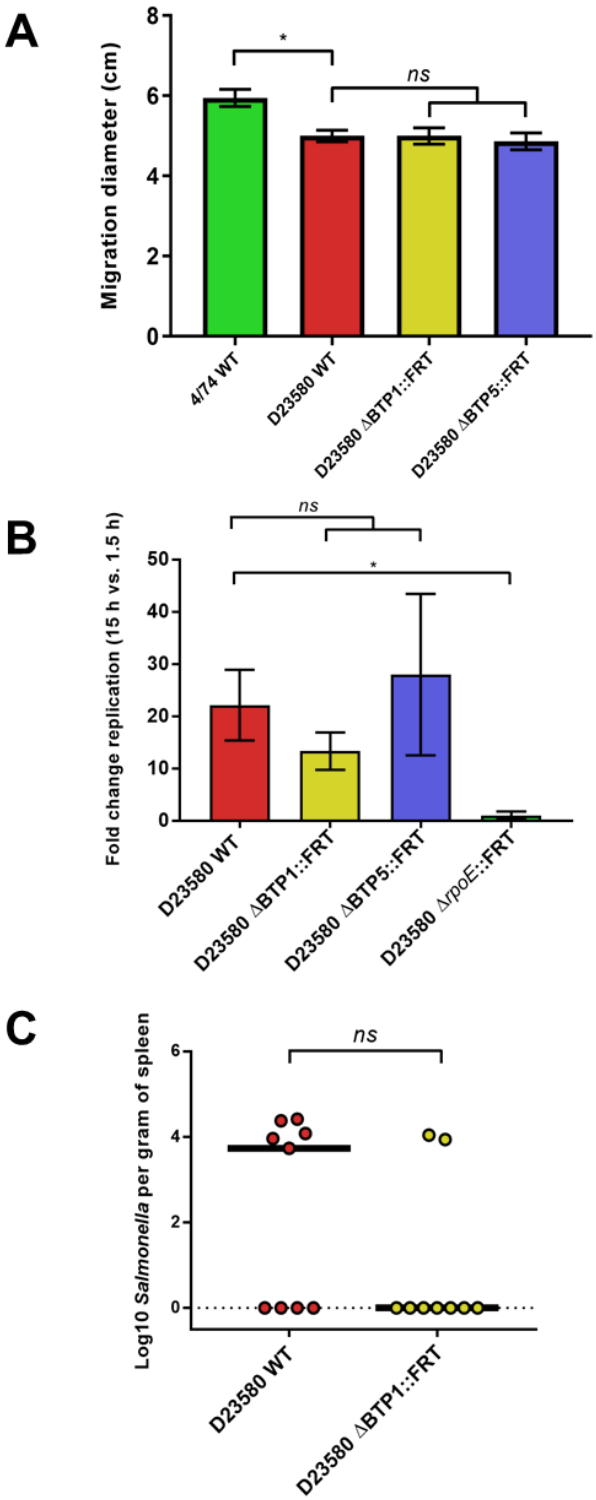



**Figure 6.1 Assessment of virulence-associated phenotypes of the BTP1 and BTP5 prophages.** A. Swimming motility of strain 4/74, D23580 WT, D23580 ΔBTP1::FRT and D23580 ΔBTP5::FRT. Migration diameter was measured after 5 hours (Chapter 2.3.2) B. Intramacrophage proliferation of D23580 WT, D23580 ΔBTP1::FRT and D23580 ΔBTP5::FRT. D23580 Δ*rpoE*::FRT is shown as a negative control. Proliferation is given as the fold-change in CFU/ml between 1.5 h vs 15 h of infection in RAW 264.7 murine macrophages (Chapter 2.10.1). C. Virulence of D23580 WT and D23580 ΔBTP1 in the chicken infection model. 2 week old chicks were inoculated orally with D23580 WT or D23580 ΔBTP1::FRT (Chapter 2.10.2). The CFU/ml of bacteria in the spleen was enumerated 2 days post infection.

## 6.3 Assigning function to the *ST313-td* gene of BTP1

The *ST313-td* gene of *S.* Typhimurium ST313 has previously been reported to confer a virulence phenotype in mice and cultured human macrophages (Herrero-Fresno et al., 2014). Transcriptome analysis of the BTP1 prophage showed that the *ST313-td* gene is co-transcribed from the same promoter as the *cI* prophage repressor gene from a promoter analogous in position to the $P_{RM}$ promoter of phage λ (Hammarlöf et al., 2017). The consequence of this transcriptional organisation is that the expression of the *ST313-td* gene is implicitly associated with prophage lysogeny, and this gene is unlikely to be expressed during lytic replication (when the Cro protein blocks transcription from $P_{RM}$ (Ptashne, 2004)). In phage P22 no additional genes are encoded between the *cI* gene and $P_L$ (the left-ward lytic promoter), whereas in phage λ this region contains the *rexAB* genes, which are also co-transcribed with the repressor gene. The *rexAB* genes are not required for phage Lambda replication, but, in the λ lysogen, mediate infection exclusion to certain phages, and increase the growth rate of λ lysogens in glucose-limited conditions (Lin et al., 1977; Parma et al., 1992). However, the *rexAB* genes are not conserved in other lambdoid phages and the genetic space between the end of the *cI* gene and the $P_L$ promoter has been noted to be a region of high mosaicism in lambdoid prophage (Degnan et al., 2007). It has been hypothesised that exchange of different alleles in this region may confer λ lysogens fitness in different niches (Degnan et al., 2007). Certainly, this hypothesis would be consistent with the reported contribution of the *ST313-td* gene to the virulence of African *S.* Typhimurium ST313. However, when the *ST313-td* gene was initially described, the significance of its genetic context within prophage BTP1 had not been appreciated. Therefore, the contribution of *ST313-td* towards the biology of the BTP1 prophage was investigated.

The *ST313-td* coding sequence is 924 bp long, encoding a putative protein of 308 amino acids. blastn analysis of the NCBI non-redundant nucleotide database shows the *ST313-td* gene is present in at least one isolate of *Salmonella enterica* serovars Typhimurium, Dublin, Anatum and Bardo. Additionally, a gene with 85% identity to *ST313-td* is present in a single isolate of *Klebsiella pneumoniae* (Accession number CP021960). The ST313-td protein sequence does not contain significant homology to any characterised proteins or domains (as judged by blastp analysis of the SwissProt database). However, the NCBI non-redundant protein database contains proteins with more than 90% identity to ST313-td in a number of Enterobacteriaceae genomes including isolates of *Salmonella enterica, Klebsiella pneumoniae, Cronobacter sakazakii* and *Citrobacter braakii.*

### 6.3.1 ST313-td confers immunity to infection by phage P22

Figure 6.2A shows a detailed transcriptomic context of the *ST313-td* gene region. In order to investigate whether the *ST313-td* gene confers infection immunity to heterologous phages like the λ *rexAB* system, two stop codons were introduced into the beginning of the coding sequence (Figure 6.2). This mutation was designed to prevent synthesis of functional ST313-td protein, but not affect the transcription of the gene. The strain containing the mutated *ST313-td* gene was designated D23580$^{ST313\text{-}td\text{STOP}}$. The methods used to construct the D23580$^{ST313\text{-}td\text{STOP}}$ strain can be found in Chaper 2.6.2.

The mutation of the *ST313-td* gene did not affect the spontaneous induction frequency of the BTP1 prophage, and culture supernatants from D23580 WT and D23580$^{ST313\text{-}td\text{STOP}}$ contained similar levels of BTP1 phage (Figure 6.3A).

To test the effect of the *ST313-td* premature stop codon mutation on phage immunity, a high titre P22 WT phage stock was plated onto bacterial lawns of D23580 WT and D23580$^{ST313\text{-}td\text{-}\text{STOP}}$. As has been observed previously (Kintz et al., 2015), D23580 WT showed a high degree of immunity to P22 infection, with only a faint zone of growth inhibition observed at the highest P22 concentration tested (undiluted stock) (Figure 6.3B). Contrastingly, for D23580$^{ST313\text{-}td\text{STOP}}$, which differs from D23580 WT by just 4 nucleotides, zones of confluent lysis were visible at lower phage dilutions, and individual phage plaques were seen at the lowest P22 phage dilution ($10^{-6}$) (Figure 6.3B). It has previously been described that LPS-acetyltransferase activity encoded by the *gtrC* gene of BTP1 mediates a high degree of immunity in a BTP1 phage lysogen to infection by P22 (Kintz et al., 2015), and other studies have shown that acetylation, as well as glycosylation, of the O-antigen rhamnose moiety can block P22 phage infection (Wollin et al., 1987). These results have identified the *ST313-td* gene as a second factor mediating immunity to P22 in the BTP1 lysogen.

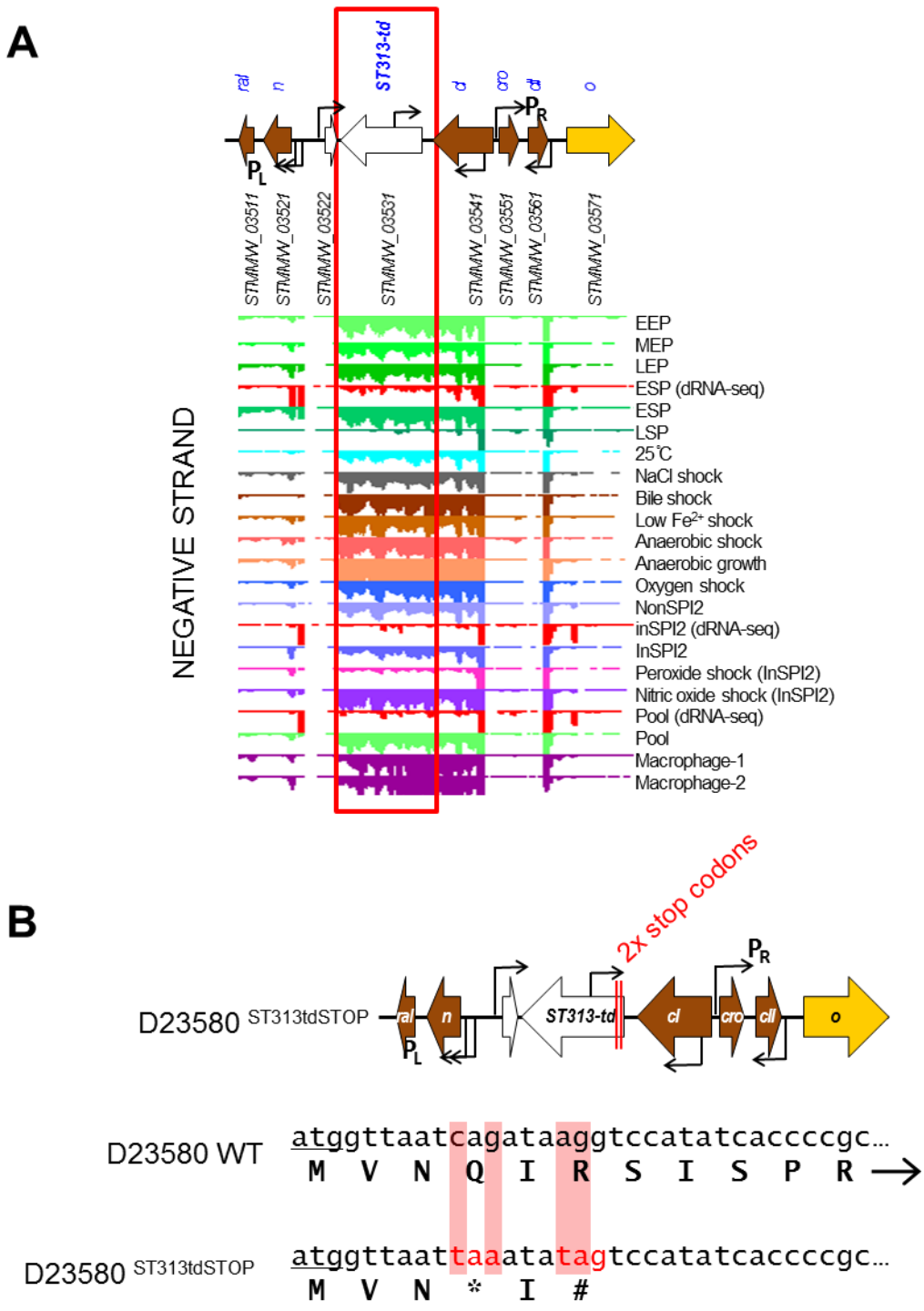



**Figure 6.2 A transcriptomic view of the *ST313-td* gene.** A. The transcriptomic region of the *ST313-td* gene expanded from Figure 5.1. B. The mutation strategy taken to study the role of the *ST313-td* gene. Two stop codons were introduced into the beginning of the coding sequence representing a total of 4 nucleotide substitutions (Chapter 2.6.2).

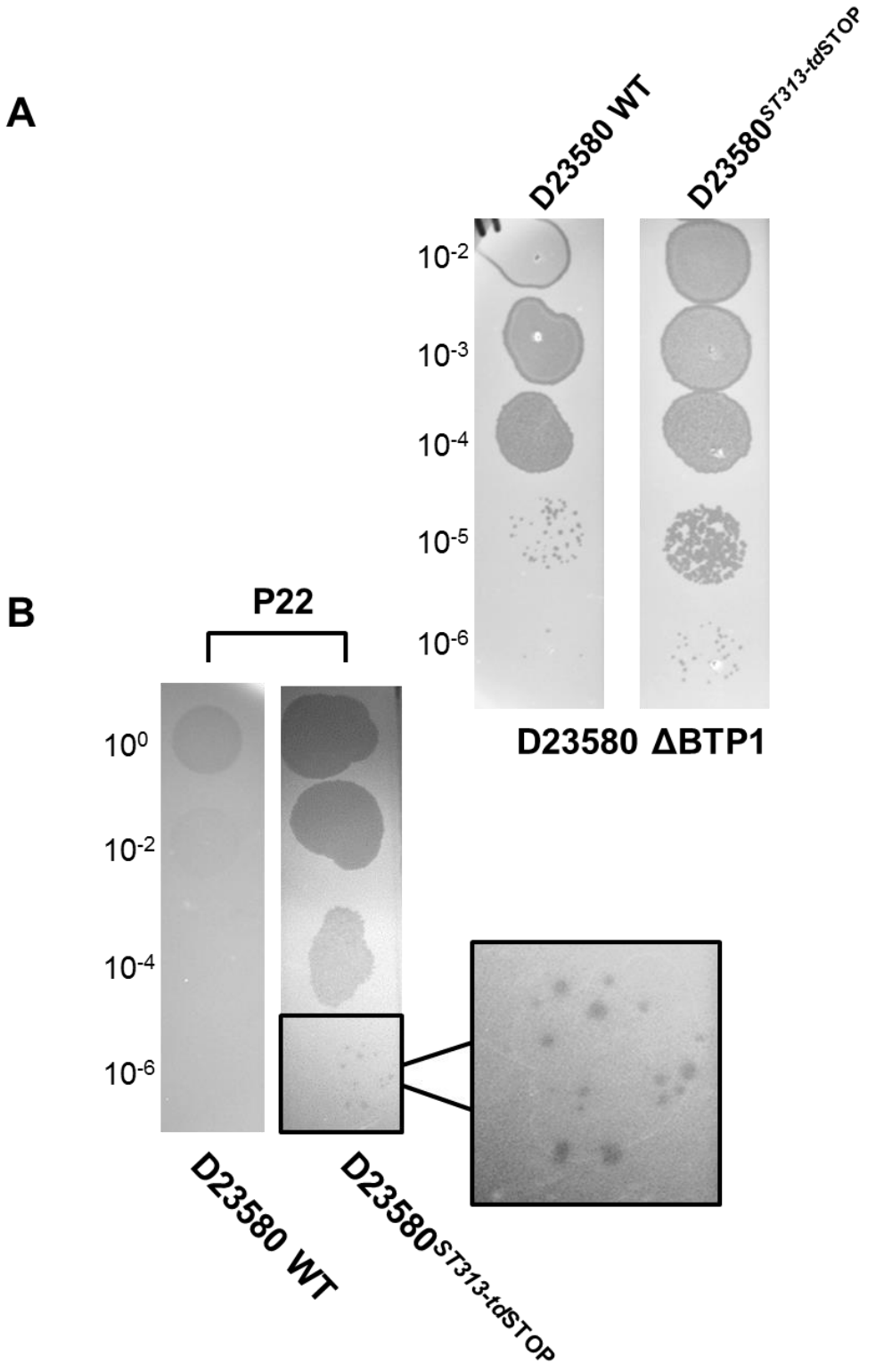



**Figure 6.3 The ST313-td protein is not required for spontaneous induction of prophage BTP1, but reduces susceptibility to P22 infection.** A. Titration of the over-night culture supernatant of strain D23580 WT and D23580 $^{ST313td}$STOP using D23580 ΔBTP1 as a host. B. Plaque assay of phage P22 using strain D23580 WT and strain D23580 $^{ST313td}$STOP as a host.

To gain further insight into the mechanism of ST313-td-mediated phage immunity, the WT and defective alleles of the *ST313-td* gene, were inserted into the $pP_L$ expression plasmid under the control of the strong $P_{LlacO-1}$ promoter (Figure 6.4A). A gentamicin resistant derivative of the $pP_L$ plasmid was constructed (Table 2.5, Table 2.7), as D23580 is resistant to ampicillin, which is the selectable marker in the original $pP_L$ (pJV300) plasmid (Sittka et al., 2007). $pP_L$ constructs result in over-expression of the inserted gene due to the high activity of the $P_{LlacO-1}$ promoter and the high copy number of the plasmid. Studying the WT and defective versions of the gene in a genetic background independent of the BTP1 prophage allowed interrogation of whether the *ST313-td* gene conferred immunity alone or in concert with other elements of the BTP1 phage. The defective, stop-codon-containing, version of the gene controls for the possibility that the *ST313-td* gene transcript, rather than the protein, contributes to the immunity phenotype. The plasmids, together with an empty-vector control plasmid, were transformed into a D23580 ΔBTP1 background. The $pP_L$-*ST313-td* plasmid conferred a high degree of immunity to P22 infection in the D23580 ΔBTP1 background, but the $pP_L$-*ST313-td*STOP plasmid did not provide any immunity to P22 infection, showing that the ST313-td protein alone is sufficient to mediate immunity to P22 (Figure 6.4B). Interestingly, the $pP_L$-*ST313-td* plasmid did not provide any immunity to BTP1 infection in the D23580 ΔBTP1 (Figure 6.4B). This result suggests that either the BTP1 prophage is not susceptible to infection interference by ST313-td, or that the BTP1 phage contains an immunity factor to prevent self-interference by ST313-td.

To control for interference, or contributing effects of other prophages in the D23580 genome, the experiment was repeated in a D23580 ΔΦ background, where all five full-length prophages of D23580 have been removed (Figure 6.4C). The effect of the $pP_L$-*ST313-td* plasmid on immunity to P22 and BTP1 was identical to the D23580 ΔBTP1 background, showing that the function of ST313-td is not dependent on the presence of other D23580 prophages.

