GENOMIC ANALYSIS OF INVASIVE NON- TYPHOIDAL *SALMONELLA* (NTS) FROM SUB-SAHARAN AFRICA

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

By

CHRISTINA BRONOWSKI

Department of Clinical Infection, Microbiology and Immunology
Institute of Infection and Global Health
University of Liverpool

November 2010

To my Mum

DECLARATION

The work in this thesis is the result of my own work. The material presented here has not been presented and will not be presented, wether wholly or in part for another degree or qualification. Some of the technical procedures were carried out in collaboration with other people and reference has been made to their work and/or data where appropriate.
During the PhD I was based in the Institute of Infection and Global Health, Department of Clinical Infection and Microbiology, where most of my work was carried out in the laboratory of Dr C. Winstanley. Cell Invasion assays were performed in the Laboratory of Dr P. Wigley at the Department of Infection Biology. Genome sequencing and analysis was carried out by the Department of Functional and Comparative Genomics in the Institute of Integrative Biology, University of Liverpool and at the Wellcome Trust Sanger Institute, Pathogen Genomics.

This PhD was funded by the Medical Research Council, Illumina sequencing was funded by the Wellcome Trust.

Christina Bronowski

ABSTRACT

Non-typhoidal *Salmonella* (NTS) are an important cause of bacteraemia in children and HIV-infected adults in sub-Saharan Africa. The most commonly isolated NTS serovars are Typhimurium and Enteritidis. Our NTS collection from Malawi showed that *S.* Bovismorbificans accounted for 1 percent of the total number of invasive bacteraemia NTS isolated over a six year period (1998 to 2004), making it the third most common serovar and suggesting an increasing importance as an invasive serovar. NTS strains from these areas exhibit extensive antimicrobial resistance with no identifiable animal host, which has lead to the assumption that they differ from classic gastroenteritis strains.

Suppression Subtractive Hybridisation (SSH) is a technique capable of identifying genetic sequences present in one strain but absent from another. SSH was applied to investigate four representative Malawian paediatric bacteraemia NTS strains. *S.* Typhimurium D26104, *S.* Heidelberg D23734 and *S.* Bovismorbificans 3114, were subtracted against *S.* Typhimurium LT2 (G639). *S.* Enteritidis D21685 was subtracted against the *S.* Enteritidis laboratory strain NCTC13349. The distribution of a selection of SSH sequences among a panel of strains from the UK, Africa and other tropical locations was assessed by PCR and Dot Blot analysis.

Expression of a number of *S.* Heidelberg SSH sequences was tested in a subset of strains. The genome of S. Bovismorbificans 3114 was determined using 454 pyrosequencing and the genomes of a further 14 Malawian bacteraemia and 3 UK veterinary strains were investigated using Illumina sequencing. SSH sequences for *S.* Heidelberg and *S.* Bovismorbificans were more diverse in nature than those identified in *S.* Typhimurium. *S.* Typhimurium subtracted sequences were mainly bacteriophage- and plasmid-related, whilst *S.* Enteritidis produced few subtracted sequences, indicating limited additional sequences present in D21685 compared to NCTC13349. Distribution data suggests that some subtracted sequences are specific to certain groups of NTS serovars, while others appeared to be distributed according to geographical source. There was no evidence for major differences in the accessory genome content between African invasive and gastrointestinal isolates of *S.* Heidelberg. Six of the *S.* Heidelberg SSH sequences were part of fimbrial operons. The *tcf* operon, associated with host specificity of *S.* Typhi, and the *stk* operon, reported previously in *S.* Paratyphi, were both present in the majority of *S.* Heidelberg isolates, but had restricted distributions amongst other serovars tested. Expression of the *stk* operon amongst isolates of *S.* Heidelberg was variable. Three of the seven targeted genes were not expressed in a UK veterinary isolate of *S.* Heidelberg.

MLST analysis of *S.* Typhimurium strains from sub-Saharan Africa confirmed the presence of a distinct sequence type, ST313, associated with invasive disease in these regions, while MLST of *S.* Bovismorbificans showed that the majority of Malawian isolates were of the main lineage ST142. Three major regions of difference (RODs) have been identified within the *S.* Bovismorbificans 3114 genome compared to other *Salmonella* serovars, which showed variable distribution between the 18 Bovismorbificans strains, as well as the presence of a *S.* Bovismorbificans virulence plasmid, carrying the *spv* (*Salmonella* plasmid virulence) cassette. The *S.* Bovismorbificans 3114 genome showed some evidence for genome degradation through pseudogene formation, previously reported for other *Salmonella* serovars which may have led to host specialization.

ACKNOWLEDGEMENTS

This work would not have been possible without a long list of people, who have helped, advised, encouraged and supported me. Everyone I have met and had a chance to work with has been friendly and kind.