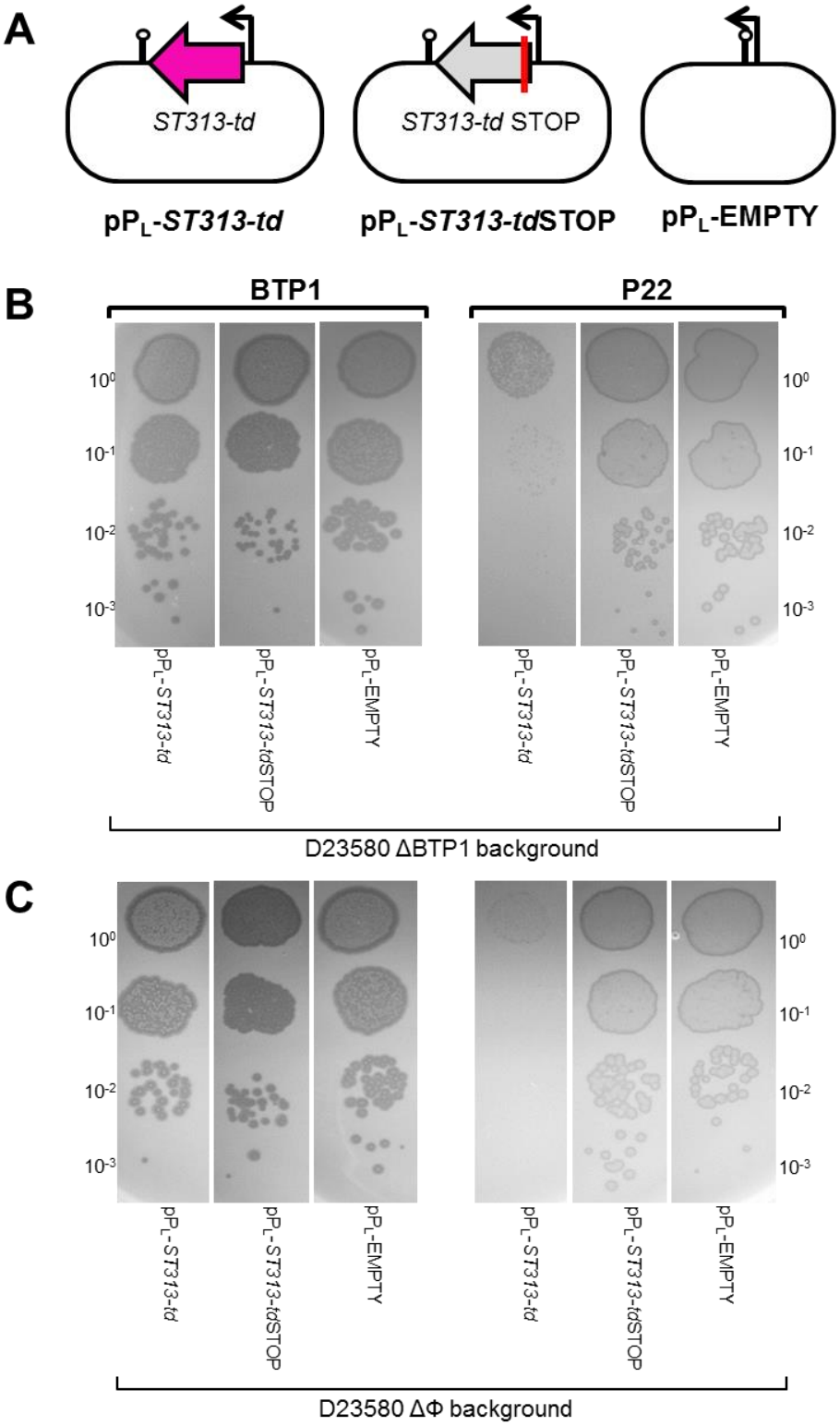



**Figure 6.4 Heterologous expression of ST313-td protein confers immunity to P22 infection but not BTP1 infection.** A. Schematic illustration of plasmid constructs used to investigate *ST313-td* function. B. Plaque assay of BTP1 phage and P22 phage on D23580 ΔBTP1 strains containing the plasmid constructs described in *A.* C. Plaque assay of BTP1 phage and P22 phage on D23580 ΔΦ strains containing the plasmid constructs described in *A.*

### 6.3.2 ST313-td does not affect P22 adsorption

The first event during the interaction between a phage and its host is the adsorption of phage to the host receptor. Therefore, the most direct way to gain immunity to phage infection is to prevent phage adsorption via receptor modification or loss (Seed, 2015). To determine whether the immunity conferred by ST313-td to P22 infection is mediated by adsorption interference, the adsorption of P22 to D23580 ΔBTP1 containing the three test plasmids described above (Figure 6.4A) was measured (as described in Chatper 2.4.6). Adsorption can be easily assayed by quantifying the free-phage in supernatant following addition of host bacteria; if adsorption occurs, there will be a reduction in free phage particles in the supernatant once bacteria are removed. Figure 6.5 shows that P22 could adsorb to all three test strains, regardless of the presence of the $pP_L$-*ST313-td* plasmid. The results indicate that ST313-td-mediated immunity to P22 infection is not associated with prevention of phage adsorption.

### 6.3.3 The ST313-td protein is not effective in a heterologous host background

The BTP1 prophage is strongly associated with African *S.* Typhimurium ST313 isolates, though the prophage has also been found in another sequence type of *S.* Typhimurium associated with passerine birds (ST568) (Chapter 3.3.2). To test whether the immunity conferred by the ST313-td protein is effective even in other *Salmonella* strains that do not natively contain the BTP1 prophage, the three test plasmids were transformed into *S.* Typhimurium strain 4/74. Unexpectedly, in the 4/74 background, the plasmid-borne *ST313-td* gene was not effective at preventing P22 infection (Figure 6.6A). Though the efficiency of plating of P22 on 4/74 $pP_L$-*ST313-td* appears slightly reduced compared to that on 4/74 $pP_L$-*ST313-td*STOP, the effect was much less dramatic to that observed in the D23580 strain background.

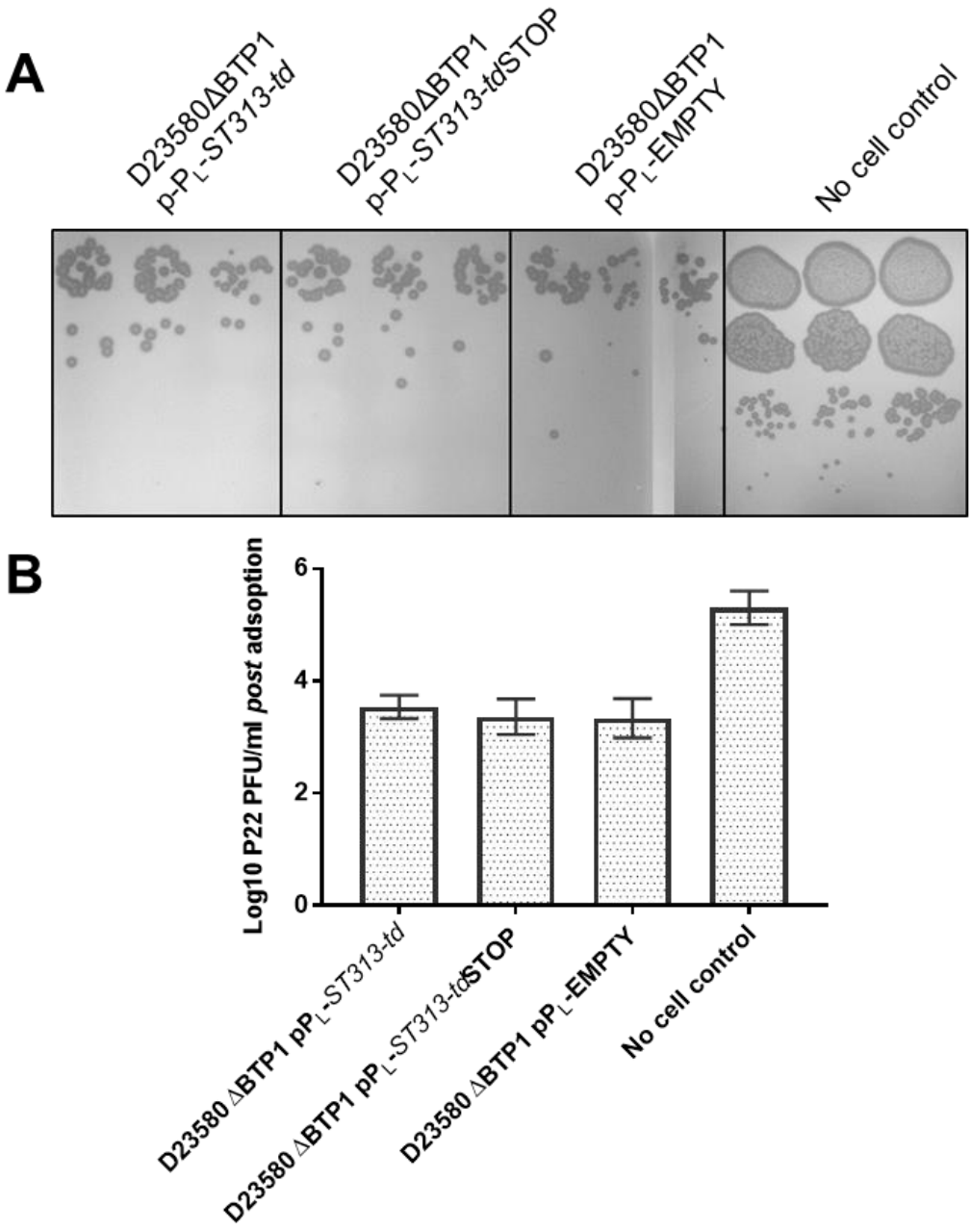



**Figure 6.5 Heterologous expression of ST313-td protein does not affect P22 adsorption to D23580 ΔBTP1.** A. Plaque assay of free phage after adsorption to D23580 ΔBTP1. B. Graph showing there is no difference in the number of phage remaining *post* adsorption to the test any of the test strains. The assays were conducted as described in Chapter 2.4.6.

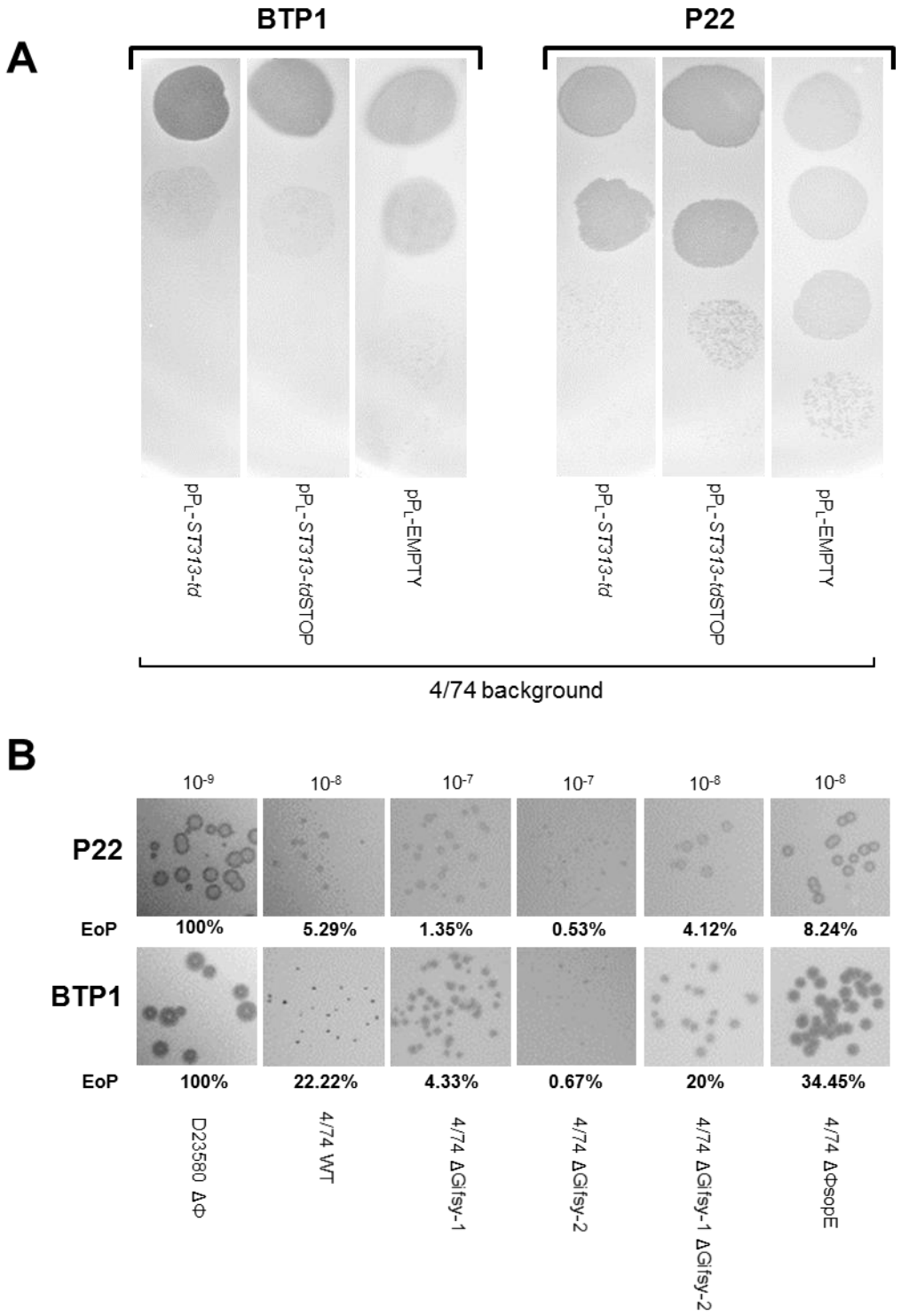



**Figure 6.6 ST313-td-mediated immunity to P22 infection is not functional in strain 4/74.** A. Plasmid constructs described in Figure 5.14 were transformed into strain 4/74 and these strains were used as host strains for plaque assays of BTP1 phage and P22 phage. B. The effect of 4/74 native prophages on immunity to infection by P22 or BTP1. Prophage mutants of 4/74 were assayed for the efficiency of plating (EoP) and plaquing morphology, upon six host strains specified at the bottom of the panel.

To test whether the poor efficacy of the ST313-td protein in strain 4/74 could be the result of interference by the native prophages of strain 4/74, BTP1 and P22 efficiency of plating and plaque morphologies were tested on various prophage mutants of strain 4/74 (Figure 6.6B). The efficiency of plating of both phages is reduced in strain 4/74 WT compared to control strain D23580 ΔΦ, which contains no prophages. This result suggests that the native prophages of 4/74 (or another element of the accessory genome, such as plasmids) confers some immunity to the BTP1 and P22 phages. Deletion of the Gifsy-1 prophage of 4/74 increased the size of BTP1 & P22 plaques but decreased the efficiency of plating (EoP) for both phages. Deletion of the Gifsy-2 prophage of 4/74 did not affect plaque morphology but reduced the EoP of BTP1 & P22 even more than the deletion of Gifsy-1. Surprisingly, deletion of both Gifsy phages produced larger plaque morphologies and did not affect the EoP. Lastly, deletion of the ΦsopE prophage of 4/74 significantly increased the EoP and resulted in plaque morphologies comparable to those obtained on positive control strain D23580 ΔΦ. This suggests that the ΦsopE prophage confers resistance to both the P22 and BTP1 phages. These findings raise the possibility that the mechanism of ST313-td-mediated immunity to P22 infection may be redundant in 4/74 because a similar system is conferred by the ΦsopE prophage.

More work is required to clarify the role of the *ST313-td* gene of prophage BTP1. In particular, the mechanism by which ST313-td confers infection immunity to P22 is not known. Other mechanisms to inhibit phage infection include: blocking DNA entry into the cell, destroying phage DNA once it is inside the cell (e.g. restriction modification systems), inhibiting phage assembly, and mediating cell death (e.g. abortive infection) (Seed, 2015). Assessment of the range of phages to which the ST313-td protein confers immunity is needed, and may offer clues into the mechanism of action.

## 6.4 Assigning function to the STnc6030 antisense transcript of BTP1

Small, non-coding RNAs (commonly referred to as sRNAs or ncRNAs) are short RNA transcripts (usually 50-500 nt) that do not encode for proteins (Gottesman and Storz, 2011). Instead, ncRNAs often have regulatory roles in bacteria. Recent advances in sequencing technologies have led to a huge expansion of research into the non-coding genome, and it is now widely accepted that ncRNAs are abundant in bacterial genomes and play an integral role in the molecular circuitry of the cell (Han et al., 2013; Luco, 2013; Vogel, 2009). In *Salmonella*, sRNAs have been shown to regulate the transcription of master regulatory genes, for example the sRNA ArcZ regulates the alternative sigma factor RpoS (σ38) which is required for *Salmonella* infection (Hébrard et al., 2012). By regulating the transcription of such important genes, sRNAs have a global impact upon *Salmonella* metabolism and virulence (Vogel, 2009).

Transcriptomic analysis of the prophages of D23580 identified a number of putative novel ncRNAs in the BTP1 prophage. One of the novel RNAs, designated STnc6030, was selected for further investigation for a number of reasons. STnc6030 is highly expressed in the majority of conditions for which RNA-seq data are available (Chapter 5.2.1). Additionally, the putative RNA is of an unusually large size >700 nucleotides in length and is positioned antisense to the BTP1 tailspike gene (STMMW_03901) and a putative DNA injection protein (STMMW_03891). It should be noted that the region where the STnc6030 asRNA is located is an area of unusual transcription, with 10 TSS (sense and antisense) defined within the tailspike gene alone. Due to the length and antisense position of the STnc6030 transcript, the putative RNA was hypothesised to be an asRNA species. The majority of asRNAs that have been identified in bacteria function as inhibitors of target RNA function (Wagner et al., 2002), and asRNAs are commonly found in accessory genome elements such as phages and plasmids (Thomason and Storz, 2010). I hypothesised that STnc6030 interacted with the transcription of the BTP1 tailspike gene, and the function of STnc6030 was further investigated.

### 6.4.1 STnc6030 encodes an approximately 500 nt transcript

Figure 6.7 shows a detailed view of the STnc6030 transcript region within the BTP1 transcriptome. The RNA is encoded antisense to the end of the putative DNA injection gene (*STMMW_03891*) and the tailspike gene (*STMMW_03901*). The beginning of the STnc6030 transcript corresponds to the beginning of RNA-seq reads mapping to

the tailspike gene on the sense strand, which may suggest antisense interference with the tailspike gene transcript.

The putative STnc6030 transcript as identified in the BTP1 transcriptomic data was cloned into the $pP_L$ expression plasmid under the control of the constitutive $P_{LlacO-1}$ promoter, and an empty vector control plasmid was used as a negative control in all experiments (Chapter 2.7). To confirm the presence of the STnc6030 transcript, an anti-STnc6030 riboprobe was synthesised in order to detect the RNA by Northern blot (Figure 6.8). The riboprobe was designed to cover the totality of the STnc6030 transcript as estimated from the transcriptomic data, allowing the detection of any transcripts corresponding to this region.

Transcripts corresponding to the STnc6030 RNA were detected by Northern blot using both an agarose and an acrylamide gel to allow optimal visualization and size estimation. The anti-STnc6030 riboprobe detected a number of transcripts for D23580 WT RNA, whilst no transcripts were detected by the probe using D23580 ΔBTP1 RNA, showing that the detected bands do not reflect non-specific probe binding. The largest transcript detected by the anti-STnc6030 probe in the D23580 WT RNA was approximately 500 nt in length, significantly shorter than the putative length identified from the transcriptomic data (786 nt). At least 2 other smaller transcripts were detected, of ~500 nt and 480 nt in length. However, these smaller transcripts may be the result of RNA processing or degradation product, and may not necessarily correspond to additional transcription start sites, as has been seen for other *Salmonella* sRNAs such as ArcZ (Papenfort et al., 2009).