First and foremost I would like to thank my supervisor Dr Craig Winstanley, without whom this project truly would not have been possible. Craig has been understanding and supportive throughout. I would also like to thank Dr Nick Thomson whose enthusiasm and positivity are infectious and who has given me the great opportunity of making this work a bit more special.

Sadly my first supervisor Prof. Tony Hart passed away, I know that this topic was very close to his heart and I hope we did him proud. I would also like to thank Dr Paul Wigley, Dr Chris Parry and Prof Alan McCarthy.

I would like to thank my family, particularly my mum who is always there for me offering encouragement without pressure, my brother for being available for all questions IT-related and Jürgen for always being available for a chat.
Sadly my beloved Oma passed away, she always believed in me and I hope she would have been proud.

I would like to thank Dirk who believed in me and who, particularly in the last months of thesis writing, has endured my moods and kept me fed, watered and in clean clothes, offering every support I needed.

The people who have come and gone from our little office have made the last four years very enjoyable and I would like to thank you all for that, particular thanks goes to Vicky and Bruno. Jo has been there from the start, being a good friend, offering help and advice both in the lab and the real world. Jackie, Fred and Clare have made my life easier and my work enjoyable.

Some of my closest friends have offered a constant stream of encouragement from the distance and I would like to say thank you to Jenny, Faye and Federica.

ABBREVIATIONS

%	percent
µg	micrograms
µl	microlitre
µM	micromolar
µm	micrometre
A	adenosine
aa	amino acids
ACT	Artemis ComparisonTool
AIDS	acquired immunodeficiency syndrome
bp	base pairs
BLAST	Basic Local Alignment Search Tool
C	cytosine
°C	degrees centigrade
cc	clonal complex
cDNA	complementary deoxyribonucleic acid
CDS	coding sequence
CFU	colony forming units
CO_2	carbon dioxide
DIG	digoxigenin
dl	decilitre
DMEM	Dulbecco's modified eagles medium
DNA	deoxyribonucleic acid
DT	definitive phage type
EBI	European Bioinformatics Institute
EDTA	ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
ESBL	extended spectrum beta-lactamase
g	gram
g	gravity
G	guanine
h	hours
HCL	hydrochloric acid
HIV	human immunodeficiency virus
kb	kilobases
kDa	kilodalton
L	litre
LB	Luria Bertani
LPS	lipopolysaccharides

M	molar
MAb	monoclonal antibody
Mb	megabases
MDR	multi-drug resistant
mg	milligram
Mg^{2+}	magnesium
min	minutes
MLST	multi-locus sequence typing
ml	millilitre
mM	millimolar
N	number
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	nanogram
nm	nanometre
nM	nanomoles
nt	nucleotide
NTS	non-typhoidal Salmonellae
OD	optical density
PAI	pathogenicity island
PBS	phosphate buffered saline
PBS-T	phosphate buffered saline with 0.05% Tween 20
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
pH	-log $[H^+]$
PT	phage type
RE	restriction enzyme
RKC	rabbit kidney cells
RLUH	Royal Liverpool University Hospital
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
ROD	region of difference
rpm	revolutions per minute
s	seconds
SDS	sodium dodecyl sulphate
SGI	*Salmonella* genomic island
SSC	saline sodium citrate
SPI	Salmonella pathogenicity island
ST	sequence type
T	thymidine
TBE	Tris/Borate/EDTA
Tris	trishydroxymethylaminomethane
TTS	type three secretion system

U	units
UV	ultra-violet
v	volts
v/v	volts per volume
w/v	weight per volume
WHO	World Health Organisation
X-Gal	5-bromo-4-chloride-3-indolyl-betaD-galactopyranoside
AML10	Amoxycllin 10µg
AMC30	Amoxycillin/clavulanic acid 30µg
CTX30	Cefotaxime 30µg
CN 10	Gentamycin 10µg
CIP 1	Ciprofloxacin 1µg
W 2.5	Trimethoprim 2.5µg
NA 30	Nalidixic acid 30µg
RL100	Sulphamethoxazole 100µg
C10	Chloramphenicol 10µg
TET30	Tetracyclin 30µg ?
CXM 5	Cefuroxime sodium 5µg
RD 2	Rifampicine 2µg
CAZ30	Ceftazidime 30µg
S 25	Streptomycin 25µg

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Supplementary data

CHAPTER 1: INTRODUCTION

1.1. Salmonellae - General description

Salmonellae are flagellated, non-spore forming, Gram-negative bacilli, belonging to the family Enterobacteriaceae. Among the Enterobacteriaceae *Salmonella* is closely related to *Escherichia coli* and *Shigella* spp. The taxonomy and nomenclature of the genus *Salmonella* has changed over the years and continues to evolve. Two species of *Salmonella* are currently recognized, *Salmonella bongori* and *Salmonella enterica*. For these two species, 50 different serogroups have been determined which can be further divided into over 2500 serovars (Popoff *et al.*, 2003). *Salmonella* serovars are determined by the Kauffman-White Scheme which lists antigenic formulae. *S. enterica* is divided into six subspecies (I - *enterica*, II - *salmae*, IIIa - *arizonae*, IIIb - *diarizonae*, IV - *houtane* and VI - *indica*) with a seventh having been described, based on multilocus enzyme electrophoresis (Boyd *et al.*, 1996). *S. bongori* was initially described as subspecies V but was later found to be sufficiently divergent to be assigned to its own species. Figure 1.1 summarizes *Salmonella* nomenclature and names some important serovars.

Out of convenience, a system has been adopted in which the species name is omitted and replaced by the serovar; therefore *Salmonella enterica* subspecies *enterica* serovar Typhimurium becomes *Salmonella* Typhimurium. This is technically incorrect but has been adopted by most of the literature, and will therefore be followed here (Hohmann, 2001).

Salmonella serovars are determined through three major types of antigens: the somatic (O) antigens, which are outer membrane lipopolysaccharides (LPS) causing the endotoxic properties of *Salmonella,* the flagellar (H) protein and a capsular polysaccharide (Vi) which is only present in some serovars, most importantly *Salmonella* Typhi and Paratyphi C. *Salmonella spp* are unique among the Enterobacteriaceae in that they possess two distinct H antigens which are co-ordinately expressed as either phase 1 or phase 2, and therefore only one type of H antigen is expressed at a time (Chiou *et al.*, 2006; Silverman *et al.*, 1979). They provide a useful epidemiological tool in the identification of a source of infection. Antigenic analysis is used in identifying the organism clinically and assigning it to one of the serogroups (A-I), each of which contains a great number of serovars.

The pathological role of a number of serovars belonging to subspecies II and VI is unknown. *Salmonella* disseminate in the environment (soil, plants and water) where they can survive for months (water) and years (soil), but do not appear to multiply significantly.

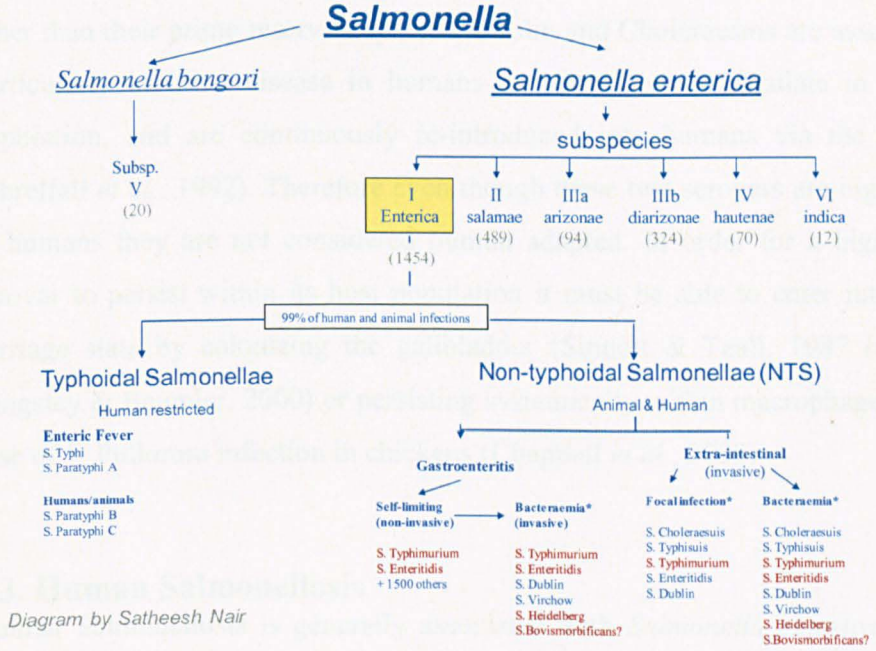


Figure 1.1 Schematic representation of *Salmonella* nomenclature. NTS serovars that this work focuses on are highlighted in red. (by Satheesh Nair, personal communication)

1.2 Host specificity

It is widely known that *Salmonella* serovars are host adapted, with some serovars such as *S.* Typhimurium having a broad host range and others such as *S.* Typhi demonstrating a narrow host range. These observations are largely based on epidemiological evidence. Studies have shown that certain serovars frequently cause illness in certain animal species (Barrow *et al.*, 1994; Gibson, 1961; Hinton, 1971): *S.* Dublin is associated with disease reported in cattle, while *S.* Choleraesuis is associated with porcine-illness. *S.* Typhimurium on the other hand is able to cause disease in a large number of species, including cattle, sheep, pigs, poultry and horses as well as humans. It is important to differentiate host adaptation and the ability of a pathogen to cause disease in an animal. *S.* Typhi, which causes typhoid fever in humans and higher primates but does not cause disease in other vertebrates, is often