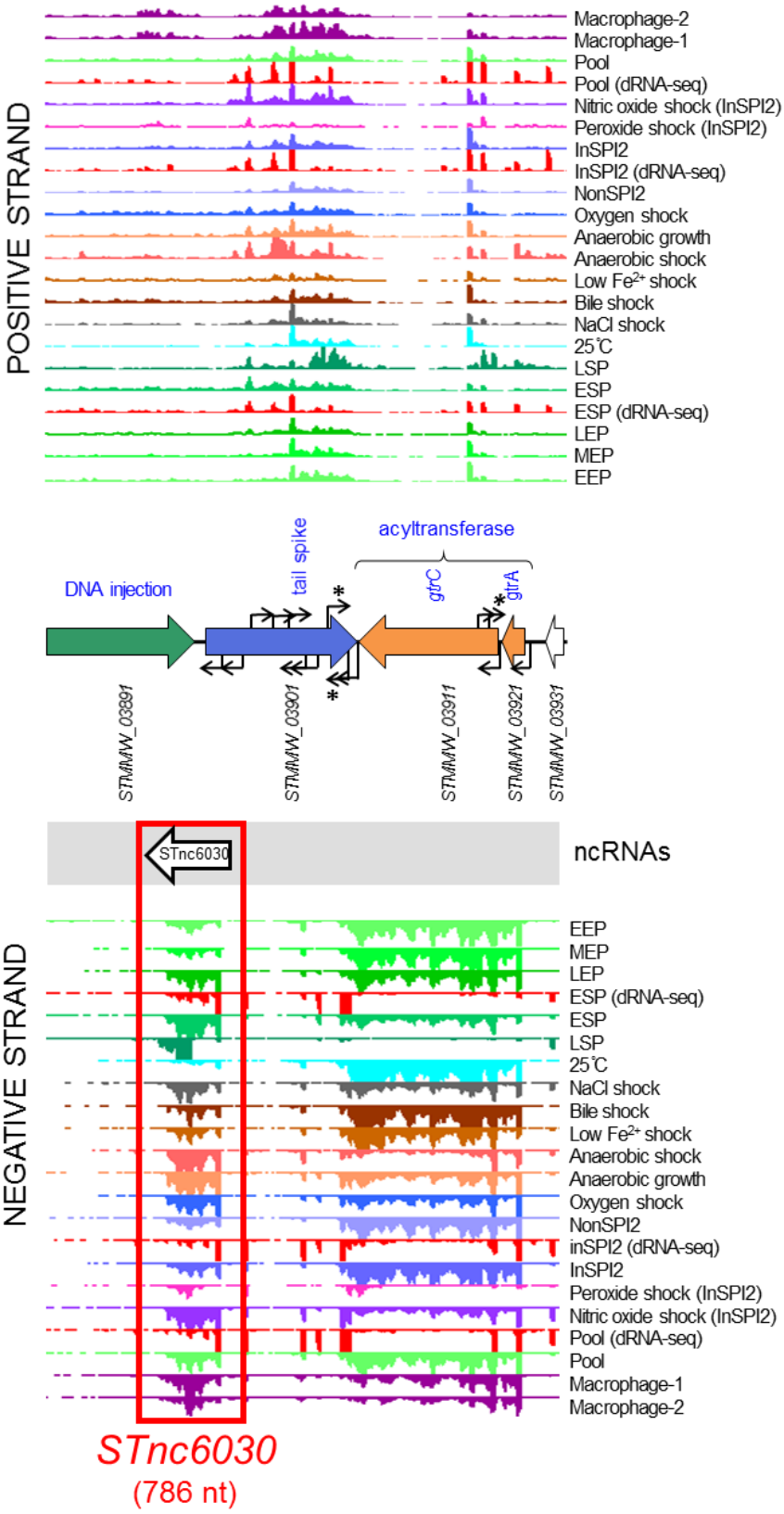



**Figure 6.7 The transcriptomic context of the STnc6030 ncRNA.** The transcriptomic data is expanded from Figure 5.1.

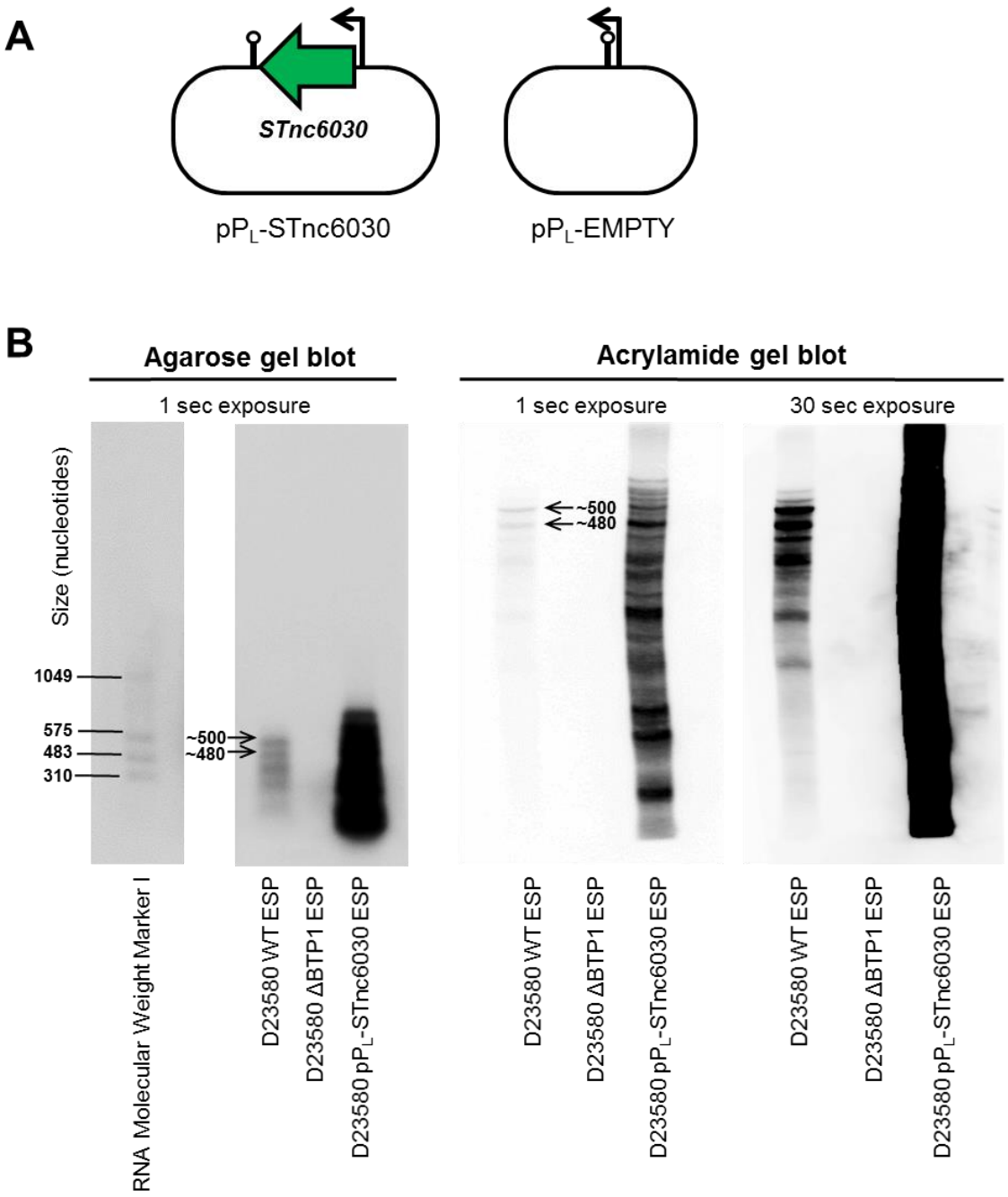



**Figure 6.8 The STnc6030 transcript can be detected by northern blot but is shorter than indicated in transcriptomic data.** A. Schematic illustration of plasmid constructs used to study the function of the STnc6030 transcript. A 786 bp region of DNA was cloned into the plasmid based on location coordinates inferred from the transcriptomic data (Figure 6.7). B. Detection of the STnc6030 transcript by northern blot on an agarose and acrylamide gel using an anti-STnc6030 DIG-labelled riboprobe (Chapter 2.9.5). The two most abundant transcripts detected by the anti-STnc6030 riboprobe are indicated along with the approximate size estimated from the molecular weight ladder employed on the agarose gel.

### 6.4.2 Heterologous expression of STnc6030 confers immunity to phage BTP1 infection in a naïve host

In order to test the function of STnc6030, the $pP_L$-STnc6030 and empty vector control plasmids were transformed into D23580 WT and D23580 ΔBTP1 backgrounds. Over-expression of the STnc6030 RNA did not affect the level of BTP1 spontaneous induction in the D23580 WT background, as there was no difference in the number of BTP1 phage in culture supernatants of D23580 WT, D23580 $pP_L$-STnc6030 and D23580 pEMPTY (Figure 6.9A).

Expression of STnc6030 in a naïve host (D23580 ΔBTP1) was found to mediate complete immunity to BTP1 infection (Figure 6.9B), but did not modulate susceptibility to P22 infection. These results were consistent with a regulatory mechanism in which STnc6030 targeted the sense transcript of the BTP1 DNA injection and tailspike genes. RNA-RNA interaction requires nucleotide complementarity, and it should be noted that the corresponding region of the P22 genome does not share similarity to BTP1 at the nucleotide level.

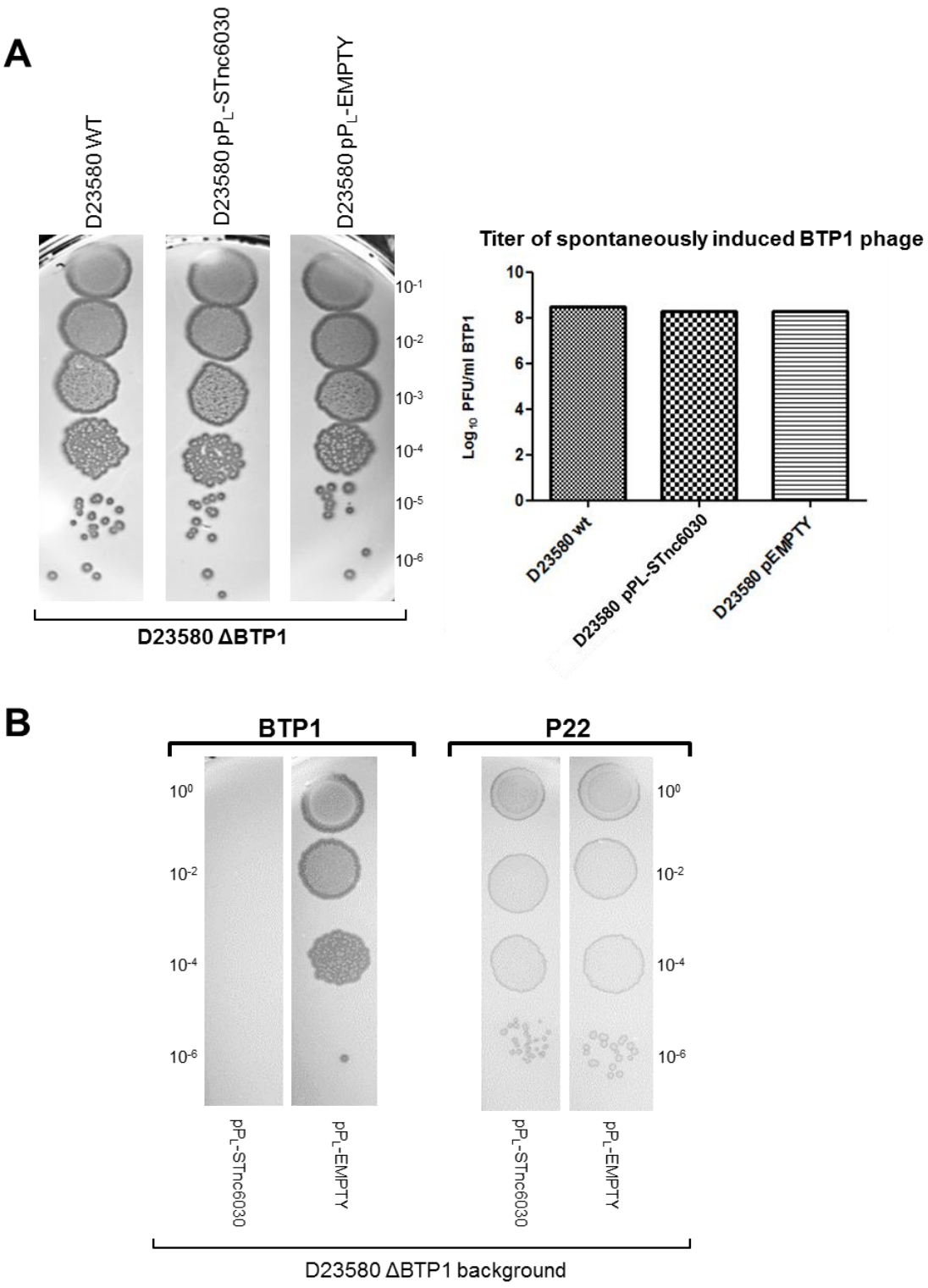


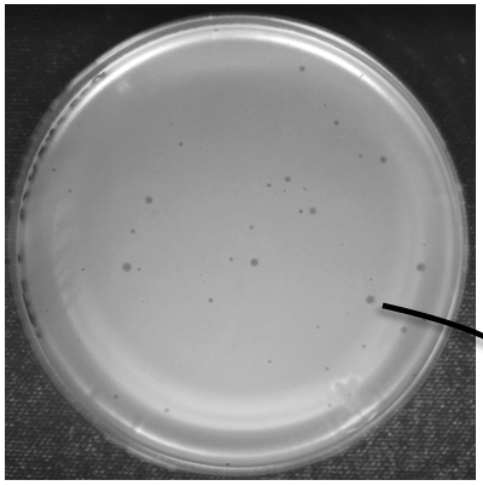

**Figure 6.9 Over-expression of the STnc6030 RNA in D23580 WT background does not affect BTP1 spontaneous induction, but heterologous expression in D23580 ΔBTP1 completely protects against BTP1, but not P22 infection.** A. Plaque assay of overnight culture supernatants of D23580 WT, D23580 pPL-STnc6030 and D23580 pEMPTY on host strain D23580 ΔBTP1. B. Plaque assay of BTP1 and P22 phage on D23580 ΔBTP1 strains containing the pPL-STnc6030 expression plasmid or the negative control plasmid.

### 6.4.3 The isolation of STnc6030 escape mutants indicate that the 3' end of the STnc6030 RNA is functionally active

To better understand the functionality of the STnc6030 transcript, a high titre BTP1 phage stock was screened for the presence of naturally occurring mutants resistant to interference by STnc6030 (Figure 6.10). The BTP1 phage stock was plated on lawns of D23580 ΔBTP1 containing the $pP_L$-STnc6030. Escape phages were estimated to occur at a frequency of $4x10^{-8}$ in the BTP1 phage stock, and were associated with varying plaque sizes (Figure 6.10A). Five escape phages were isolated and purified, and a nested PCR strategy was used to amplify the STnc6030 transcript region in the escape phages. Sequencing of the STnc6030 region revealed that each escape phage contained one single nucleotide polymorphism (SNP) relative to the BTP1 WT sequence (Figure 6.10B). The two escape phages associated with small plaque morphologies were found to harbour an identical SNP. A total of 4 unique SNPs were identified that conferred resistance to STnc6030-mediated immunity. Interestingly, the 4 SNPs clustered within a 36 bp region corresponding to the 3' end of the STnc6030 transcript, antisense to the putative DNA injection gene *STMMW_03891* (Figure 6.10C). These results suggest that the 3' end of STnc6030 is a functionally active "seed" region of the RNA, which interacts with the antisense target (*STMMW_03891*). The 4 SNPs also represent non-synonymous amino acid substitutions to the STMMW_03891 protein (Figure 6.10C). It is possible that the substitution associated with SNP identified in escape phages 3 and 4 results in reduced function of the STMMW_03891, resulting in a less efficient infection process and the reduced plaque size observed.

To confirm the biological role of the four nucleotides described in Figure 6.10B, in the future, complementary nucleotide substitutions should be constructed in the STMMW_03891 gene.

A



**Phage that are resistant to STnc6030 occur naturally at a frequency of $4x10^{-8}$ (STnc6030 escape phages)**

5 escape phages purified & STnc6030 region amplified and sequenced

D23580 ΔBTP1 $pP_L$-STnc6030

B

| Escape phage | Notes | Sequencing results | S / nS* | |
|---|---|---|---|---|
| 1 | Big plaques | Position 404055 in D23580 genome C → G | nS (T → K) | |
| 2 | Big plaques | Position 404076 in D23580 genome T → A | nS (V → E) | |
| 3 | Pin-point plaques | Position 404046 in D23580 genome G → A | nS (G → D) | Same SNP |
| 4 | Pin-point plaques | Position 404046 in D23580 genome G → A | nS (G → D) | |
| 5 | Big plaques | Position 404040 in D23580 genome T → C | nS (I → T) | |

C

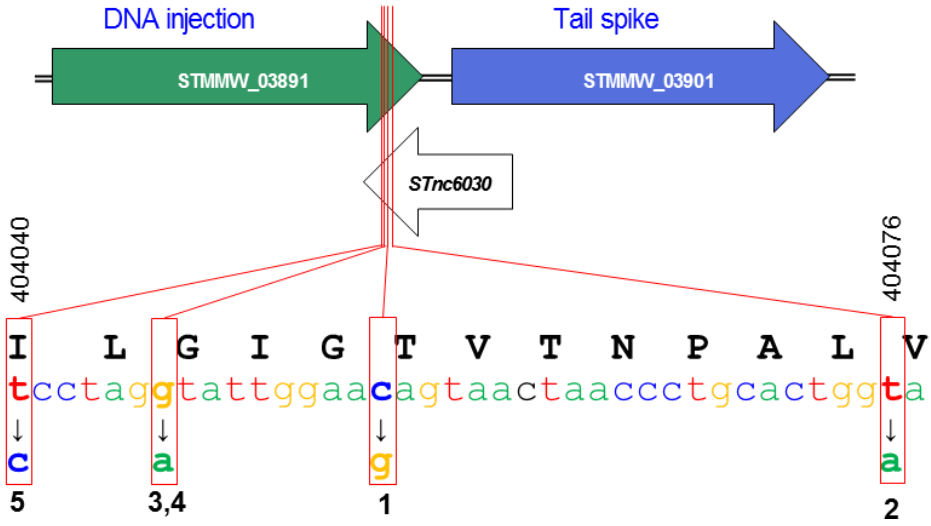



**Figure 6.10 Isolation of STnc6030-escape mutants**. A. A high titer BTP1 phage stock was used to identify naturally occurring BTP1 phage mutants that were immune to inhibition by STnc6030. Escape phages were estimated to occur at a frequency of approximately $4^{-9}$. B. Five escape phages of varying plaque morphologies were selected for sequencing. The sequence of the STnc6030 region of the escape phages identified SNPs. The position and substitution are shown. *S/nS = synonymous or non-synonymous with respect to the STMMW_03891 protein. C. The SNPs identified to confer immunity to STnc6030 interference were clustered within a 36bp region (shown) corresponding to the 3' end of the STnc6030 transcript and the 3' end of the STMMW_03891 gene.

### 6.4.4 The biological role of the STnc6030 asRNA remains unclear

The data presented here suggest that STnc6030 effectively excludes BTP1 phage replication specifically by antisense inhibition of the putative DNA injection gene *STMMW_03891*. However, the significance of the STnc6030-*STMMW_03891* interaction for the biology of the BTP1 prophage is not clear. Over-expression of STnc6030 had no effect on the production of BTP1 phage via spontaneous induction, and yet expression of the RNA in a naïve host mediated total immunity to BTP1 infection. How does the STnc6030 RNA differentiate between phage replication through induction and phage replication through infection? Further work is required to reveal the biological role of the STnc6030 RNA in the life history of the BTP1 prophage.

## 6.5 Discussion

To assess whether any of the highly-transcribed regions of the BTP1 or BTP5 prophage contribute to the virulence of African ST313 strains, three assays of virulence-related phenotypes were conducted using full prophage knock-out strains. Motility is an essential part of *Salmonella* virulence, and functioning flagella have been implicated in adherence to and colonisation of the gut (Josenhans and Suerbaum, 2002). Intracellular replication is a defining feature of *Salmonella* infection, and therefore replication within macrophages has been used as a model to explore *Salmonella* virulence (Srikumar et al., 2015). Lastly, infection of chickens is an *in vivo* infection model, and therefore a direct measure of virulence potential, that has previously been used to identify the enhanced virulence of African *S.* Typhimurium ST313 (Parsons et al., 2013). The whole prophage knock-out mutants of BTP1 and BTP5 did not show attenuated phenotypes compared to wild-type strains in motility or macrophage replication assays. In the chicken infection model, D23580 ΔBTP1::FRT showed a trend of slightly reduced colonisation of the spleen compared with D23850 WT, however the results were not statistically significant. The number of chickens infected in the two groups was low (n=9), and therefore it is possible that with higher sample numbers and more statistical power, this difference may reflect a real biological effect. Further work will be required to define the contribution of the BTP1 & BTP5 prophages in the chicken infection model.

Though the results of all three virulence-related assays did not show statistically significant effects of deleting either prophage, it remains possible that the prophages contribute to niche adaption of African *S.* Typhimurium ST313. A recent analysis revealed that African *S.* Typhimurium ST313 strains show only subtle differences to non-African ST19 strains in a number of animal infection models (Ramachandran et al., 2017), suggesting that an accurate model of ST313-mediated iNTS disease has yet to be established. As a defining characteristic of iNTS disease is the association with immune-compromised populations, perhaps the true disease potential of iNTS-associated ST313 strains will only be evident in an infection model which mimics the environment of an immune-compromised human host.

Of course, the novel prophages BTP1 & BTP5 may not contribute to the virulence of African ST313 strains, but rather contribute to another, as yet unknown, element of their adaption to the sub-Saharan African human niche.

An example of an accessory phage function that could contribute to the fitness of a bacterium in a specific environment is immunity to phage infection. Indeed,

superinfection immunity systems are commonly found in temperate prophages (Seed, 2015). In this chapter, the *ST313-td* gene was found to mediate immunity to P22 infection in the BTP1 lysogen. Though the mechanism of action remains unknown (except for the ruling-out adsorption interference) the presence of the *ST313-td* gene in the BTP1 prophage makes strain D23580 (and probably all African ST313 strains) highly resistant to P22 infection.

Though the *ST313-td* gene was originally reported to have a virulence function, increasing systemic infection in a mouse model, the results presented in section 6.3 suggest a more complex function. Perhaps *ST313-td* enhances virulence as well as conferring immunity to phage infection? Parallels could be drawn to the glycosyltransferase (Gtr) enzymes of P22-like phages, which modify the phage receptor of the lysogen to prevent the infection of self-phages and heterologous phages that utilise the same receptor (Davies et al., 2013). The mechanism of Gtr-mediated superinfection immunity is glycosylation of the bacterial O-antigen (the phage receptor). Whilst Gtr systems undeniably serve a phage-function by conferring immunity, there is increasing evidence that the modification by Gtr enzymes represent an essential component of bacterial surface modification for immune evasion (Kintz et al., 2017). Further, a number of core-genome encoded Gtr enzymes are thought to have been co-opted by the bacterial host from once functional prophages (Davies et al., 2013). In light of this, the mechanism by *ST313-td* confers immunity to phage infection may be the same by which it contributes to virulence. The adsorption-independent mechanism of the *ST313-td* remains to be described, and further work is required to understand the contribution of the ST313-td protein to the biology of African *S.* Typhimurium ST313 strains.

Like the *ST313-td* gene, the asRNA STnc6030 also conferred an immunity phenotype on the BTP1 lysogen. The ~ 500 nt transcript is located antisense to the 3' end of the putative DNA-injection gene *STMMW_03891* and 5' end of the tailspike gene *STMMW_03901*. Over-expression of STnc6030 completely abolished susceptibility to infection by BTP1, but not the related phage P22. The spontaneously induced titer of BTP1 phage was not affected when the STnc6030 transcript was overexpressed, suggesting that the asRNA does not interfere with spontaneously induced BTP1 replication. The data suggest that the functionally-active region of the asRNA is the 3' end (corresponding to antisense of *STMMW_03891*), as naturally occurring mutants of BTP1 phage that were immune to the action of STnc6030 had single nucleotide substitutions in a 3'-located 36 bp region. asRNAs in bacteria (also known as *cis*-encoded RNAs, are usually found antisense to annotated coding genes and thus

share extensive genetic complementarity with the corresponding transcripts (Gottesman and Storz, 2011). asRNAs can affect the stability of complementary mRNA transcripts by base-pair interactions. Native endoribonucleases such as RNase III and RNase E target double-stranded RNA complexes formed by base pairing of asRNA and target mRNAs (Thomason and Storz, 2010). As double-stranded RNA molecules are substrates for endoribonucleases, the effect of asRNA targeting is to increase the degradation of an mRNA transcript and so reduce levels of the gene product. Alternatively, base-pairing of two RNA species can block an RNase recognition site, leading to increased stability of the target mRNA (Thomason and Storz, 2010). If the mechanism of the STnc6030 asRNA is a base-pairing interaction with the DNA-injection gene *STMMW_03891*, that decreases the stability of the mRNA, how is the effect on phage immunity mediated? The consequence of this interaction would likely be the formation of BTP1 phage particles that lack the DNA-injection protein required for successful infection of a new host. In this scenario, the initial infected cell (e.g. D23580 ΔΦ $pP_L$STnc6030) would be lysed as a result of the termination of lytic replication, but the resulting phages would not be infective, raising the possibility that STnc6030 acts as an abortive infection (abi) system. However, as the asRNA is natively located within the BTP1 prophage, which contains other infection exclusion systems that prevent self-infection (the Gtr system), the biological role of the STnc6030 asRNA remains uncertain.

In conclusion, phenotypes associated with virulence could not be convincingly assigned to either the BTP1 or BTP5 prophage. It remains possible that BTP1 or BTP5 contribute to virulence in a way not detectable in these models, or that lysogenic conversion by BTP1 and BTP5 does not relate to virulence. Indeed, investigation into the function of two highly expressed accessory elements of the BTP1 prophage, the *ST313-td* gene and the STnc6030 transcript both demonstrated immunity functions. *ST313-td* was convincingly shown to mediate immunity to the related lambdoid phage P22. Could the immunity function conferred by *ST313-td* explain the strong association between prophage BTP1 and the African *S.* Typhimurium ST313 lineages? Though the ecology of phages in the natural environment is not well studied, it is logical to assume that lambdoid prophages are highly abundant due to the frequency at which they are isolated and studied. Perhaps the particular immunity conferred by the *ST313-td* gene provides a significant selective advantage in the sub-Saharan African environment. Certainly, if the ST313-td protein is proven to confer immunity to a broad range of different phages, it would lend support to this hypothesis. Alternatively, the physiological phenotype of the *ST313-td* mediated immunity may

provide a selective advantage in a particular niche, which may or may not relate directly to infection of a human host.

Some caution has been taken over the interpretation of the immunity phenotype of the STnc6030 asRNA. Though the heterologous expression of the RNA was shown to mediate complete immunity to BTP1 infection, it remains possible that these results are experimental artefacts. This is because the exact coordinates of the asRNA were predicted based on RNA-seq data, though the size of the largest transcript detected by Northern blot (~500 nt) is considerably shorter than the RNA predicted by RNA-seq (789 nt). Consequently, the asRNA expressed from the $pP_L$-STnc6030 construct is likely an extended version of the STnc6030 asRNA, and could be considered synthetic. In light of this, the possibility cannot be ruled out that the immunity phenotype conferred by the synthetic STnc6030 expression vector is simply an artefact of synthetic antisense silencing of the complementary mRNA sequence. An important aspect of further study will be the experimental identification of the 5' and 3' ends of the STnc6030 RNA transcript(s) by 5' rapid amplification of cDNA ends (5' RACE). Once the extent of the STnc6030 has been defined, the immunity phenotype of the real STnc6030 RNA must be retested. However, evidence that STnc6030 performs a genuine biological role was provided by the sequencing of escape phage mutants that could overcome infection interference by STnc6030; if the immunity phenotype was simply a synthetic antisense-silencing effect, mutations that destabilise the base pairing interaction would not necessary be localised to a particular part of the target region. In fact, all escape phage sequenced possessed mutations within a 36 bp region. The function of the STnc6030 RNA will be a topic for further study.

The fact that the *ST313-td* gene, and perhaps the STnc6030 asRNA, mediate important phenotypes in terms of ST313 biology means that the BTP1 prophage does mediate lysogenic conversion of African *S.* Typhimurium ST313.

# Chapter 7

## General discussion

## 7.1 The context of this thesis

*Salmonella enterica* is an important pathogen that causes disease across the world. The diseases caused by *Salmonella enterica* isolates range from mild, self-limiting gastroenteritis, to fatal systemic infection. Typically in humans, 'typhoidal' *S. enterica* serovars including *S.* Typhi and *S.* Paratyphi A are associated with systemic disease, whilst 'non-typhoidal' serovars such as *S.* Typhimurium and *S.* Enteritidis are associated with gastrointestinal infections. However, in recent years this clinical distinction has weakened, as it has become increasingly appreciated that there are clades of traditionally non-typhoidal *S. enterica* serovars which are specifically associated with systemic disease in infant, malnourished or immunocompromised hosts (Feasey et al., 2016; Kingsley et al., 2009). The syndrome of systemic infection caused by non-typhoidal *S. enterica* serovars is known as invasive non-typhoidal *Salmonella* (iNTS) disease (Feasey et al., 2012). Whilst it was appreciated early on that iNTS-associated clades of *S. enterica* are strongly associated with multi-drug resistance, the genetic determinants mediating the enhanced propensity to cause systemic infection compared to closely-related non-typhoidal clades are unclear. This study set out to investigate whether or not novel prophages associated with iNTS-causing clades of *S.* Typhimurium (ST313) effect the biology of the host in a way that could contribute to the niche adaption, whether through modulating virulence or other phenotypes.

## 7.2 Summary of findings

In Chapter 3, I showed that the African ST313 representative strain D23580 contains 5 full length prophages. Prophage BTP1 and BTP5 are undescribed, novel prophages specific to African ST313 strains, whilst the remainder; Gifsy-2, ST64B and Gifsy-1 have been previously described and are found in other strains of *S.* Typhimurium. Of the five prophages, only BTP1 and BTP5 showed evidence for functional phage production, and mutations inactivating Gifsy-2, ST64B and Gifsy-1 were identified. The BTP1 prophage spontaneously induced at a high rate, estimated to cause the phage-mediated lysis of approximately 0.2% of the lysogenic population. A GFP reporter system was developed to visualise the spontaneous induction of the BTP1 prophage at the single-cell level. Though the BTP5 phage could not be studied using traditional plaque assay methodology, there was evidence that the BTP5 prophage was capable of forming viable BTP5 phage that could lysogenise naïve hosts.

In Chapter 4, I analysed the genomes of *S.* Typhimurium ST313 genomes from the UK, and found that in fact, ST313 in the UK represents a distinct population of

antibiotic susceptible strains associated with gastrointestinal infection. The discovery of the UK-ST313 population revolutionised our understanding of the evolution of the two African ST313 lineages, because UK-strains shared more recent common ancestry to each of the two African lineages than the African lineages share with each other. Additionally the UK-ST313 strains lacked the BTP1 and BTP5 prophages. Together these results indicated that the BTP1 and BTP5 prophages were acquired independently by the two African ST313 lineages, showing convergent evolution to acquire and conserve the BTP1 and BTP5 prophages.

In light of this, in the remaining two chapters I explored the potential functional contribution of the BTP1 and BTP5 prophages that could drive their conservation in African ST313. In Chapter 5, I presented evidence that the prophages, in particular BTP5, affect the global gene expression of representative strain D23580 based on the results of a transcriptomic 'target hunt' experiment. Finally, in Chapter 6, I showed that the *ST313-td* gene of BTP1 mediates lysogenic conversion by functioning as a superinfection immunity factor against infection by *Salmonella* phage P22.

An overall model illustrating the prophage evolution and functional contribution to ST313 lineages described in this thesis is summarised in Figure 7.1.

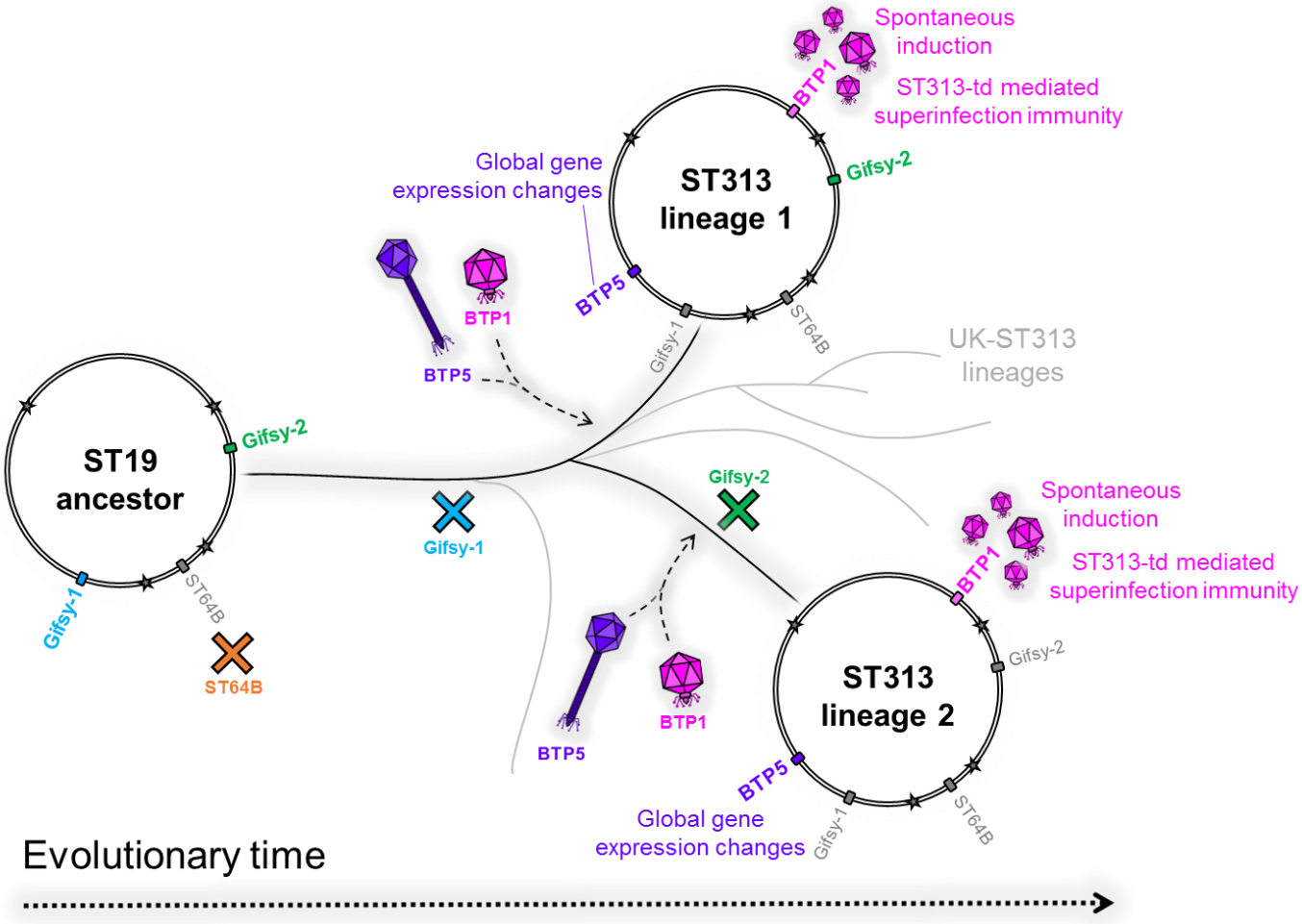



**Figure 7.1 Model summarising the evolution of the *S.* Typhimurium ST313 prophage repertoire and its putative functional contribution.**

## 7.3 Implications of this work

My findings highlight the importance of characterising prophages beyond the genomic level. Putative prophages identified in genome sequencing studies are frequently disregarded, despite the overwhelming evidence that prophages contribute greatly to bacterial biology and pathogenesis. Characterising novel prophages is challenging, particularly in bacterial systems that are not genetically tractable or where no related phages or prophages have been characterised. Studying the prophages of D23580 at the functional level revealed the surprising discovery that the BTP1 prophage yields the highest spontaneously induced phage-titer so far described in the literature for any prophage.

Though spontaneous prophage induction is not well studied and the novelty of this finding is not entirely clear, the spontaneously induced phage titer produced by the BTP1 lysogen is surprising given that a predicted 0.2% of the cellular population are lysed as a result. The transcriptomic data presented in Chapter 5 suggest that BTP1 spontaneous induction occurs in all growth conditions, including intracellular growth inside murine macrophages. Spontaneous cell lysis during infection of mammalian hosts could hypothetically lead to increased LPS (endotoxin) release and therefore an enhanced inflammatory response (Zivot and Hoffman, 1995). Whether or not BTP1-mediated spontaneous induction and cell lysis interacts with the human immune response to infection by *S.* Typhimurium ST313 is beyond the scope of this study, but remains an important topic for future investigation. Indeed, *Salmonella* are known to exploit gut inflammation by utilising tetrathionate produced by oxidation of thiosulphate in the presence of reactive oxygen species during infection of the gut (ROS) (Chirullo et al., 2015; Winter et al., 2010), so a heightened inflammatory response could plausibly modulate *Salmonella* pathogenesis. How exactly the host pathologies associated with iNTS such as AIDS, malnutrition and malaria coinfection may affect the host response to BTP1-mediated cell lysis should also be considered.

The mechanism behind the high frequency spontaneous induction of the BTP1 prophage is another question that remains unanswered. As induction of a prophage is generally the result of repressor inactivation or repressor titration (Campbell, 2006), two hypotheses arise. Firstly, the BTP1 CI repressor protein may have a higher turnover rate, and be intrinsically less stable than the repressor of P22 or Lambda. Secondly, the BTP1 CI repressor may be intrinsically less abundant in the cell. Assuming there is activation of the SOS response in a fraction of the cellular population, as has been demonstrated in other bacteria (Nanda et al., 2014), both these eventualities would result in the balance between induction and maintenance

of lysogeny more readily swinging towards the former. Future experiments involving functional exchange of the BTP1 and P22 repressor proteins may provide evidence for the role of the repressor protein in BTP1 spontaneous induction.

There are a number of examples in the literature where spontaneous prophage induction has been demonstrated to directly contribute to the disease potential of a bacterial pathogen. Most notably, in the case of Shiga toxin producing *E. coli* (STEC). The Shiga toxin, Stx, is carried by a number of different lambdoid prophages which, upon infection of susceptible *E. coli*, lysogenically convert the strain to cause fatal renal pathology due to the synthesis of the Stx toxin (Allison, 2007). However, no secretion system has been identified for the Stx toxin, and therefore synthesis and release of Stx toxin (encoded by a gene located in the prophage early lytic gene operon) depends entirely on prophage induction and cell lysis (Livny and Friedman, 2004). As spontaneous prophage induction in essential for the function of the Stx toxin, it is likely that a heightened sensitivity of the Stx-prophage to induction signals would significantly affect STEC infection and disease progression. Increased spontaneous induction has been observed in Stx-phage lysogens compared to lysogens of non-Stx encoding prophages, suggesting that increased prophage induction is a phenotype that may be selected for when phage-mediated lysis is necessary for release of a cellular product (Livny and Friedman, 2004). However, no phenotype for the BTP1 lysogen has yet been detected that would be consistent with a role for spontaneous cell lysis as a release of a cellular product.

Furthermore, there are numerous examples of prophages which contribute to virulence in animal models without a known mechanism, highlighting the challenge of identifying prophage virulence factors. For example, the SPC-P1 prophage is required for virulence of *Salmonella* Paratyphi C, as judged by the attenuation of a strain lacking the prophage, but no virulence factors have been identified (Zou et al., 2010). Similarly, the *Salmonella* Enteritidis prophage element ΦSE12 is required for virulence in mice though the mechanism by which the prophage promotes virulence is unknown (Araya et al., 2010). Clearly, further functional studies are required to understand the relationship between temperate phage and bacterial pathogens.

Overall, whilst this study clearly showed the strong association between novel prophages BTP1 and BTP5 and the African *S.* Typhimurium ST313 lineages, no evidence was found that the prophages contribute to pathogenicity. Furthermore the implication of the functional inactivation of the three common *S.* Typhimurium prophages Gifsy-2, ST64B and Gifsy-1 in ST313 isolates remains unclear. On the

other hand, I have presented clear evidence that prophage BTP1 mediates a superinfection immunity function against other phages via the function of the ST313-td potein, as well as data that suggests both prophages BTP1 and BTP5 modulate global gene expression of the lysogen. These results suggest that the role of prophages in the evolution of bacterial pathogens is diverse and complex, and may not be as simple as the addition of a single virulence factor.

In conclusion, this thorough exploration of the prophage repertoire of *S.* Typhimurium ST313 poses interesting questions about the potential fitness costs and benefits of novel prophages in epidemic S. Typhimurium ST313. The phenotypic contributions of the prophages represents novel biology in this clinically important sequence type, and it remains to be seen how prophages modulate the behaviour of the pathogen in terms of ecological niche, host range or invasiveness in humans.

# Bibliography


Abedon, S.T., LeJeune, J.T., 2007. Why Bacteriophage Encode Exotoxins and other Virulence Factors. Evol. Bioinforma. Online 1, 97–110.

Achtman, M., Wain, J., Weill, F.X., Nair, S., Zhou, Z., Sangal, V., Krauland, M.G., Hale, J.L., Harbottle, H., Uesbeck, A., Dougan, G., Harrison, L.H., Brisse, S., Group, S.E.M.S., 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. PLoS Pathog. 8, e1002776. doi:10.1371/journal.ppat.1002776

Ackermann, H.-W., 2009. Phage classification and characterization. Methods Mol. Biol. 501, 127–140. doi:10.1007/978-1-60327-164-6_13

Ackermann, H.-W., 2006. Classification of Bacteriophages, in: The Bacteriophages. Oxford University Press, Oxford.