cited as an example of host adaptation. However, *Salmonella* host adaptation and pathogenicity is more complicated when considering other serovars. Whilst *S.* Gallinarum and *S.* Pullorum are highly adapted to galliform birds causing typhoid-like disease, a number of serovars which are regarded as host specific (Chappell *et al.*, 2009), such as Dublin and Choleraesuis, are still able to cause disease in species other than their prime reservoir species. Dublin and Choleraesuis are associated with particularly invasive disease in humans but they do not circulate in the human population, and are continuously re-introduced into humans via the food chain (Threlfall *et al.*, 1992). Therefore even though these two serovars are highly virulent in humans they are not considered human adapted. In order for a highly adapted serovar to persist within its host population it must be able to enter into a chronic carriage state by colonizing the gallbladder (Sinnott & Teall, 1987 reviewed by Kingsley & Baumler, 2000) or persisting systemically within macrophages , as in the case of *S.* Pullorum infection in chickens (Chappell *et al.*, 2009)

1.3. Human Salmonellosis

Human Salmonellosis is generally associated with *Salmonella enterica* subspecies *enterica* (subspecies I), which comprises over 1400 serovars (Brenner *et al.*, 2000). These serovars are commonly subdivided into two groups according to their mode of infection: serovars causing enteric fever (typhoidal) and serovars causing gastrointestinal disesase, the non-typhoidal Salmonellae (NTS). *Salmonella* are also a recognized cause of bacteraemia and focal infections. The outcome of *Salmonella* infection depends on a number of factors including inoculation dose, immune status, underlying diseases of the host and genetic composition of the infecting bacterium.

In European countries, NTS outbreaks are predominantly caused by *Salmonella enterica* serovars Typhiumurium and Enteritidis and present as diarrhoeal disease caused by food poisoning. Foodstuffs such as poultry, eggs, beef, pork, milk, cheese, chocolate, fish and shellfish, vegetables and juices have been implicated as transmitting vehicles, whereby contamination can occur at multiple stages along the food chain. The Enter-net Surveillance network was introduced in 1994 (http://www.hpa.org.uk/AboutTheHPA/WhatTheHealthProtectionAgencyDoes/InternationalWork/EnterNet/). Ever since laboratory-confirmed human cases of *Salmonella* infection have been collected and reported by the 27 participating

countries. The number of reported cases peaked around 1997, and has since been in decline.

1.3.1. Enteric Fever

There are five *Salmonella* serovars commonly associated with fever: *S.* Typhi, *S.* Paratyphi A, *S.* Paratyphi B, *S.* Paratyphi C and *S.* Sendai. These *Salmonella* serovars are exclusively human pathogens and the cause of enteric (typhoid and paratyphoid) fever and are rarely associated with gastroenteritis. Serovar Typhi on its own, and serovars Paratyphi A and Sendai together, form genetically homogenous groups, whilst serovars Paratyphi B and Paratyphi C are genetically heterogeneous (Selander *et al.*, 1990). Spread is from person to person via contaminated food or water, with an incubation period of 10 to 14 days. Enteric fever is a systemic invasive disease. Symptoms are non-specific and can include malaise, body aches and respiratory symptoms, resembling a flu-like illness. Diarrhoea and constipation are equally possible. An uncomplicated illness will last for 4-6 weeks without treatment; however there are a multitude of possible complications which can involve gastrointestinal bleeding and perforation, toxaemia and encephalopathy. The case fatality rate was 10-20% in untreated cases (Woodward *et al.*, 1948; Woodward, 1949). However, particularly in children and severe hospitalised cases, mortality can rise to 30-50% (Hoffman *et al.*, 1984; Punjabi *et al.*, 1988; Rogerson *et al.*, 1991). Case fatality can be reduced to <1 per cent with prompt treatment with appropriate modern antibiotics (Parry *et al.*, 2002). 1-3 per cent of patients will become chronic carriers, which can cause continuous outbreaks in areas with poor public health and sewage disposal and no clean water supply (Ohl & Miller, 2001). Much of our understanding of typhoid fever is derived from systemic *S.* Typhimurium infections in mouse models (McClelland *et al.*, 2001). Initial gastrointestinal infection causes a short lived gastroenteritis. In the small intestine the bacteria penetrate the gut mucosa; M cells, specialized epithelial cells overlying the Peyer's patches are probably the site of internalisation. From there the bacteria translocate to the intestinal lymphoid tissue, and some will be passed on to the reticuloendothelial cells of the liver and spleen.