Adamo, S.A., Webster, J.P., 2013. Neural parasitology: how parasites manipulate host behaviour. J. Exp. Biol. 216, 1–2. doi:10.1242/jeb.082511

Alikhan, N.-F., Petty, N.K., Ben Zakour, N.L., Beatson, S.A., 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12, 402. doi:10.1186/1471-2164-12-402

Allison, H.E., 2007. Stx-phages: drivers and mediators of the evolution of STEC and STEC-like pathogens. Future Microbiol. 2, 165–174. doi:10.2217/17460913.2.2.165

Almeida, F., Seribelli, A.A., da Silva, P., Medeiros, M.I.C., Dos Prazeres Rodrigues, D., Moreira, C.G., Allard, M.W., Falcão, J.P., 2017. Multilocus sequence typing of *Salmonella* Typhimurium reveals the presence of the highly invasive ST313 in Brazil. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 51, 41–44. doi:10.1016/j.meegid.2017.03.009

Andersen, J.L., He, G.-X., Kakarla, P., KC, R., Kumar, S., Lakra, W.S., Mukherjee, M.M., Ranaweera, I., Shrestha, U., Tran, T., Varela, M.F., 2015. Multidrug Efflux Pumps from Enterobacteriaceae, *Vibrio cholerae* and *Staphylococcus aureus* Bacterial Food Pathogens. Int. J. Environ. Res. Public. Health 12, 1487–1547. doi:10.3390/ijerph120201487

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Ao, T.T., Feasey, N.A., Gordon, M.A., Keddy, K.H., Angulo, F.J., Crump, J.A., 2015. Global burden of invasive nontyphoidal *Salmonella* disease, 2010(1). Emerg. Infect. Dis. 21. doi:10.3201/eid2106.140999

Aramli, L.A., Teschke, C.M., 1999. Single amino acid substitutions globally suppress the folding defects of temperature-sensitive folding mutants of phage P22 coat protein. J. Biol. Chem. 274, 22217–22224.

Araya, D.V., Quiroz, T.S., Tobar, H.E., Lizana, R.J., Quezada, C.P., Santiviago, C.A., Riedel, C.A., Kalergis, A.M., Bueno, S.M., 2010. Deletion of a prophage-like element causes attenuation of *Salmonella enterica* serovar Enteritidis and promotes protective immunity. Vaccine 28, 5458–5466. doi:10.1016/j.vaccine.2010.05.073

Ashton, P.M., Nair, S., Peters, T.M., Bale, J.A., Powell, D.G., Painset, A., Tewolde, R., Schaefer, U., Jenkins, C., Dallman, T.J., de Pinna, E.M., Grant, K.A., 2016. Identification of *Salmonella* for public health surveillance using whole genome sequencing. PeerJ 4, e1752. doi:10.7717/peerj.1752

Baggesen, D.L., Sørensen, G., Nielsen, E.M., Wegener, H.C., 2010. Phage typing of *Salmonella* Typhimurium – is it still a useful tool for surveillance and outbreak investigation? Euro Surveill. 15.

Baharoglu, Z., Mazel, D., 2014. SOS, the formidable strategy of bacteria against aggressions. FEMS Microbiol. Rev. 38, 1126–1145. doi:10.1111/1574-6976.12077
Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. J. Comput. Mol. Cell Biol. 19, 455–477. doi:10.1089/cmb.2012.0021
Barksdale, L., Arden, S.B., 1974. Persisting Bacteriophage Infections, Lysogeny, and Phage Conversions. Annu. Rev. Microbiol. 28, 265–300. doi:10.1146/annurev.mi.28.100174.001405
Battistuzzi, F.U., Feijao, A., Hedges, S.B., 2004. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. BMC Evol. Biol. 4, 44. doi:10.1186/1471-2148-4-44
Bäumler, A., Fang, F.C., 2013. Host Specificity of Bacterial Pathogens. Cold Spring Harb. Perspect. Med. 3, a010041. doi:10.1101/cshperspect.a010041
Bhavsar, A.P., Brown, N.F., Stoepel, J., Wiermer, M., Martin, D.D.O., Hsu, K.J., Imami, K., Ross, C.J., Hayden, M.R., Foster, L.J., Li, X., Hieter, P., Finlay, B.B., 2013. The *Salmonella* Type III Effector SspH2 Specifically Exploits the NLR Co-chaperone Activity of SGT1 to Subvert Immunity. PloS Pathog. 9, e1003518. doi:10.1371/journal.ppat.1003518
Blank, K., Hensel, M., Gerlach, R.G., 2011. Rapid and highly efficient method for scarless mutagenesis within the *Salmonella enterica* chromosome. PLoS ONE 6, e15763. doi:10.1371/journal.pone.0015763
Bobay, L.-M., Touchon, M., Rocha, E.P.C., 2014. Pervasive domestication of defective prophages by bacteria. Proc. Natl. Acad. Sci. 111, 12127–12132. doi:10.1073/pnas.1405336111
Bogomolnaya, L.M., Santiviago, C.A., Yang, H.-J., Baumler, A.J., Andrews-Polymenis, H.L., 2008. “Form variation” of the O12 antigen is critical for persistence of *Salmonella* Typhimurium in the murine intestine. Mol. Microbiol. 70, 1105–1119. doi:10.1111/j.1365-2958.2008.06461.x
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. doi:10.1093/bioinformatics/btu170
Bordet, J., 1925. Le problème de l’autolyse microbienne transmissible ou du bactériophage. Ann. Inst. Pasteur 711–763.
Bordet, J., 1923. The Cameron Prize Lecture on Microbic Transmissible Autolysis. Br. Med. J. 1, 175–178.
Bossi, L., Figueroa-Bossi, N., 2005. Prophage Arsenal of *Salmonella enterica* Serovar Typhimurium, in: Phages.
Bossi, L., Fuentes, J.A., Mora, G., Figueroa-Bossi, N., 2003. Prophage contribution to bacterial population dynamics. J. Bacteriol. 185, 6467–6471.
Bronowski, C., Fookes, M.C., Gilderthorp, R., Ashelford, K.E., Harris, S.R., Phiri, A., Hall, N., Gordon, M.A., Wain, J., Hart, C.A., Wigley, P., Thomson, N.R., Winstanley, C., 2013. Genomic Characterisation of Invasive Non-Typhoidal *Salmonella enterica* Subspecies enterica Serovar Bovismorbificans Isolates from Malawi. PLoS Negl. Trop. Dis. 7, e2557. doi:10.1371/journal.pntd.0002557
Brumby, A.M., Lamont, I., Dodd, I.B., Egan, J.B., 1996. Defining the SOS operon of coliphage 186. Virology 219, 105–114.
Brüssow, H., Canchaya, C., Hardt, W.-D., 2004. Phages and the Evolution of Bacterial Pathogens: from Genomic Rearrangements to Lysogenic Conversion. Microbiol. Mol. Biol. Rev. 68, 560–602. doi:10.1128/MMBR.68.3.560-602.2004

Bryksin, A.V., Matsumura, I., 2010. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. BioTechniques 48, 463–465. doi:10.2144/000113418
Butela, K., Lawrence, J.G., 2012. Genetic manipulation of pathogenicity loci in non-Typhimurium *Salmonella*. J. Microbiol. Methods 91, 477–482. doi:10.1016/j.mimet.2012.09.013
Campbell, A., 2006. General Aspects of Lysogeny, in: The Bacteriophages. Oxford University Press, Oxford, pp. 66–73.
Campoy, S., Hervas, A., Busquets, N., Erill, I., Teixido, L., Barbe, J., 2006. Induction of the SOS response by bacteriophage lytic development in *Salmonella enterica*. Virology 351, 360–7. doi:10.1016/j.virol.2006.04.001
Canals, R., Hammarlöf, D.L., Kröger, C., Owen, S.V., Lacharme-Lora, L., Fong, W.Y., Rowe, W.P.M., Hokamp, K., Gordon, M.A., Hinton, J.C.D., 2018. Comparative multi-condition inter-strain transcriptomics identifies unexpected differences between global and African sequence types of *Salmonella* Typhimurium. Prep.
Carden, S., Okoro, C., Dougan, G., Monack, D., 2015. Non-typhoidal *Salmonella* Typhimurium ST313 isolates that cause bacteremia in humans stimulate less inflammasome activation than ST19 isolates associated with gastroenteritis. Pathog. Dis. 73. doi:10.1093/femspd/ftu023
Carden, S.E., Walker, G.T., Honeycutt, J., Lugo, K., Pham, T., Jacobson, A., Bouley, D., Idoyaga, J., Tsolis, R.M., Monack, D., 2017. Pseudogenization of the Secreted Effector Gene sseI Confers Rapid Systemic Dissemination of *S.* Typhimurium ST313 within Migratory Dendritic Cells. Cell Host Microbe 21, 182–194. doi:10.1016/j.chom.2017.01.009
Carrolo, M., Frias, M.J., Pinto, F.R., Melo-Cristino, J., Ramirez, M., 2010. Prophage spontaneous activation promotes DNA release enhancing biofilm formation in *Streptococcus pneumoniae*. PLoS ONE 5, e15678. doi:10.1371/journal.pone.0015678
Carver, T., Berriman, M., Tivey, A., Patel, C., Bohme, U., Barrell, B.G., Parkhill, J., Rajandream, M.-A., 2008. Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. Bioinformatics 24, 2672–2676. doi:10.1093/bioinformatics/btn529
Casjens, S., 2003. Prophages and bacterial genomics: what have we learned so far? Mol. Microbiol. 49, 277–300.
Casjens, S.R., 2011. A Plethora of Putative Phages and Prophages, in: The Lure of Bacterial Genetics. American Society of Microbiology.
Casjens, S.R., Hendrix, R.W., 2015. Bacteriophage lambda: Early pioneer and still relevant. Virology, 60th Anniversary Issue 479–480, 310–330. doi:10.1016/j.virol.2015.02.010
Casjens, S.R., Thuman-Commike, P.A., 2011. Evolution of mosaically related tailed bacteriophage genomes seen through the lens of phage P22 virion assembly. Virology 411, 393–415. doi:10.1016/j.virol.2010.12.046
Cenens, W., Makumi, A., Govers, S.K., Lavigne, R., Aertsen, A., 2015. Viral Transmission Dynamics at Single-Cell Resolution Reveal Transiently Immune Subpopulations Caused by a Carrier State Association. PLoS Genet. 11, e1005770. doi:10.1371/journal.pgen.1005770
Cenens, W., Mebrhatu, M.T., Makumi, A., Ceyssens, P.J., Lavigne, R., Van Houdt, R., Taddei, F., Aertsen, A., 2013. Expression of a novel P22 ORFan gene reveals the phage carrier state in *Salmonella* typhimurium. PLOS Genet. 9, e1003269. doi:10.1371/journal.pgen.1003269
Centers for Disease Control and Prevention (CDC), 2016. National Enteric Disease Surveillance: *Salmonella* Annual Report, 2013. CDC, Atlanta, Georgia: US Department of Health and Human Services.

Chakraborty, S., Gogoi, M., Chakravortty, D., 2014. Lactoylglutathione lyase, a critical enzyme in methylglyoxal detoxification, contributes to survival of *Salmonella* in the nutrient rich environment. Virulence 6, 50–65. doi:10.4161/21505594.2014.983791
Chen, Y., Golding, I., Sawai, S., Guo, L., Cox, E.C., 2005. Population Fitness and the Regulation of *Escherichia coli* Genes by Bacterial Viruses. PLoS Biol. 3. doi:10.1371/journal.pbio.0030229
Chirullo, B., Pesciaroli, M., Drumo, R., Ruggeri, J., Razzuoli, E., Pistoia, C., Petrucci, P., Martinelli, N., Cucco, L., Moscati, L., Amadori, M., Magistrali, C.F., Alborali, G.L., Pasquali, P., 2015. *Salmonella* Typhimurium exploits inflammation to its own advantage in piglets. Front. Microbiol. 6. doi:10.3389/fmicb.2015.00985
Christie, G.E., Haggård-Ljungquist, E., Feiwell, R., Calendar, R., 1986. Regulation of bacteriophage P2 late-gene expression: the ogr gene. Proc. Natl. Acad. Sci. U. S. A. 83, 3238–3242.
Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., Barrell, B.G., 2001. Massive gene decay in the leprosy *bacillus*. Nature 409, 1007–1011. doi:10.1038/35059006
Coombes, B.K., Wickham, M.E., Brown, N.F., Lemire, S., Bossi, L., Hsiao, W.W.L., Brinkman, F.S.L., Finlay, B.B., 2005. Genetic and molecular analysis of GogB, a phage-encoded type III-secreted substrate in *Salmonella enterica* serovar typhimurium with autonomous expression from its associated phage. J. Mol. Biol. 348, 817–830. doi:10.1016/j.jmb.2005.03.024
Crump, J.A., Mintz, E.D., 2010. Global trends in typhoid and paratyphoid fever. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 50, 241–246. doi:10.1086/649541
Cumby, N., Davidson, A.R., Maxwell, K.L., 2012. The moron comes of age. Bacteriophage 2, 225–228. doi:10.4161/bact.23146
Cunha, B.A., 2004. The death of Alexander the Great: malaria or typhoid fever? Infect. Dis. Clin. North Am. 18, 53–63. doi:10.1016/S0891-5520(03)00090-4
d'Hérelle, F., 1922. The bacteriophage, its role in immunity. Baltimore, Williams & Wilkins Company.
d'Hérelle, F., 1917. Sur un microbe invisible antagoniste des bacilles dysentériques. Comptes Rendus Académie Sci. 373–5.
Datsenko, K.A., Wanner, B.L., 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. 97, 6640–5. doi:10.1073/pnas.120163297
Datta, S., Costantino, N., Court, D.L., 2006. A set of recombineering plasmids for gram-negative bacteria. Gene 379, 109–15. doi:10.1016/j.gene.2006.04.018
Davies, M.R., Broadbent, S.E., Harris, S.R., Thomson, N.R., van der Woude, M.W., 2013. Horizontally acquired glycosyltransferase operons drive salmonellae lipopolysaccharide diversity. PLOS Genet. 9, e1003568. doi:10.1371/journal.pgen.1003568
Degnan, P.H., Michalowski, C.B., Babić, A.C., Cordes, M.H.J., Little, J.W., 2007. Conservation and diversity in the immunity regions of wild phages with the immunity specificity of phage λ: Immunity regions of lambda-immunity phages. Mol. Microbiol. 64, 232–244. doi:10.1111/j.1365-2958.2007.05650.x
DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B.,

Altshuler, D., Daly, M.J., 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498. doi:10.1038/ng.806

Desai, P.T., Porwollik, S., Long, F., Cheng, P., Wollam, A., Clifton, S.W., Weinstock, G.M., McClelland, M., 2013. Evolutionary Genomics of *Salmonella enterica* Subspecies. mBio 4, e00579-12. doi:10.1128/mBio.00579-12

Doolittle, R.F., Feng, D.F., Tsang, S., Cho, G., Little, E., 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271, 470–477.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. doi:10.1186/1471-2148-7-214

Ellermeier, C.D., Slauch, J.M., 2006. The Genus *Salmonella*, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes: Volume 6: Proteobacteria: Gamma Subclass. Springer New York, New York, NY, pp. 123–158. doi:10.1007/0-387-30746-X_7

Erill, I., Campoy, S., Barbé, J., 2007. Aeons of distress: an evolutionary perspective on the bacterial SOS response. FEMS Microbiol. Rev. 31, 637–656. doi:10.1111/j.1574-6976.2007.00082.x

Fàbrega, A., Vila, J., 2013. *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. Clin. Microbiol. Rev. 26, 308–341. doi:10.1128/CMR.00066-12

Feasey, N.A., Cain, A.K., Msefula, C.L., Pickard, D., Alaerts, M., Aslett, M., Everett, D.B., Allain, T.J., Dougan, G., Gordon, M.A., Heyderman, R.S., Kingsley, R.A., 2014. Drug Resistance in *Salmonella enterica* ser. Typhimurium Bloodstream Infection, Malawi. Emerg. Infect. Dis. 20, 1957–1959. doi:10.3201/eid2011.141175

Feasey, N.A., Dougan, G., Kingsley, R.A., Heyderman, R.S., Gordon, M.A., 2012. Invasive non-typhoidal *Salmonella* disease: an emerging and neglected tropical disease in Africa. Lancet 379, 2489–99. doi:10.1016/S0140-6736(11)61752-2

Feasey, N.A., Hadfield, J., Keddy, K.H., Dallman, T.J., Jacobs, J., Deng, X., Wigley, P., Barquist, L., Langridge, G.C., Feltwell, T., Harris, S.R., Mather, A.E., Fookes, M., Aslett, M., Msefula, C., Kariuki, S., Maclennan, C.A., Onsare, R.S., Weill, F.-X., Le Hello, S., Smith, A.M., McClelland, M., Desai, P., Parry, C.M., Cheesbrough, J., French, N., Campos, J., Chabalgoity, J.A., Betancor, L., Hopkins, K.L., Nair, S., Humphrey, T.J., Lunguya, O., Cogan, T.A., Tapia, M.D., Sow, S.O., Tennant, S.M., Bornstein, K., Levine, M.M., Lacharme-Lora, L., Everett, D.B., Kingsley, R.A., Parkhill, J., Heyderman, R.S., Dougan, G., Gordon, M.A., Thomson, N.R., 2016. Distinct *Salmonella* Enteritidis lineages associated with enterocolitis in high-income settings and invasive disease in low-income settings. Nat. Genet. 48, 1211–1217. doi:10.1038/ng.3644

Feiner, R., Argov, T., Rabinovich, L., Sigal, N., Borovok, I., Herskovits, A.A., 2015. A new perspective on lysogeny: prophages as active regulatory switches of bacteria. Nat. Rev. Microbiol. 13, 641–650. doi:10.1038/nrmicro3527

Felix, A., Callow, B.R., 1943. Typing Of Paratyphoid B Bacilli By Means Of Vi Bacteriophage. Br. Med. J. 2, 127–130.