Salmonella organisms are able to survive and multiply within mononuclear phagocytic cells. During the bacteraemic phase, bacteria are released from their intracellular environment into the blood stream and disseminate widely. Common

sites of secondary infection are the spleen, liver, bone marrow, gallbladder and the Peyer's patches of the terminal ileum (House *et al.*, 2001). Bacteraemia caused by non-typhoidal *Salmonella* serovars are due to a typical pyogenic host response that is characterized by exudative intestinal inflammation, neutrophilia and septic shock during bacteraemia (Tsolis *et al.*, 2008). LPS contributes to mortality by eliciting a rapid TLR-4 dependent production of TNF-α (Engelberts *et al.*, 1991; Wilson *et al.*, 2008). In turn, TNF-α in combination with IL-1 β and IFN increases production of nitric oxide by inducing expression of nitric oxide synthase. Nitric oxide is a powerful vasodilator contributing to hypotension (Kilbourn & Belloni, 1990; Petros *et al.*, 1991). During sepsis fibrin is deposited in the microvasculature, resulting in intravascular coagulopathy that can result in organ failure (Waage *et al.*, 1991). The coagulation system is activated by an increase in tissue factor mediated by TNF-α-dependent increases in tissue factor expression on monocytes (Carlsen *et al.*, 1988).
The host response elicited by typhoidal serovars differs considerably from that, S. Typhi employs virulence factors that prevent classical antibacterial host responses, making it atypical. S. Typhi evades the TLR-4 immune response by expressing the Vi-capsule antigen. Vi-antigen markedly reduced expression of TLR-4 dependent production of TNF-α in a mouse sepsis model (Wilson *et al.*, 2008). It may accomplish this by physically masking or preventing access to the antigenic surface structures such as LPS (Tsolis *et al.*, 2008). The development of a vaccine is desirable; the Vi capsular polysaccharide antigen is the major focus of *S.* Typhi vaccine research (Hale *et al.*, 2006). Vi-negative strains of *S.* Typhi appear to be less virulent then Vi-positive strains (Parry *et al.*, 2002).

1.3.2. Gastroenteritis/Salmonellosis

Non-typhoidal Salmonellae (NTS) are important pathogens, causing four clinical syndromes in humans: diarrhoeal disease, invasive bacteraemic illness, focal infections and asymptomatic carriage. All *Salmonella* infections initially start with ingestion of the pathogen. Transmission to humans is strongly associated with agricultural products. Water-borne transmission is less common but does occur and person-to-person spread results from sharing contaminated food with an infected person. The incubation period of food-borne infections ranges from 12 to 48 hours. Symptoms of gastrointestinal illness caused by NTS infection include diarrhoea, chills, abdominal cramps, fever and head and body aches as well as nausea and

vomiting (www.hpa.org). These infections are usually self-limiting and antimicrobial treatment is not recommended. In an immunocompetent host the untreated case fatality rate is 0.1%. NTS diarrhoeal disease is often described as gastroenteritis; in fact NTS predominantly causes colitis with very minor small bowel involvement (Dr M. Gordon, personal communication). Complications can occur in the young and in the elderly, where severe dehydration can become life-threatening (Trevejo *et al.*, 2003). Bloodstream infections are another severe complication. Bacteraemias can be divided into primary and secondary bacteraemia. Primary bacteraemias principally occur in immunocompromised and HIV patients (Grant *et al*, 1997) but are also the major cause of morbidity and mortality in infants in Sub-Saharan Africa (Peters *et al*, 2004). Secondary bacteraemias on the other hand are a complication in 5-10% of gastroenteritis cases; both require urgent antibiotic treatment with fluoroquinolones or third generation cephalosporins (Dr M. Gordon and Dr C. Parry, personal communication).

Some of the risk factors for salmonellosis include the extremes of age, achlorhydria, atrophic gastritis, alteration of the endogenous bowel flora of the intestine (e.g. as a result of antimicrobial therapy or surgery), diabetes, malignancy, rheumatological disorders, reticuloendothelial blockade, HIV infection, and therapeutic immunosuppression of all types (Giannella *et al.*, 1971; Hohmann, 2001; Kunz & Waddell, 1956).

In European countries, NTS outbreaks are predominantly caused by *S. enterica* serovar Typhiumurium and serovar Enteritidis and present as diarrhoeal disease caused by food poisoning. Generally *S.* Typhimurium infections are more closely associated with contaminated pig, poultry or bovine meat, while *S.* Enteritidis is more closely associated with contaminated egg and poultry meat (http://www.efsa.europa.eu/en/scdocs/doc/1496.pdf). A report by Fisher from 1998 to 2003 showed that in the 24 countries, which are mostly European, participating in the study, *S.* Typhimurium and Enteritidis are still the most common isolates. There was a slight rise in *S.* Typhimurium cases in 2001; however the total number of *Salmonella* cases is still in decline. Over this six year period the number of Enteritidis cases fell by 36.2%, the number of Typhimurium cases fell by 26.6% and the number of other isolates such as *S.* Heidelberg fell by 35.3% (Fisher, 2004). The ESFA report from 2004-2008 registered a total of 131468 human confirmed *Salmonella* cases from 27 European countries, overall the number of reported cases

has been falling steadily from 2004 across the EU. Austria and Slovenia showed the greatest average decline of cases per year, while Cyprus and Estonia reported the highest average rise of cases per year. Reports generally peak within summer and autumn and decline in winter, in keeping with the consumption of fresh food and methods of cooking. The seasonal peak of *S.* Enteritidis was much more pronounced than that of *S.* Typhimurium. The foodstuffs most commonly implicated were broiler, pig and bovine meat, as well as fish, eggs and fruit and vegetables (http://www.efsa.europa.eu/en/scdocs/doc/1496.pdf).