Ferreira, R.B.R., Valdez, Y., Coombes, B.K., Sad, S., Gouw, J.W., Brown, E.M., Li, Y., Grassl, G.A., Antunes, L.C.M., Gill, N., Truong, M., Scholz, R., Reynolds, L.A., Krishnan, L., Zafer, A.A., Sal-Man, N., Lowden, M.J., Auweter, S.D., Foster, L.J., Finlay, B.B., 2015. A Highly Effective Component Vaccine against Nontyphoidal *Salmonella enterica* Infections. mBio 6, e01421-01415. doi:10.1128/mBio.01421-15

Figueroa-Bossi, N., Bossi, L., 2004. Resuscitation of a Defective Prophage in *Salmonella* Cocultures. J. Bacteriol. 186, 4038–4041. doi:10.1128/JB.186.12.4038-4041.2004
Figueroa-Bossi, N., Bossi, L., 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Mol. Microbiol. 33, 167–76.
Figueroa-Bossi, N., Uzzau, S., Maloriol, D., Bossi, L., 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. Mol. Microbiol. 39, 260–71.
Fokine, A., Rossmann, M.G., 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage 4. doi:10.4161/bact.28281
Fookes, M., Schroeder, G.N., Langridge, G.C., Blondel, C.J., Mammina, C., Connor, T.R., Seth-Smith, H., Vernikos, G.S., Robinson, K.S., Sanders, M., Petty, N.K., Kingsley, R.A., Bäumler, A.J., Nuccio, S.-P., Contreras, I., Santiviago, C.A., Maskell, D., Barrow, P., Humphrey, T., Nastasi, A., Roberts, M., Frankel, G., Parkhill, J., Dougan, G., Thomson, N.R., 2011. *Salmonella bongori* Provides Insights into the Evolution of the Salmonellae. PLoS Pathog. 7. doi:10.1371/journal.ppat.1002191
Fortier, L.-C., Sekulovic, O., 2013. Importance of prophages to evolution and virulence of bacterial pathogens. Virulence 4, 354–365. doi:10.4161/viru.24498
Freeman, V.J., 1951. Studies on the Virulence of Bacteriophage-Infected Strains of *Corynebacterium Diphtheriae*. J. Bacteriol. 61, 675–688.
Galan, J.E., 2016. Typhoid toxin provides a window into typhoid fever and the biology of *Salmonella* Typhi. Proc. Natl. Acad. Sci. 113, 6338–6344. doi:10.1073/pnas.1606335113
Garrity, G., Brenner, D.J., Krieg, N.R., Staley, J.R., 2007. Bergey's Manual® of Systematic Bacteriology: Volume 2: The Proteobacteria, Part B: The Gammaproteobacteria. Springer US.
Georgiades, K., Raoult, D., 2011. Genomes of the Most Dangerous Epidemic Bacteria Have a Virulence Repertoire Characterized by Fewer Genes but More Toxin-Antitoxin Modules. PLoS ONE 6, e17962. doi:10.1371/journal.pone.0017962
Gibbons, H.S., Kalb, S.R., Cotter, R.J., Raetz, C.R.H., 2005. Role of Mg2+ and pH in the modification of *Salmonella* lipid A after endocytosis by macrophage tumour cells. Mol. Microbiol. 55, 425–440. doi:10.1111/j.1365-2958.2004.04409.x
Gilchrist, J.J., MacLennan, C.A., Hill, A.V.S., 2015. Genetic susceptibility to invasive *Salmonella* disease. Nat. Rev. Immunol. 15, 452–463. doi:10.1038/nri3858
Gil-Cruz, C., Bobat, S., Marshall, J.L., Kingsley, R.A., Ross, E.A., Henderson, I.R., Leyton, D.L., Coughlan, R.E., Khan, M., Jensen, K.T., Buckley, C.D., Dougan, G., MacLennan, I.C.M., Lopez-Macias, C., Cunningham, A.F., 2009. The porin OmpD from nontyphoidal *Salmonella* is a key target for a protective B1b cell antibody response. Proc. Natl. Acad. Sci. 106, 9803–9808. doi:10.1073/pnas.0812431106
Gilks, C.F., Brindle, R.J., Otieno, L.S., Simani, P.M., Newnham, R.S., Bhatt, S.M., Lule, G.N., Okelo, G.B., Watkins, W.M., Waiyaki, P.G., 1990. Life-threatening bacteraemia in HIV-1 seropositive adults admitted to hospital in Nairobi, Kenya. Lancet 336, 545–549.
Gödeke, J., Paul, K., Lassak, J., Thormann, K.M., 2011a. Phage-induced lysis enhances biofilm formation in *Shewanella oneidensis* MR-1. ISME J. 5, 613–626. doi:10.1038/ismej.2010.153
Goering, R.V., 2010. Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 10, 866–875. doi:10.1016/j.meegid.2010.07.023

Goh, Y.S., MacLennan, C.A., 2013. Invasive African nontyphoidal *Salmonella* requires high levels of complement for cell-free antibody-dependent killing. J. Immunol. Methods 387, 121–129. doi:10.1016/j.jim.2012.10.005
Gordon, M.A., 2011. Invasive nontyphoidal *Salmonella* disease: epidemiology, pathogenesis and diagnosis. Curr. Opin. Infect. Dis. 24, 484–489. doi:10.1097/QCO.0b013e32834a9980
Gottesman, S., Storz, G., 2011. Bacterial Small RNA Regulators: Versatile Roles and Rapidly Evolving Variations. Cold Spring Harb. Perspect. Biol. 3, a003798. doi:10.1101/cshperspect.a003798
Grimont, P.A., Weill, F.-X., 2007. Antigenic formulae of the *Salmonella* serovars. WHO Collab. Cent. Ref. Res. *Salmonella* Inst. Pasteur Paris Fr.
Gruenewald, R., Blum, S., Chan, J., 1994. Relationship between human immunodeficiency virus infection and salmonellosis in 20- to 59-year-old residents of New York City. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 18, 358–363.
Guerin, P.J., Grais, R.F., Rottingen, J.A., Valleron, A.J., 2007. Using European travellers as an early alert to detect emerging pathogens in countries with limited laboratory resources. BMC Public Health 7, 8. doi:10.1186/1471-2458-7-8
Günster, R.A., Matthews, S.A., Holden, D.W., Thurston, T.L.M., 2017. SseK1 and SseK3 Type III Secretion System Effectors Inhibit NF-κB Signaling and Necroptotic Cell Death in *Salmonella*-Infected Macrophages. Infect. Immun. 85. doi:10.1128/IAI.00010-17
Hammarlöf, D.L., Kröger, C., Owen, S.V., Canals, R., Lacharme-Lora, L., Wenner, N., Wells, T.J., Henderson, I.J., Wigley, P., Hokamp, K., Feasey, N.A., Gordon, M.A., Hinton, J.C.D., 2017. The role of a single non-coding nucleotide in the evolution of an epidemic African clade of *Salmonella*. bioRxiv 175265. doi:10.1101/175265
Han, Y., Liu, L., Fang, N., Yang, R., Zhou, D., 2013. Regulation of pathogenicity by noncoding RNAs in bacteria. Future Microbiol. 8, 579–591. doi:10.2217/fmb.13.20
Hatfull, G.F., 2008. Bacteriophage Genomics. Curr. Opin. Microbiol. 11, 447–453. doi:10.1016/j.mib.2008.09.004
Hautefort, I., Proenca, M.J., Hinton, J.C.D., 2003. Single-Copy Green Fluorescent Protein Gene Fusions Allow Accurate Measurement of *Salmonella* Gene Expression In Vitro and during Infection of Mammalian Cells. Appl. Environ. Microbiol. 69, 7480–7491. doi:10.1128/AEM.69.12.7480-7491.2003
Hébrard, M., Kröger, C., Srikumar, S., Colgan, A., Händler, K., Hinton, J.C.D., 2012. sRNAs and the virulence of *Salmonella enterica* serovar Typhimurium. RNA Biol. 9, 437–445. doi:10.4161/rna.20480
Hendrix, R., Roberts, J., Stahl, F., Weisberg, R., 1983. Lambda II.
Herrero-Fresno, A., Wallrodt, I., Leekitcharoenphon, P., Olsen, J.E., Aarestrup, F.M., Hendriksen, R.S., 2014. The Role of the st313-td Gene in Virulence of *Salmonella* Typhimurium ST313. PLoS ONE 9. doi:10.1371/journal.pone.0084566
Hershey, A.D., Chase, M., 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. J. Gen. Physiol. 36, 39–56.
Hiley, L., Fang, N.-X., Micalizzi, G.R., Bates, J., 2014. Distribution of Gifsy-3 and of Variants of ST64B and Gifsy-1 Prophages amongst *Salmonella enterica* Serovar Typhimurium Isolates: Evidence that Combinations of Prophages Promote Clonality. PLoS ONE 9, e86203. doi:10.1371/journal.pone.0086203
Ho, T.D., Figueroa-Bossi, N., Wang, M., Uzzau, S., Bossi, L., Slauch, J.M., 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella enterica* serovar Typhimurium. J. Bacteriol. 184, 5234–9.

Holmes, R.K., 2000. Biology and Molecular Epidemiology of Diphtheria Toxin and the *tox* Gene. J. Infect. Dis. 181, S156–S167. doi:10.1086/315554
Hughes, L.A., Wigley, P., Bennett, M., Chantrey, J., Williams, N., 2010. Multi-locus sequence typing of *Salmonella enterica* serovar Typhimurium isolates from wild birds in northern England suggests host-adapted strain. Lett. Appl. Microbiol. 51, 477–479. doi:10.1111/j.1472-765X.2010.02918.x
Hyman, P., Abedon, S.T., 2009. Practical methods for determining phage growth parameters. Methods Mol. Biol. Clifton NJ 501, 175–202. doi:10.1007/978-1-60327-164-6_18
Inouye, M., Conway, T.C., Zobel, J., Holt, K.E., 2012. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics 13, 338. doi:10.1186/1471-2164-13-338
Iwashita, S., Kanegasaki, S., 1973. Smooth specific phage adsorption: endorhamnosidase activity of tail parts of P22. Biochem. Biophys. Res. Commun. 55, 403–409.
Jones, T.F., Ingram, L.A., Cieslak, P.R., Vugia, D.J., Tobin-D'Angelo, M., Hurd, S., Medus, C., Cronquist, A., Angulo, F.J., 2008. Salmonellosis outcomes differ substantially by serotype. J. Infect. Dis. 198, 109–114. doi:10.1086/588823
Josenhans, C., Suerbaum, S., 2002. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. 291, 605–614. doi:10.1078/1438-4221-00173
Joshi, N.A., Fass J. N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files.
Kalionis, B., Pritchard, M., Egan, J.B., 1986. Control of gene expression in the P2-related temperate coliphages. IV. Concerning the late control gene and control of its transcription. J. Mol. Biol. 191, 211–220.
Kariuki, S., Okoro, C., Kiiru, J., Njoroge, S., Omuse, G., Langridge, G., Kingsley, R.A., Dougan, G., Revathi, G., 2015. Ceftriaxone-resistant *Salmonella enterica* serotype typhimurium sequence type 313 from Kenyan patients is associated with the blaCTX-M-15 gene on a novel IncHI2 plasmid. Antimicrob. Agents Chemother. 59, 3133–3139. doi:10.1128/AAC.00078-15
Kauffmann., F., 1940. Zur Serologie des I-Antigens in der *Salmonella*-Group. Acta Pathol. Microbiol. Scand. 17, 135–144. doi:10.1111/j.1699-0463.1940.tb01473.x
Kingsley, R.A, Msefula, C.L., Thomson, N.R., Kariuki, S., Holt, K.E., Gordon, M. a, Harris, D., Clarke, L., Whitehead, S., Sangal, V., Marsh, K., Achtman, M., Molyneux, M.E., Cormican, M., Parkhill, J., MacLennan, C. a, Heyderman, R.S., Dougan, G., 2009. Epidemic multiple drug resistant *Salmonella* Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. Genome Res. 19, 2279–87. doi:10.1101/gr.091017.109
Kingsley, R.A., Kay, S., Connor, T., Barquist, L., Sait, L., Holt, K.E., Sivaraman, K., Wileman, T., Goulding, D., Clare, S., Hale, C., Seshasayee, A., Harris, S., Thomson, N.R., Gardner, P., Rabsch, W., Wigley, P., Humphrey, T., Parkhill, J., Dougan, G., 2013. Genome and Transcriptome Adaptation Accompanying Emergence of the Definitive Type 2 Host-Restricted *Salmonella enterica* Serovar Typhimurium Pathovar. mBio 4, e00565-13-e00565-13. doi:10.1128/mBio.00565-13
Kintz, E., Davies, M.R., Hammarlöf, D.L., Canals, R., Hinton, J.C.D., van der Woude, M.W., 2015. A BTP1 prophage gene present in invasive non-typhoidal *Salmonella* determines composition and length of the O-antigen of the lipopolysaccharide. Mol. Microbiol. n/a-n/a. doi:10.1111/mmi.12933
Kintz, E., Heiss, C., Black, I., Donohue, N., Brown, N., Davies, M.R., Azadi, P., Baker, S., Kaye, P.M., van der Woude, M., 2017. *Salmonella enterica* Serovar Typhi Lipopolysaccharide O-Antigen Modification Impact on Serum Resistance and Antibody Recognition. Infect. Immun. 85. doi:10.1128/IAI.01021-16

Koskiniemi, S., Pranting, M., Gullberg, E., Nasvall, J., Andersson, D.I., 2011. Activation of cryptic aminoglycoside resistance in *Salmonella enterica*. Mol. Microbiol. 80, 1464–1478. doi:10.1111/j.1365-2958.2011.07657.x

Kreuzer, K.N., 2013. DNA Damage Responses in Prokaryotes: Regulating Gene Expression, Modulating Growth Patterns, and Manipulating Replication Forks. Cold Spring Harb. Perspect. Biol. 5, a012674. doi:10.1101/cshperspect.a012674

Krinke, L., Wulff, D.L., 1987. OOP RNA, produced from multicopy plasmids, inhibits lambda cII gene expression through an RNase III-dependent mechanism. Genes Dev. 1, 1005–1013.

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S.K., Hammarlöf, D.L., Canals, R., Grissom, J.E., Conway, T., Hokamp, K., Hinton, J.C.D., 2013. An Infection-Relevant Transcriptomic Compendium for *Salmonella enterica* Serovar Typhimurium. Cell Host Microbe 14, 683–695. doi:10.1016/j.chom.2013.11.010

Kröger, C., Dillon, S.C., Cameron, A.D., Papenfort, K., Sivasankaran, S.K., Hokamp, K., Chao, Y., Sittka, A., Hébrard, M., Händler, K., Colgan, A., Leekitcharoenphon, P., Langridge, G.C., Lohan, A.J., Loftus, B., Lucchini, S., Ussery, D.W., Dorman, C.J., Thomson, N.R., Vogel, J., Hinton, J.C., 2012. The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. Proc. Natl. Acad. Sci. 109, E1277-86. doi:10.1073/pnas.1201061109

Kropinski, A.M., Sulakvelidze, A., Konczy, P., Poppe, C., 2007. *Salmonella* phages and prophages--genomics and practical aspects. Methods Mol. Biol. 394, 133–75. doi:10.1007/978-1-59745-512-1_9

Laible, M., Boonrod, K., 2009. Homemade site directed mutagenesis of whole plasmids. United States.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. doi:10.1038/nmeth.1923

Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K., Ritchie, M.E., 2016. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Research 5, 1408. doi:10.12688/f1000research.9005.1

Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29. doi:10.1186/gb-2014-15-2-r29

Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R., Monack, D.M., 2006. Genome-Wide Screen for *Salmonella* Genes Required for Long-Term Systemic Infection of the Mouse. PLOS Pathog. 2, e11. doi:10.1371/journal.ppat.0020011

Lawrence, J.G., Hendrix, R.W., Casjens, S., 2001. Where are the pseudogenes in bacterial genomes? Trends Microbiol. 9, 535–540.

Lemire, S., Figueroa-Bossi, N., Bossi, L., 2011. Bacteriophage crosstalk: coordination of prophage induction by trans-acting antirepressors. PLOS Genet. 7, e1002149. doi:10.1371/journal.pgen.1002149

Lemire, S., Figueroa-Bossi, N., Bossi, L., 2008. Prophage Contribution to *Salmonella* Virulence and Diversity, in: Hensel, M., Schmidt, H. (Eds.), Horizontal Gene Transfer in the Evolution of Pathogenesis: Cambridge University Press, pp. 159–192.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinforma. Oxf. Engl. 30, 923–930. doi:10.1093/bioinformatics/btt656
Lilleengen, K., 1948. Typing of *Salmonella* Typhimurium by means of bacteriophage. Acta Pathol. Microbiol. Scand. 25, 1–39.
Lin, L., Bitner, R., Edlin, G., 1977. Increased reproductive fitness of *Escherichia coli* lambda lysogens. J. Virol. 21, 554–559.
Lisbonne, M., Carrére, L., 1922. Antagonisme microbien et lyse transmissible du Bacille de Shiga. Comptes Rendus Seances Soc. Biol. 569–570.
Liu, X., Jiang, H., Gu, Z., Roberts, J.W., 2013. High-resolution view of bacteriophage lambda gene expression by ribosome profiling. Proc. Natl. Acad. Sci. 110, 11928–11933. doi:10.1073/pnas.1309739110
Livny, J., Friedman, D.I., 2004. Characterizing spontaneous induction of Stx encoding phages using a selectable reporter system. Mol. Microbiol. 51, 1691–1704. doi:10.1111/j.1365-2958.2003.03934.x
Llagostera, M., Barbé, J., Guerrero, R., 1986. Characterization of SE1, a New General Transducing Phage of *Salmonella* typhimurium. Microbiology 132, 1035–1041. doi:10.1099/00221287-132-4-1035
Lokken, K.L., Walker, G.T., Tsolis, R.M., 2016. Disseminated infections with antibiotic-resistant non-typhoidal *Salmonella* strains: contributions of host and pathogen factors. Pathog. Dis. 74. doi:10.1093/femspd/ftw103
Luco, R.F., 2013. The non-coding genome: a universe in expansion for fine-tuning the coding world. Genome Biol. 14, 314. doi:10.1186/gb4140
Lwoff, A., 1953. Lysogeny. Bacteriol. Rev. 17, 269–337.
Lwoff, A., Gutmann, A., 1950. Recherches sur un Bacillus megaterium lysogene. Ann. Inst. Pasteur 711–739.
Ma, M., Eaton, J.W., 1992. Multicellular oxidant defense in unicellular organisms. Proc. Natl. Acad. Sci. U. S. A. 89, 7924–7928.
MacLennan, C.A., Levine, M.M., 2013. Invasive nontyphoidal *Salmonella* disease in Africa: current status. Expert Rev. Anti Infect. Ther. 11, 443–446. doi:10.1586/eri.13.27
Maiden, M.C.J., van Rensburg, M.J.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Micro 11, 728–736.
Majowicz, S.E., Musto, J., Scallan, E., Angulo, F.J., Kirk, M., O'Brien, S.J., Jones, T.F., Fazil, A., Hoekstra, R.M., 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am. 50, 882–889. doi:10.1086/650733
Malorny, B., Hauser, E., Dieckmann, R., 2011. New Approaches in Subspecies-level *Salmonella* Classification, in: *Salmonella*: From Genome to Function. Caister Academic Press, Norfolk, UK, pp. 1–25.
Maloy, S.R., 1990. Experimental techniques in bacterial genetics. Jones and Bartlett, Boston.
Martínez-García, E., de Lorenzo, V., 2011. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. Environ. Microbiol. 13, 2702–2716. doi:10.1111/j.1462-2920.2011.02538.x
Mather, A.E., Lawson, B., de Pinna, E., Wigley, P., Parkhill, J., Thomson, N.R., Page, A.J., Holmes, M.A., Paterson, G.K., 2016. Genomic Analysis of *Salmonella enterica* Serovar Typhimurium from Wild Passerines in England and Wales. Appl. Environ. Microbiol. 82, 6728–6735. doi:10.1128/AEM.01660-16
McClelland, M., Sanderson, K.E., Clifton, S.W., Latreille, P., Porwollik, S., Sabo, A., Meyer, R., Bieri, T., Ozersky, P., McLellan, M., Harkins, C.R., 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., Wilson, R.K., 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. Nat. Genet. 36, 1268–1274. doi:10.1038/ng1470

McNally, A., Thomson, N.R., Reuter, S., Wren, B.W., 2016. "Add, stir and reduce": *Yersinia* spp. as model bacteria for pathogen evolution. Nat. Rev. Microbiol. 14, 177–190. doi:10.1038/nrmicro.2015.29

Mead, P.S., Slutsker, L., Dietz, V., McCaig, L.F., Bresee, J.S., Shapiro, C., Griffin, P.M., Tauxe, R. V., 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. 5, 607–625. doi:10.3201/eid0506.990625

Michel, B., 2005. After 30 Years of Study, the Bacterial SOS Response Still Surprises Us. PLOS Biol. 3, e255. doi:10.1371/journal.pbio.0030255

Mirold, S., Rabsch, W., Rohde, M., Stender, S., Tschape, H., Russmann, H., Igwe, E., Hardt, W.D., 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella* typhimurium strain. Proc. Natl. Acad. Sci. 96, 9845–9850.

Moreno Switt, A.I., den Bakker, H.C., Cummings, C.A., Rodriguez-Rivera, L.D., Govoni, G., Raneiri, M.L., Degoricija, L., Brown, S., Hoelzer, K., Peters, J.E., Bolchacova, E., Furtado, M.R., Wiedmann, M., 2012. Identification and Characterization of Novel *Salmonella* Mobile Elements Involved in the Dissemination of Genes Linked to Virulence and Transmission. PLoS ONE 7, e41247. doi:10.1371/journal.pone.0041247

Muir, P., Li, S., Lou, S., Wang, D., Spakowicz, D.J., Salichos, L., Zhang, J., Weinstock, G.M., Isaacs, F., Rozowsky, J., Gerstein, M., 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biol. 17, 53. doi:10.1186/s13059-016-0917-0

Murray, N.E., 2006. The impact of phage lambda: from restriction to recombineering. Biochem. Soc. Trans. 34, 203–207. doi:10.1042/BST20060203

Nanda, A.M., Heyer, A., Krämer, C., Grünberger, A., Kohlheyer, D., Frunzke, J., 2014. Analysis of SOS-Induced Spontaneous Prophage Induction in *Corynebacterium glutamicum* at the Single-Cell Level. J. Bacteriol. 196, 180–188. doi:10.1128/JB.01018-13

Nanda, A.M., Thormann, K., Frunzke, J., 2015. Impact of Spontaneous Prophage Induction on the Fitness of Bacterial Populations and Host-Microbe Interactions. J. Bacteriol. 197, 410–419. doi:10.1128/JB.02230-14

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. doi:10.1093/molbev/msu300

Nicol, J.W., Helt, G.A., Blanchard, S.G., Raja, A., Loraine, A.E., 2009. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinforma. Oxf. Engl. 25, 2730–2731. doi:10.1093/bioinformatics/btp472

Nilsson, A.S., Haggård-Ljungquist, E., 2006. The P2-like Bacteriophages, in: The Bacteriophages. Oxford University Press, New York, NY.