A number of NTS serovars appear to have a greater tendency to cause invasive disease. These include serovars Virchow, Enteritidis, Choleraeseuis, Hadar, Dublin and Heidelberg (Saphra & Winter, 1957; Threlfall *et al.*, 1992; Vugia *et al.*, 2004; Weinberger *et al.*, 2004). Focal infections due to NTS appear to be associated with severe underlying disease and lead to increased mortality (Galofre *et al.*, 1994).

1.3.3. NTS in HIV infected patients

In 2008 an estimated 33.4 million people worldwide were living with HIV infections, 22.4 million of which in sub-Saharan Africa. A total of 2 million AIDS-related deaths were reported worldwide, with 1.4 million occurring in sub-Saharan Africa (UNAIDS www.unaids.org/en/KnowledgeCentre/ HIVData/GlobalReport/ 2008/). The incidence of NTS bacteraemia is 20 to 100-fold higher in HIV- infected patients than in HIV-uninfected patients in both the developed and developing world (Gordon, 2008; Hickey & Shanson, 1993; Meyer *et al.*, 1994). Low CD4 counts appear to be responsible and mortality in these patients is high, if they do not have access to appropriate antimicrobial treatment (Gordon *et al.*, 2002; Thamlikitkul *et al.*, 1996; Vugia *et al.*, 1993). Recurrent NTS bacteraemia infection is considered as one of several AIDS-defining bacterial infections (CDC, 1985). During the pre-HAART (Highly Active Anti-Retroviral Therapy) era up to 43% of patients experienced recurrent episodes of NTS bacteraemia. Hung and co-workers found that the incidence of NTS bacteraemia can be reduced by 96% in patients who responded to HAART treatment (Hung *et al.*, 2007).

1.3.4. NTS infection in Sub-Saharan Africa

In contrast to the situation in industrialized countries, diseases caused by NTS are one of the leading causes of morbidity and mortality in Sub-Saharan Africa,

1.3.4.1. NTS in immunocompromised adults

HIV-related infections in Africa differ from those in the developed world. Whereas opportunistic parasitic infections and fungal infections are commonly seen in the developed world, in Africa bacterial infections and tuberculosis dominate the clinical picture. A study from Uganda showed *Salmonella* spp. to be among the commonest identifiable causes of diarrhoeal disease in HIV-infected patients (Lule *et al.*, 2009) and NTS are the commonest bacteraemia isolates in African countries with high HIV-prevalence (Gilks *et al.*, 1990; Gordon *et al.*, 2001; Gordon *et al.*, 2002; Vugia *et al.*, 1993)

Febrile illness is the most common symptom on presentation to a Malawian hospital where HIV-seroprevalence is 72%. Reports from other African countries showed that the rate of malaria as a cause of fever in adults is often overestimated where microbiological facilities are lacking (Dougle *et al.*, 1997; Gordon *et al.*, 2001; Petit *et al.*, 1995). Table 1.1 summarizes studies of NTS bacteraemia in sub-Saharan Africa.

1.3.4.2. NTS in children

In young children NTS are the commonest cause of bacteraemia, the second commonest cause of neonatal meningitis and an important cause of septic arthritis (Berkley *et al.*, 2005). In some sub-Saharan African communities the burden of mortality due to paediatric NTS bacteraemia may be greater than that of malaria (Berkley *et al.*, 2005).

Clinical diagnosis of invasive NTS infections in African children is difficult without suitable microbiological facilities, as the clinical factors overlap with those of other conditions. Extraintestinal NTS infections are difficult to distinguish from malaria and are virtually indistinguishable from *S.* Typhi infections in terms of clinical features (Green & Cheesbrough, 1993). Cheesbrough and co-workers defined invasive *Salmonella* infection by: (1) illness requiring hospital admission in the opinion of an experienced paediatrician, (2) a history of fever for more than 5 days, (3) no focus of infection on clinical examination, and (4) negative or only scanty positive thick film for malarial parasites (Cheesbrough *et al.*, 1997). It is notable that few children present with *S.* Typhi infections. A study by Brent *et al* (1996) in Kenya showed that during the 4 years of their study not one of the 16750 children admitted

to the hospital had *S.* Typhi bacteraemia. Other studies in sub-Saharan Africa confirmed that the overwhelming majority of *Salmonella* bacteraemias are caused by NTS (Brent *et al.*, 2006) (Table 1.1). Case-fatality rates for bacteraemia in children in Africa have been estimated to lie between 4.4 and 27% (Brent *et al.*, 2006; Enwere *et al.*, 2006; Graham *et al.*, 2000; Walsh *et al.*, 2000). The fatality rate for NTS meningitis is likely to be higher than for any other bacterial pathogen; in one study in Malawi 64% of neonates with NTS meningitis died, compared to 26% of group B Streptococcal meningitis (Milledge *et al.*, 2005).

Table 1.1 A summary of important NTS bacteraemia studies in Africa

Country	Year	Reference	no of specimens	HIV prevalence	bacteraemia specimens	% NTS bacteraemia	adult/ children	Comments
Central African Republic	2001 2003	Bahwere et al, 2001 Kassa-Kelembho et al, 2003	779 131	? 85%	124 49	73% 45%	C A	high amp and gent resistance mostly S. Typhimurium
Cote d'Ivoire	1995 1993 2001	Lee et al, 1995 Vugia et al, 1993 Attia et al, 2001	264 319 270	6% ? 100%	36 43 9	64% ? 37%	C A A	all NTS bacteraemias HIV neg all patients enrolled were HIV-pos
Gambia	1999 2006 1994 1987	Mulholland et al, 1999 Enwere et al, 2006 O'Dempsey et al, 1994 Mabey et al , 1987	697 7369 1162 ?	? low	38 355 186 247	13% 27% 10% 29%	C C C C	64 infants died - 6% *S.* Typhi 18% *S.* Typhi
Kenya	1990 2005 2006 2009	Gilks et al, 1990 Berkley et al, 2005 Brent et al, 2006 Wiliams et al, 2009	506 19339 16570 38441	19% ? 18% ?	51 1094 2157	24% 166 (100%) 10%	A C C C	*S.* Typhimurium most common only tested for NTS, *S.* Enteritidis (52%), 6% also had confirmed meningitis 0.4% *S.* Typhi
Malawi	2007 2000 2001	Bronzan et al, 2007 2123 Walsh et al, 2000	1388 ? 238	? ? 73%	64 365 68	58% 38% 41%	C C A	2% *S.* Typhi 4% *S.* Typhi 2 cases of *S.*Typhi

	2004 2001	Bell et al, 2001 Peters et al, 2004 Gordon et al, 2001	352 2789	83% 92% in NTS	128 449	36% 36.5%	A A	1 case of S. Typhi 2.7% *S.*Typhi
Mozambique	2009	Sigauque et al, 2009	19896	?	1550	26%	C	0.2% *S.*Typhi
Nigeria	2009	Falade et al, 2009	330	?	95	16%	C	no *S.* Typhi
Rwanda	1987	Lepage et al, 1987	14032	?	112	32%	C	42% *S.*Typhi
Tanzania	2010 1998	Mtove et al, 2010 Archibald et al, 1998	1502 517	~7% 55%	156 145	29% 19%	C A	9% *S.* Typhi infections in older children Mycobacterium commonest isolate
Uganda	2006 1998	Bachou et al, 2006 Ssali et al, 1998	445 299	? 76%	76 75	37% 17%	C A	7% *S.* Typhi Mycobacterium and *S.* pneumonae commonest
Zaire (DRC)	2001 1997 1993	Bahwere et al, 2001 Cheesbrough et al, 1997 Green et al, 1993	779 120 ?	? ? ?	124 55 206	43% 63% 84%	C C C	2% 20% 17%
Zimbabwe	1996	Nathoo et al, 1996	309	54%	95	10.5%	C	Gram positive bacteria more common

1.3.4.3. Risk factors

Risk factors for NTS bacteraemias differ between Sub-Saharan Africa and other parts of the world. A long list of risk factors has been identified for invasive NTS. In the developed world the main risk factors for NTS bacteraemias include old age, often with underlying disease such as liver cirrhosis, systemic lupus erythematosus, immunodeficiency, atherosclerosis and solid organ cancers (Hsu & Lin, 2005).

In sub-Saharan Africa it is universally accepted that the high risk groups for invasive NTS infection are young children and HIV-infected adults. Most studies focus on these groups, which makes it difficult to assess the actual incidence of invasive NTS disease in Africa. It is likely to vary by population prevalence of HIV-infection, age distribution and local conditions, but is likely to be higher than that of typhoid fever caused by *S.* Typhi infection, which in turn is estimated to be 50 cases per 100,000 person years. Clinical studies appear to suggest that typhoid infections are less common than invasive NTS disease (Crump *et al.*, 2004; Mweu & English, 2008).

HIV is a recognized risk factor for invasive NTS infection in the developed world and among African adults. NTS are the commonest blood culture isolates from adults in countries where HIV prevalence is high. Mortality of NTS bacteraemia among HIV positive adults can be as high as 80% (Gordon *et al.*, 2001). However, the majority of African children with NTS bacteraemia are HIV negative. It is thought that HIV is particularly associated with relapses after appropriate treatment (Gordon *et al.*, 2002). This was not observed for the majority of children with NTS bacteraemia (Brent *et al.*, 2006).