Nuccio, S.-P., Bäumler, A.J., 2014. Comparative Analysis of *Salmonella* Genomes Identifies a Metabolic Network for Escalating Growth in the Inflamed Gut. mBio 5, e00929-14. doi:10.1128/mBio.00929-14

Okoro, C.K., Barquist, L., Connor, T.R., Harris, S.R., Clare, S., Stevens, M.P., Arends, M.J., Hale, C., Kane, L., Pickard, D.J., Hill, J., Harcourt, K., Parkhill, J., Dougan, G., Kingsley, R.A., 2015. Signatures of adaptation in human invasive *Salmonella* Typhimurium ST313 populations from sub-Saharan Africa. PLoS Negl. Trop. Dis. 9, e0003611. doi:10.1371/journal.pntd.0003611

Okoro, C.K., Kingsley, R. a, Connor, T.R., Harris, S.R., Parry, C.M., Al-Mashhadani, M.N., Kariuki, S., Msefula, C.L., Gordon, M. a, de Pinna, E., Wain, J., Heyderman, R.S., Obaro, S., Alonso, P.L., Mandomando, I., Maclennan, C. a, Tapia, M.D., Levine, M.M., Tennant, S.M., Parkhill, J., Dougan, G., 2012a. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. Nat. Genet. 1–9. doi:10.1038/ng.2423

Okoro, C.K., Kingsley, R.A., Connor, T.R., Harris, S.R., Parry, C.M., Al-Mashhadani, M.N., Kariuki, S., Msefula, C.L., Gordon, M.A., de Pinna, E., Wain, J., Heyderman, R.S., Obaro, S., Alonso, P.L., Mandomando, I., MacLennan, C.A., Tapia, M.D., Levine, M.M., Tennant, S.M., Parkhill, J., Dougan, G., 2012b. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. Nat. Genet. 44, 1215–1221. doi:10.1038/ng.2423

Osterhout, R.E., Figueroa, I.A., Keasling, J.D., Arkin, A.P., 2007. Global analysis of host response to induction of a latent bacteriophage. BMC Microbiol. 7, 82. doi:10.1186/1471-2180-7-82

Ottemann, K.M., Miller, J.F., 1997. Roles for motility in bacterial-host interactions. Mol. Microbiol. 24, 1109–1117.

Owen, S.V., Wenner, N., Canals, R., Makumi, A., Hammarlöf, D.L., Gordon, M.A., Aertsen, A., Feasey, N.A., Hinton, J.C.D., 2017. Characterization of the Prophage Repertoire of African *Salmonella* Typhimurium ST313 Reveals High Levels of Spontaneous Induction of Novel Phage BTP1. Front. Microbiol. 8, 235. doi:10.3389/fmicb.2017.00235

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., 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. doi:10.1093/nar/gkn050

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

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T.G., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C.F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527. doi:10.1038/35097083

Parma, D.H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E., Gold, L., 1992. The Rex system of bacteriophage lambda: tolerance and altruistic cell death. Genes Dev. 6, 497–510.

Parsons, B.N., Humphrey, S., Salisbury, A.M., Mikoleit, J., Hinton, J.C.D., Gordon, M.A., Wigley, P., 2013. Invasive Non-Typhoidal *Salmonella* Typhimurium ST313 Are Not Host-Restricted and Have an Invasive Phenotype in Experimentally Infected Chickens. PLoS Negl. Trop. Dis. 7, e2487. doi:10.1371/journal.pntd.0002487

Pires, S.M., Vieira, A.R., Hald, T., Cole, D., 2014. Source attribution of human salmonellosis: an overview of methods and estimates. Foodborne Pathog. Dis. 11, 667–676. doi:10.1089/fpd.2014.1744

Powell, D.R., 2017. Degust: Visualize, explore and appreciate RNA-seq differential gene-expression data. Victorian Bioinformatics Consortium, Monash University.

Ptashne, M., 2004. A genetic switch : phage lambda revisited. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Rabinovich, L., Sigal, N., Borovok, I., Nir-Paz, R., Herskovits, A.A., 2012. Prophage excision activates Listeria competence genes that promote phagosomal escape and virulence. Cell 150, 792–802. doi:10.1016/j.cell.2012.06.036
Rabsch, W., 2007. *Salmonella* typhimurium phage typing for pathogens. Methods Mol. Biol. Clifton NJ 394, 177–211. doi:10.1007/978-1-59745-512-1_10
Rabsch, W., Truepschuch, S., Windhorst, D., Gerlach, R.G., 2011. Typing Phage and Prophages of *Salmonella*, in: *Salmonella*: From Gemome to Function. Caister Academic Press, Norfolk, UK, pp. 25–48.
Ramachandran, G., Aheto, K., Shirtliff, M.E., Tennant, S.M., 2016. Poor biofilm-forming ability and long-term survival of invasive *Salmonella* Typhimurium ST313. Pathog. Dis. 74. doi:10.1093/femspd/ftw049
Ramachandran, G., Panda, A., Higginson, E.E., Ateh, E., Lipsky, M.M., Sen, S., Matson, C.A., Permala-Booth, J., DeTolla, L.J., Tennant, S.M., 2017. Virulence of invasive *Salmonella* Typhimurium ST313 in animal models of infection. PLoS Negl. Trop. Dis. 11, e0005697. doi:10.1371/journal.pntd.0005697
Ramachandran, G., Perkins, D.J., Schmidlein, P.J., Tulapurkar, M.E., Tennant, S.M., 2015. Invasive *Salmonella* Typhimurium ST313 with Naturally Attenuated Flagellin Elicits Reduced Inflammation and Replicates within Macrophages. PLoS Negl. Trop. Dis. 9. doi:10.1371/journal.pntd.0003394
Ranade, K., Poteete, A.R., 1993. A switch in translation mediated by an antisense RNA. Genes Dev. 7, 1498–1507.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47–e47. doi:10.1093/nar/gkv007
Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25. doi:10.1186/gb-2010-11-3-r25
Ruska, H., 1940. Die Sichtbarmachung der bakteriophagen Lyse im Übermikroskop. Naturwissenschaften 28, 45–46. doi:10.1007/BF01486931
Salisbury, A.-M., Bronowski, C., Wigley, P., 2011. *Salmonella* Virchow isolates from human and avian origins in England--molecular characterization and infection of epithelial cells and poultry. J. Appl. Microbiol. 111, 1505–1514. doi:10.1111/j.1365-2672.2011.05152.x
Sasikaran, J., Ziemski, M., Zadora, P.K., Fleig, A., Berg, I.A., 2014. Bacterial itaconate degradation promotes pathogenicity. Nat. Chem. Biol. 10, 371–377. doi:10.1038/nchembio.1482
Sassanfar, M., Roberts, J.W., 1990. Nature of the SOS-inducing signal in *Escherichia coli*. The involvement of DNA replication. J. Mol. Biol. 212, 79–96. doi:10.1016/0022-2836(90)90306-7
Schmieger, H., 1972. Phage P22-mutants with increased or decreased transduction abilities. Mol. Gen. Genet. MGG 119, 75–88.
Seed, K.D., 2015. Battling Phages: How Bacteria Defend against Viral Attack. PLOS Pathog. 11, e1004847. doi:10.1371/journal.ppat.1004847
Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma. Oxf. Engl. 30, 2068–2069. doi:10.1093/bioinformatics/btu153
Shearwin, K.E., Brumby, A.M., Egan, J.B., 1998. The Tum protein of coliphage 186 is an antirepressor. J. Biol. Chem. 273, 5708–5715.
Shearwin, K.E., Egan, J.B., 2000. Establishment of Lysogeny in Bacteriophage 186: DNA binding and transcriptional activation by the CII protein. J. Biol. Chem. 275, 29113–29122. doi:10.1074/jbc.M004574200

Shimizu, T., Ohta, Y., Noda, M., 2009. Shiga toxin 2 is specifically released from bacterial cells by two different mechanisms. Infect. Immun. 77, 2813–2823. doi:10.1128/IAI.00060-09

Singletary, L.A., Karlinsey, J.E., Libby, S.J., Mooney, J.P., Lokken, K.L., Tsolis, R.M., Byndloss, M.X., Hirao, L.A., Gaulke, C.A., Crawford, R.W., Dandekar, S., Kingsley, R.A., Msefula, C.L., Heyderman, R.S., Fang, F.C., 2016. Loss of Multicellular Behavior in Epidemic African Nontyphoidal *Salmonella enterica* Serovar Typhimurium ST313 Strain D23580. mBio 7, e02265. doi:10.1128/mBio.02265-15

Sittka, A., Pfeiffer, V., Tedin, K., Vogel, J., 2007. The RNA chaperone Hfq is essential for the virulence of *Salmonella* typhimurium. Mol. Microbiol. 63, 193–217. doi:10.1111/j.1365-2958.2006.05489.x

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., Holmes, I.H., 2009. JBrowse: A next-generation genome browser. Genome Res. 19, 1630–1638. doi:10.1101/gr.094607.109

Slauch, J.M., Lee, A.A., Mahan, M.J., Mekalanos, J.J., 1996. Molecular characterization of the oafA locus responsible for acetylation of *Salmonella* typhimurium O-antigen: oafA is a member of a family of integral membrane trans-acylases. J. Bacteriol. 178, 5904–5909.

Srikumar, S., Kröger, C., Hébrard, M., Colgan, A., Owen, S.V., Sivasankaran, S.K., Cameron, A.D., Hokamp, K., Hinton, J.C., 2015. RNA-seq Brings New Insights to the Intra-Macrophage Transcriptome of *Salmonella* Typhimurium. PLoS Pathog. 11, e1005262.

Stanley, T.L., Ellermeier, C.D., Slauch, J.M., 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar typhimurium survival in Peyer's patches. J. Bacteriol. 182, 4406–13.

Steinbacher, S., Miller, S., Baxa, U., Budisa, N., Weintraub, A., Seckler, R., Huber, R., 1997. Phage P22 tailspike protein: crystal structure of the head-binding domain at 2.3 A, fully refined structure of the endorhamnosidase at 1.56 A resolution, and the molecular basis of O-antigen recognition and cleavage. J. Mol. Biol. 267, 865–880. doi:10.1006/jmbi.1997.0922

Stevens, R.H., de Moura Martins Lobo Dos Santos, C., Zuanazzi, D., de Accioly Mattos, M.B., Ferreira, D.F., Kachlany, S.C., Tinoco, E.M.B., 2013. Prophage induction in lysogenic *Aggregatibacter actinomycetemcomitans* cells co-cultured with human gingival fibroblasts, and its effect on leukotoxin release. Microb. Pathog. 54, 54–59. doi:10.1016/j.micpath.2012.09.005

Surber, M.W., Maloy, S., 1998. The PutA protein of *Salmonella* typhimurium catalyzes the two steps of proline degradation via a leaky channel. Arch. Biochem. Biophys. 354, 281–287. doi:10.1006/abbi.1998.0697

Swaminathan, B., Barrett, T.J., Hunter, S.B., Tauxe, R.V., CDC PulseNet Task Force, 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. 7, 382–389. doi:10.3201/eid0703.010303

Taylor, K., Wegrzyn, G., 1995. Replication of coliphage lambda DNA. FEMS Microbiol. Rev. 17, 109–119.

Tenbroeck, C., 1920. A group of Paratyphoid Bacilli from Animals Closely Resembling those Found in Man. J. Exp. Med. 32, 19–31.

Tennant, S.M., Diallo, S., Levy, H., Livio, S., Sow, S.O., Tapia, M., Fields, P.I., Mikoleit, M., Tamboura, B., Kotloff, K.L., Nataro, J.P., Galen, J.E., Levine, M.M., 2010. Identification by PCR of non-typhoidal *Salmonella enterica* serovars associated with invasive infections among febrile patients in Mali. PLoS Negl. Trop. Dis. 4, e621. doi:10.1371/journal.pntd.0000621

The, H.C., Thanh, D.P., Holt, K.E., Thomson, N.R., Baker, S., 2016. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. Nat Rev Micro 14, 235–250.

The Salmonella Subcommittee of the Nomenclature Committee of the International Society for Microbiology, 1934. The Genus *Salmonella* Lignières, 1900. J. Hyg. (Lond.) 34, 333–350.

Thomason, M.K., Storz, G., 2010. Bacterial antisense RNAs: How many are there and what are they doing? Annu. Rev. Genet. 44, 167–188. doi:10.1146/annurev-genet-102209-163523

Thomson, N., Baker, S., Pickard, D., Fookes, M., Anjum, M., Hamlin, N., Wain, J., House, D., Bhutta, Z., Chan, K., Falkow, S., Parkhill, J., Woodward, M., Ivens, A., Dougan, G., 2004. The Role of Prophage-like Elements in the Diversity of *Salmonella enterica* Serovars. J. Mol. Biol. 339, 279–300. doi:10.1016/j.jmb.2004.03.058

Tindall, B.J., Grimont, P.A.D., Garrity, G.M., Euzeby, J.P., 2005. Nomenclature and taxonomy of the genus *Salmonella*. Int. J. Syst. Evol. Microbiol. 55, 521–524. doi:10.1099/ijs.0.63580-0

Tsolis, R.M., Xavier, M.N., Santos, R.L., Baumler, A.J., 2011. How To Become a Top Model: Impact of Animal Experimentation on Human *Salmonella* Disease Research. Infect. Immun. 79, 1806–1814. doi:10.1128/IAI.01369-10

Tucker, C.P., Heuzenroeder, M.W., 2004. ST64B is a defective bacteriophage in *Salmonella enterica* serovar Typhimurium DT64 that encodes a functional immunity region capable of mediating phage-type conversion. Int. J. Med. Microbiol. IJMM 294, 59–63. doi:10.1016/j.ijmm.2003.12.001

Turnbull, L., Toyofuku, M., Hynen, A.L., Kurosawa, M., Pessi, G., Petty, N.K., Osvath, S.R., Carcamo-Oyarce, G., Gloag, E.S., Shimoni, R., Omasits, U., Ito, S., Yap, X., Monahan, L.G., Cavaliere, R., Ahrens, C.H., Charles, I.G., Nomura, N., Eberl, L., Whitchurch, C.B., 2016. Explosive cell lysis as a mechanism for the biogenesis of bacterial membrane vesicles and biofilms. Nat. Commun. 7, 11220. doi:10.1038/ncomms11220

Twort, F.W., 1915. An Investigation on the Nature of Ultra-Microscopic Viruses. The Lancet, Originally published as Volume 2, Issue 4814 186, 1241–1243. doi:10.1016/S0140-6736(01)20383-3

Uche, I.V., MacLennan, C.A., Saul, A., 2017. A Systematic Review of the Incidence, Risk Factors and Case Fatality Rates of Invasive Nontyphoidal *Salmonella* (iNTS) Disease in Africa (1966 to 2014). PLoS Negl. Trop. Dis. 11, e0005118. doi:10.1371/journal.pntd.0005118

Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Bernard, S., Casadesús, J., Platt, D.J., Olsen, J.E., 2000. Host adapted serotypes of *Salmonella enterica*. Epidemiol. Infect. 125, 229–255.

Vågene, Å.J., Campana, M.G., Robles García, N.M., Warinner, C., Spyrou, M.A., Andrades Valtueña, A., Huson, D., Tuross, N., Herbig, A., Bos, K.I., Krause, J., 2017. *Salmonella enterica* genomes recovered from victims of a major 16th century epidemic in Mexico. bioRxiv. doi:10.1101/106740

Veses-Garcia, M., Liu, X., Rigden, D.J., Kenny, J.G., McCarthy, A.J., Allison, H.E., 2015. Transcriptomic Analysis of Shiga-Toxigenic Bacteriophage Carriage Reveals a Profound Regulatory Effect on Acid Resistance in *Escherichia coli*. Appl. Environ. Microbiol. 81, 8118–8125. doi:10.1128/AEM.02034-15

Villafane, R., Zayas, M., Gilcrease, E.B., Kropinski, A.M., Casjens, S.R., 2008. Genomic analysis of bacteriophage epsilon 34 of *Salmonella enterica* serovar Anatum (15+). BMC Microbiol. 8, 227. doi:10.1186/1471-2180-8-227

Vogel, J., 2009. A rough guide to the non-coding RNA world of *Salmonella*. Mol. Microbiol. 71, 1–11. doi:10.1111/j.1365-2958.2008.06505.x

Wagner, E.G.H., Altuvia, S., Romby, P., 2002. Antisense RNAs in Bacteria and Their Genetic Elements. Adv. Genet., Homology Effects 46, 361–398. doi:10.1016/S0065-2660(02)46013-0
Wagner, G.P., Kin, K., Lynch, V.J., 2013. A model based criterion for gene expression calls using RNA-seq data. Theory Biosci. Theor. Den Biowissenschaften 132, 159–64. doi:10.1007/s12064-013-0178-3
Wagner, G.P., Kin, K., Lynch, V.J., 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. Theor. Den Biowissenschaften 131, 281–5. doi:10.1007/s12064-012-0162-3
Wain, J., Keddy, K.H., Hendriksen, R.S., Rubino, S., 2013. Using next generation sequencing to tackle non-typhoidal *Salmonella* infections. J. Infect. Dev. Ctries. 7, 1–5.
Wallecha, A., Munster, V., Correnti, J., Chan, T., van der Woude, M., 2002. Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. J. Bacteriol. 184, 3338–3347.
Wang, Q., Frye, J.G., McClelland, M., Harshey, R.M., 2004. Gene expression patterns during swarming in *Salmonella* typhimurium: genes specific to surface growth and putative new motility and pathogenicity genes. Mol. Microbiol. 52, 169–187. doi:10.1111/j.1365-2958.2003.03977.x
Wang, X., Kim, Y., Ma, Q., Hong, S.H., Pokusaeva, K., Sturino, J.M., Wood, T.K., 2010. Cryptic prophages help bacteria cope with adverse environments. Nat. Commun. 1, 147. doi:10.1038/ncomms1146
Weisbroth, S.H., 1979. Chapter 9 - Bacterial and Mycotic Diseases, in: The Laboratory Rat. Academic Press, pp. 193–241. doi:10.1016/B978-0-12-074901-0.50016-2
White, A.P., Gibson, D.L., Kim, W., Kay, W.W., Surette, M.G., 2006. Thin aggregative fimbriae and cellulose enhance long-term survival and persistence of *Salmonella*. J. Bacteriol. 188, 3219–3227. doi:10.1128/JB.188.9.3219-3227.2006
Widal, F., 1896. Le sérodiagnostic de la fièvre typhoïde. Bull. Société Médicale Hôp. Paris 690.
Winter, S.E., Thiennimitr, P., Winter, M.G., Butler, B.P., Huseby, D.L., Crawford, R.W., Russell, J.M., Bevins, C.L., Adams, L.G., Tsolis, R.M., Roth, J.R., Bäumler, A.J., 2010. Gut inflammation provides a respiratory electron acceptor for *Salmonella*. Nature 467, 426–429. doi:10.1038/nature09415
Wollin, R., Stocker, B.A., Lindberg, A.A., 1987. Lysogenic conversion of *Salmonella* typhimurium bacteriophages A3 and A4 consists of O-acetylation of rhamnose of the repeating unit of the O-antigenic polysaccharide chain. J. Bacteriol. 169, 1003–1009.
Wong, V.K., Baker, S., Connor, T.R., Pickard, D., Page, A.J., Dave, J., Murphy, N., Holliman, R., Sefton, A., Millar, M., Dyson, Z.A., Dougan, G., Holt, K.E., Consortium, I.T., Parkhill, J., Feasey, N.A., Kingsley, R.A., Thomson, N.R., Keane, J.A., Weill, F.-X., Hello, S.L., Hawkey, J., Edwards, D.J., Harris, S.R., Cain, A.K., Hadfield, J., Hart, P.J., Thieu, N.T.V., Klemm, E.J., Breiman, R.F., Watson, C.H., Edmunds, W.J., Kariuki, S., Gordon, M.A., Heyderman, R.S., Okoro, C., Jacobs, J., Lunguya, O., Msefula, C., Chabalgoity, J.A., Kama, M., Jenkins, K., Dutta, S., Marks, F., Campos, J., Thompson, C., Obaro, S., MacLennan, C.A., Dolecek, C., Keddy, K.H., Smith, A.M., Parry, C.M., Karkey, A., Dongol, S., Basnyat, B., Arjyal, A., Mulholland, E.K., Campbell, J.I., Dufour, M., Bandaranayake, D., Toleafoa, T.N., Singh, S.P., Hatta, M., Newton, P.N., Dance, D., Davong, V., Onsare, R.S., Isaia, L., Thwaites, G., Wijedoru, L., Crump, J.A., Pinna, E.D., Nair, S., Nilles, E.J., Thanh, D.P., Turner, P., Soeng, S., Valcanis, M., Powling, J., Dimovski, K., Hogg, G., Farrar, J., Mather, A.E., Amos, B., 2016. An extended genotyping

framework for *Salmonella enterica* serovar Typhi, the cause of human typhoid. Nat. Commun. 7, ncomms12827. doi:10.1038/ncomms12827

Wong, V.K., Baker, S., Pickard, D.J., Parkhill, J., Page, A.J., Feasey, N.A., Kingsley, R.A., Thomson, N.R., Keane, J.A., Weill, F.-X., Edwards, D.J., Hawkey, J., Harris, S.R., Mather, A.E., Cain, A.K., Hadfield, J., Hart, P.J., Thieu, N.T.V., Klemm, E.J., Glinos, D.A., Breiman, R.F., Watson, C.H., Kariuki, S., Gordon, M.A., Heyderman, R.S., Okoro, C., Jacobs, J., Lunguya, O., Edmunds, W.J., Msefula, C., Chabalgoity, J.A., Kama, M., Jenkins, K., Dutta, S., Marks, F., Campos, J., Thompson, C., Obaro, S., MacLennan, C.A., Dolecek, C., Keddy, K.H., Smith, A.M., Parry, C.M., Karkey, A., Mulholland, E.K., Campbell, J.I., Dongol, S., Basnyat, B., Dufour, M., Bandaranayake, D., Naseri, T.T., Singh, S.P., Hatta, M., Newton, P., Onsare, R.S., Isaia, L., Dance, D., Davong, V., Thwaites, G., Wijedoru, L., Crump, J.A., De Pinna, E., Nair, S., Nilles, E.J., Thanh, D.P., Turner, P., Soeng, S., Valcanis, M., Powling, J., Dimovski, K., Hogg, G., Farrar, J., Holt, K.E., Dougan, G., 2015. Phylogeographical analysis of the dominant multidrug-resistant H58 clade of *Salmonella* Typhi identifies inter- and intracontinental transmission events. Nat. Genet. 47, 632–639. doi:10.1038/ng.3281

Xu, X., McAteer, S.P., Tree, J.J., Shaw, D.J., Wolfson, E.B.K., Beatson, S.A., Roe, A.J., Allison, L.J., Chase-Topping, M.E., Mahajan, A., Tozzoli, R., Woolhouse, M.E.J., Morabito, S., Gally, D.L., 2012. Lysogeny with Shiga toxin 2-encoding bacteriophages represses type III secretion in enterohemorrhagic *Escherichia coli*. PLoS Pathog. 8, e1002672. doi:10.1371/journal.ppat.1002672

Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., Tang, X., Wang, J., Xiong, Z., Dong, J., Xue, Y., Zhu, Y., Xu, X., Sun, L., Chen, S., Nie, H., Peng, J., Xu, J., Wang, Y., Yuan, Z., Wen, Y., Yao, Z., Shen, Y., Qiang, B., Hou, Y., Yu, J., Jin, Q., 2005. Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. Nucleic Acids Res. 33, 6445–6458. doi:10.1093/nar/gki954

Yang, J., Barrila, J., Roland, K.L., Kilbourne, J., Ott, C.M., Forsyth, R.J., Nickerson, C.A., 2015. Characterization of the Invasive, Multidrug Resistant Non-typhoidal *Salmonella* Strain D23580 in a Murine Model of Infection. PLoS Negl. Trop. Dis. 9, e0003839. doi:10.1371/journal.pntd.0003839

Yang, J., Barrila, J., Roland, K.L., Ott, C.M., Nickerson, C.A., 2016. Physiological fluid shear alters the virulence potential of invasive multidrug-resistant non-typhoidal *Salmonella* Typhimurium D23580. Npj Microgravity 2, 16021. doi:10.1038/npjmgrav.2016.21

Yang, Z., Soderholm, A., Lung, T.W.F., Giogha, C., Hill, M.M., Brown, N.F., Hartland, E., Teasdale, R.D., 2015. SseK3 Is a *Salmonella* Effector That Binds TRIM32 and Modulates the Host's NF-κB Signalling Activity. PLOS ONE 10, e0138529. doi:10.1371/journal.pone.0138529

Yu, D., Sawitzke, J.A., Ellis, H., Court, D.L., 2003. Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. Proc. Natl. Acad. Sci. 100, 7207–7212. doi:10.1073/pnas.1232375100

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F.M., Larsen, M.V., 2012. Identification of acquired antimicrobial resistance genes. J. Antimicrob. Chemother. 67, 2640–2644. doi:10.1093/jac/dks261

Zhou, Z., Lundstrøm, I., Tran-Dien, A., Duchêne, S., Alikhan, N.-F., Sergeant, M.J., Langridge, G., Fotakis, A.K., Nair, S., Stenøien, H.K., Hamre, S.S., Casjens, S., Christophersen, A., Quince, C., Thomson, N.R., Weill, F.-X., Ho, S.Y.W., Gilbert, M.T.P., Achtman, M., 2017. Millennia of genomic stability within the invasive Para C Lineage of *Salmonella enterica*. bioRxiv. doi:10.1101/105759

Zivot, J.B., Hoffman, W.D., 1995. Pathogenic effects of endotoxin. New Horiz. Baltim. Md 3, 267–275.
Zou, Q.H., Li, Q.H., Zhu, H.Y., Feng, Y., Li, Y.G., Johnston, R.N., Liu, G.R., Liu, S.L., 2010. SPC-P1: a pathogenicity-associated prophage of *Salmonella* paratyphi C. BMC Genomics 11, 729. doi:10.1186/1471-2164-11-729

# Appendix i

Accession numbers and metadata for the 99 genome sequences used in this thesis. Where multiple isolates were collected from the same patient, a single representative isolate was chosen. This table includes information about the publicly available African genomes which were included in the analysis for context as well as Public Health England strains. Empty cells or the text N/A indicate no available data or not applicable (in the case of reference genomes that were not isolated by PHE).

| ID | Accession number | Lineage | PHE centre (if UK) | Travel to sSA | country associated with | travel info available | Sample Type | Isolate source | Report Date | AMR genes or plasmid replicons detected in genome | Antimicrobial resistance phenotype |
|---|---|---|---|---|---|---|---|---|---|---|---|
| DT2B | ERR024398 | UK | #N/A | | UK | #N/A | | Faeces | #N/A | #N/A | N/A |
| U10 | SRR1965263 | UK | Yorkshire and Humber | N | UK | Y | Human | Faeces | 03/06/2014 | FII-S1;FIB-17 | $ |
| U11 | SRR1965387 | UK | West Midlands | N | UK | Y | Human | Faeces | 05/06/2014 | FII-S1;FIB-17 | $ |
| U12 | SRR1967419 | UK | London | N | UK | Y | Human | N/A | 19/06/2014 | FII-S1;FIB-17 | $ |
| U14 | SRR1966418 | UK | East of England | N | UK | Y | Human | Blood | 26/08/2014 | FII-S1;FIB-17 | $ |
| U15 | SRR1645388 | UK | East Midlands | | UK | N | Human | Faeces | 24/05/2012 | FII-S1;FIB-17;gyrA_SET[83:S-F] | SUL256,NAL16,CIP |
| U16 | SRR1645171 | UK | Yorkshire and Humber | | UK | N | Human | Faeces | 15/06/2012 | FII-S1;FIB-17 | $ |
| U2 | SRR1646230 | UK | East of England | N | UK | Y | Human | Faeces | 02/11/2012 | FII-S1;FIB-17 | $ |
| U22 | SRR3048892 | UK | South East | | UK | N | Human | Faeces | 18/06/2014 | #N/A | $ |
| U23 | SRR1969854 | UK | South East | N | UK | Y | Human | Faeces | 10/10/2014 | FII-S1;FIB-17 | $ |
| U24 | SRR1967447 | UK | London | N | UK | Y | Human | Faeces | 23/10/2014 | FII-S1;FIB-17 | $ |
| U25 | SRR1958036 | UK | London | | UK | N | Human | Faeces | 10/11/2014 | FII-S1;FIB-17 | $ |
| U27 | SRR1958303 | UK | London | Y | Uganda | Y | Human | Faeces | 07/11/2014 | FII-S1;FIB-17 | $ |
| U28 | SRR1960363 | UK | None | | UK | N | Human | Faeces | 10/12/2014 | FII-S1;FIB-17 | $ |
| U29 | SRR1968084 | UK | London | | UK | N | Human | Faeces | 26/01/2015 | None found | $ |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| U3 | SRR1646238 | UK | East of England | | UK | N | Human | Faeces | 10/12/2012 | FII-S1;FIB-17 | $ |
| U30 | SRR1968221 | UK | #N/A | | UK | N | Food | Raw Produce | 24/02/2015 | FII-S1;FIB-17 | $ |
| U31 | SRR1967938 | UK | London | | UK | N | Human | Faeces | 18/02/2015 | FII-S1;FIB-17 | $ |
| U32 | SRR1969492 | UK | London | | UK | N | Human | Blood | 25/02/2015 | FII-S1;FIB-17 | $ |
| U34 | SRR1966635 | UK | West Midlands | N | UK | Y | Human | Faeces | 05/03/2015 | FII-S1;FIB-17 | $ |
| U36 | SRR1966593 | UK | London | N | UK | Y | Human | Faeces | 05/03/2015 | FII-S1;FIB-17 | N/A |
| U37 | SRR1968962 | UK | East of England | N | UK | Y | Human | Faeces | 12/03/2015 | FII-S1;FIB-17 | $ |
| U38 | SRR1967438 | UK | Yorkshire and Humber | N | UK | Y | Human | Faeces | 17/03/2015 | FII-S1;FIB-17 | $ |
| U4 | SRR1645545 | UK | West Midlands | N | UK | Y | Human | Faeces | 20/12/2012 | FII-S1;FIB-17 | $ |
| U40 | SRR1968061 | UK | London | | UK | N | Human | Faeces | 18/03/2015 | FII-S1;FIB-17 | $ |
| U43 | SRR5451258 | UK | Yorkshire and Humber | | UK | N | Human | Faeces | 10/06/2015 | #N/A | $ |
| U44 | SRR5451287 | UK | London | N | Pakistan | Y | Human | Faeces | 26/05/2015 | FII-S1;FIB-17 | $ |
| U46 | SRR5451286 | UK | London | N | UK | Y | Human | Faeces | 12/06/2015 | FII-S1;FIB-17 | $ |
| U47 | SRR5451276 | UK | South West | N | UK | Y | Human | Faeces | 16/06/2015 | FII-S1;FIB-17 | $ |
| U48 | SRR5451252 | UK | Yorkshire and Humber | N | UK | Y | Human | Faeces | 29/06/2015 | FII-S1;FIB-17 | $ |
| U49 | SRR5451278 | UK | London | | UK | N | Human | Faeces | 30/06/2015 | FII-S1;FIB-17 | $ |
| U5 | SRR1645768 | UK | Wales | | UK | N | Human | Faeces | 21/12/2012 | FII-S1;FIB-17 | $ |
| U50 | SRR5451285 | UK | London | | UK | N | Human | Faeces | 02/07/2015 | FII-S1;FIB-17 | $ |
| U51 | SRR3323011 | UK | London | N | UK | Y | Human | Faeces | 06/08/2015 | FII-S1;FIB-17 | $ |
| U52 | SRR3284779 | UK | London | N | UK | Y | Human | Faeces | 06/08/2015 | FII-S1;FIB-17 | $ |
| U53 | SRR3285477 | UK | East of England | | UK | N | Human | Faeces | 07/08/2015 | FII-S1;FIB-17 | $ |
| U54 | SRR5193685 | UK | South East | N | UK | Y | Human | Faeces | 10/08/2015 | None found | AMP8,SUL256,STR16,TET8 |
| U55 | SRR3322345 | UK | North West | N | UK | Y | Human | Faeces | 08/09/2015 | FII-S1;FIB-17 | $ |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| U56 | SRR5451268 | UK | Yorkshire and Humber | N | Pakistan | Y | Human | Faeces | 12/10/2015 | FII-S1;FIB-17 | $ |
| U57 | SRR5451279 | UK | London | N | UK | Y | Human | Faeces | 25/09/2015 | FII-S1;FIB-17 | AMP8 |
| U58 | SRR5451270 | UK | South West | N | UK | Y | Human | Faeces | 01/10/2015 | FII-S1;FIB-17;I1 | $ |
| U59 | SRR5451254 | UK | East of England | | UK | N | Human | Faeces | 19/10/2015 | FII-S1;FIB-17 | $ |
| U61 | SRR5451275 | UK | South East | N | UK | Y | Human | Faeces | 27/10/2015 | FII-S1;FIB-17 | $ |
| U62 | SRR5451283 | UK | East of England | | UK | N | Human | Faeces | 19/10/2015 | FII-S1;FIB-17 | $ |
| U63 | SRR5451281 | UK | London | | UK | N | Human | Blood | 22/10/2015 | FII-S1;FIB-17 | $ |
| U64 | SRR5451266 | UK | London | N | UK | Y | Human | Faeces | 27/10/2015 | FII-S1;FIB-17 | $ |
| U65 | SRR5451255 | UK | London | N | Portugal | Y | Human | Faeces | 28/10/2015 | FII-S1;FIB-17;FII-2 | $ |
| U66 | SRR3322379 | UK | London | | UK | N | Human | Faeces | 03/11/2015 | FII-S1;FIB-17 | $ |
| U69 | SRR5451261 | UK | #N/A | | UK | N | Dog | Dog | 22/12/2015 | FII-S1;FIB-17 | $ |
| U7 | SRR1966408 | UK | East of England | N | UK | Y | Human | Faeces | 14/05/2014 | FII-S1;FIB-17 | $ |
| U70 | SRR5451259 | UK | London | | UK | N | Human | Faeces | 04/01/2016 | FII-S1;FIB-17 | $ |
| U74 | SRR5451280 | UK | North West | N | Cape verde | Y | Human | Faeces | 03/02/2016 | FII-S1;FIB-17 | $ |
| U75 | SRR5451257 | UK | East of England | N | Belize | Y | Human | Faeces | 26/02/2016 | FII-S1;FIB-17 | $ |
| U76 | SRR5451251 | UK | South East | N | India | Y | Human | Faeces | 02/03/2016 | FII-S1;FIB-17 | AMP8,CHL8,CHL16,SUL256,STR16,NAL16,CIP0.64,CIP0.5 |
| U77 | SRR5451260 | UK | East of England | N | UK | Y | Human | Faeces | 29/02/2016 | FII-S1;FIB-17 | $ |
| U78 | SRR5451265 | UK | West Midlands | N | India | Y | Human | Faeces | 11/03/2016 | FII-S1;FIB-17 | $ |
| U8 | SRR1966656 | UK | London | N | UK | Y | Human | Faeces | 29/05/2014 | FII-S1;FIB-17 | $ |
| U80 | SRR5451253 | UK | East of England | N | UK | Y | Human | Faeces | 21/03/2016 | FII-S1;FIB-17 | $ |
| U81 | SRR5215816 | UK | East of England | N | UK | Y | Human | Faeces | 29/03/2016 | FII-S1;FIB-17 | $ |
| U82 | SRR5451272 | UK | London | N | UK | Y | Human | Faeces | 04/04/2016 | FII-S1;FIB-17 | $ |
| U83 | SRR5451267 | UK | South East | N | UK | Y | Human | Faeces | 22/04/2016 | FII-S1;FIB-17 | $ |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| U84 | SRR5451256 | UK | East of England | N | UK | Y | Human | Faeces | 22/04/2016 | FII-S1;FIB-17 | $ |
| U86 | SRR5451274 | UK | South East | N | UK | Y | Human | Faeces | 26/05/2016 | None | $ |
| U87 | SRR5451262 | UK | North West | N | India | Y | Human | Faeces | 26/05/2016 | FII-S1;FIB-17 | $ |
| U9 | SRR1968363 | UK | East of England | N | UK | Y | Human | N/A | 30/05/2014 | FII-S1;FIB-17 | $ |
| U26 | SRR1968109 | Unknown | London | | UK | N | Human | Faeces | 03/12/2014 | #N/A | AMP8,TET8 |
| 6325U | ERR024780 | II | #N/A | | Uganda | N/A | | Blood | #N/A | #N/A | N/A |
| A082 | ERR023770 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| A13198 | ERR023769 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| A16083 | ERR023784 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| A357 | ERR023771 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| A38589 | ERR023782 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| A39051 | ERR023797 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| C13184 | ERR023796 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| C2110 | ERR023777 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| D11578 | ERR023783 | II | #N/A | | Malawi | N/A | | Blood | #N/A | #N/A | N/A |
| I32 | ERR023810 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| I7 | ERR023812 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| J17 | ERR023794 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| J20 | ERR023805 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| J27 | ERR023795 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| J3 | ERR023793 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| M1111568 | ERR024769 | II | #N/A | | Mozambique | N/A | | Faeces | #N/A | #N/A | N/A |
| M1605206 | ERR024424 | II | #N/A | | Mozambique | N/A | | Blood | #N/A | #N/A | N/A |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| M2907772 | ERR024420 | II | #N/A | | Mozambique | N/A | | Blood | #N/A | #N/A | N/A |
| P51 | ERR023803 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| P73 | ERR023802 | II | #N/A | | Mali | N/A | | Blood | #N/A | #N/A | N/A |
| U1 | SRR1646227 | II | London | Y | Ghana | Y | Human | Blood | 28/08/2012 | FII-S1;FIB-17;TEM-1;strB,strA;dfrA-1;sul-2;sul-1;catA-1;gyrA_SET[83:S-Y] | AMP8,CHL8,SU256L,STR16,TMP2,NAL16,CIP |
| U13 | SRR1966509 | II | London | Y | Nigeria | Y | Human | Blood | 21/07/2014 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-1;sul-2;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U17 | SRR1967883 | II | London | Y | African continent (?) | Y | Human | Blood | 07/05/2014 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-2;sul-1;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U35 | SRR1966023 | II | London | Y | Nigeria | Y | Human | Blood | 04/03/2015 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-1;sul-2;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U42 | SRR3285401 | II | London | | UK (African name) | N | Human | Blood | 27/05/2015 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-2;sul-1;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U45 | SRR5451277 | II | London | | UK (STD clinic) | N | Human | Blood | 17/07/2015 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-2;sul-1 | AMP8,SUL256,STR16,TMP2 |
| U60 | SRR5451282 | II | London | Y | Kenya | Y | Human | Blood | 08/10/2015 | FII-S1;FIB-17;HI2;CTX-M-15;OXA-1;TEM-1;strA;strB;dfrA-1;dfrA-14;sul-2;sul-1;catA-1;tet(A)-1;terF;terE;terD;terC;terB | AMP8,CHL8,CHL16,SUL256,TOB2,NAL16,CIP0.064,CIP0.5,CTX0.5,CTX1 |
| U67 | SRR3323046 | II | East Midlands | Y | Nigeria | Y | Human | Blood | 23/11/2015 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-2;sul-1;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U68 | SRR5451264 | II | London | | UK (African name) | N | Human | Pus | 10/12/2015 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-1;sul-2;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U72 | SRR5451284 | II | London | N | UK | Y | Human | N/A | 03/02/2016 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-2;sul-1;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |
| U73 | SRR5451273 | II | London | N | UK (African name, African markets) | Y | Human | Blood | 03/02/2016 | FII-S1;FIB-17;TEM-1 | AMP8 |
| U85 | SRR5451269 | II | London | Y | Ghana | Y | Human | Faeces | 09/05/2016 | FII-S1;FIB-17;TEM-1;strA;strB;dfrA-1;sul-1;sul-2;catA-1 | AMP8,CHL8,CHL16,SUL256,STR16,TMP2 |