Paediatric NTS bacteraemia is associated with malaria, severe anaemia and malnutrition. Among children under 5 years of age, NTS are the commonest or second commonest cause of bacteraemias in all Sub-Saharan African countries where these have been studied (Berkley *et al.*, 2005; Gordon & Graham, 2008; Graham *et al.*, 2000; Kariuki *et al.*, 2006a) (see Table 1.1 for further references). Death from invasive NTS infection was significantly associated with age under six months (Green and Cheesbrough, 1993). The mean age for NTS bacteraemia in Malawi was 22 months (M. Gordon personal communication). Severe Anaemia (haemoglobin of < 5g/decilitre) or malarial anaemia follows a seasonality pattern, related to that of NTS in HIV-infected adults. Several studies have shown seasonal peaks of NTS disease coinciding with the rainy season (Brent *et al.*, 2006; Gordon *et al.*, 2008; Kariuki *et al.*, 2006a; Milledge *et al.*, 2005). Enterobacteriacae concentrations in

drinking water are highest during this time, suggesting waterborne infection of NTS is high during the wet season (Wright, 1986). Gordon and co-workers showed a clear relationship between rainfall and NTS incidence for both adults and children (Gordon *et al.*, 2008). The association of malnutrition and increased bacterial and parasitic infections is well established. Increased mortality coincides with the rainy season when malnutrition peaks (Bell *et al.*, 2001). It is difficult to prise the individual risk factors apart and they probably contribute to each other.

The majority of children diagnosed with NTS bacteraemia present with a history of fever, diarrhoea and illness for over 7 days. Children with NTS bacteraemia and without fever show severe signs of malnutrition and/or hypothermia (Brent *et al.*, 2006). All of the above mentioned risk factors play an important role in invasive NTS infection, but it is important to note that they are not consistently present (Berkley *et al.*, 2005).

Malaria is endemic in most parts of Sub-Saharan Africa, where 90% of malaria related deaths occur in children younger than 5 years of age. The association between malaria and NTS infection has long been recognized. In one study it was demonstrated that children with *falciparum* malaria and severe malarial anaemia are particularly prone to NTS bacteraemias (Figure 1.2.) (Brent *et al* 2006). Seventy five percent of NTS patients with anaemia showed signs of either current or recent malaria. The age prevalence of NTS bacteraemia corresponded with that of severe anaemia at 6 to 12 month. Similarly seasonal variation of NTS bacteraemia coincided with the prevalence of both severe anaemia and malaria. However, severe malnutrition is also increased at these times and therefore the causal relationship between malarial anaemia and NTS bacteraemia is not fully proven. Epidemiological data are consistent with the theory that malaria predisposes to invasive NTS infection by causing haemolysis, which will lead to increased erythrocytosis and thereby interferes with macrophage function (Brent *et al* 2006). The exact mechanism by which malaria facilitates systemic NTS infection is not well understood. It is thought that severe malarial anaemia leads to a specific defect in the immune system, which results in increased susceptibility to bacteraemia. *Salmonella* in particular, are well adapted to long term survival inside monocytes and macrophages. It is possible that latent Salmonellae are released into the bloodstream when macrophage function is disturbed (Graham *et al*, 2000). A study in Malaysia noted that children undergoing immunosuppressive treatments developed NTS bacteraemia without any preceding

gastroenteritis (Lee *et al*, 1999). Deficiency of the Th-1 cytokines such as interleukin-12 or interferon-γ, which would normally induce immune responses that control intracellular pathogens such as Salmonellae, further facilitates infection (MacLennan *et al.*, 2004).

It was suggested that increased intestinal helminth infection in HIV-infected patients may play a role in promoting NTS bacteraemia due to a bias towards Th2-mediated immune responses induced by the infection (Dowling *et al.*, 2002). There is some evidence that intestinal parasites that penetrate the gut such as *Ascaris lumbricoides* and *Strongyloides stercoralis* can carry bacteria into the blood stream when they pass through their life cycle from the host gut into the blood stream. It appears that enteric bacteria can be carried both on the surface and in the gut of the disseminated larvae (Igra-Siegman *et al.*, 1981). Invasive filariform larvae do not feed, but the rhabditiform larvae, earlier in the life cycle feed in the host gut whereby enteric bacteria can be ingested and then released at a later stage in the host blood stream (Archibald *et al.*, 1989). Invasive *Salmonella* infection often persists until helminths have been cleared by anti-schistosomal treatment such as niridazole or praziquantel (Gendrel *et al.*, 1994; Neves *et al.*, 1969). Figure 1.2 and Table 1.2 summarize the importance of the risk factors described above.

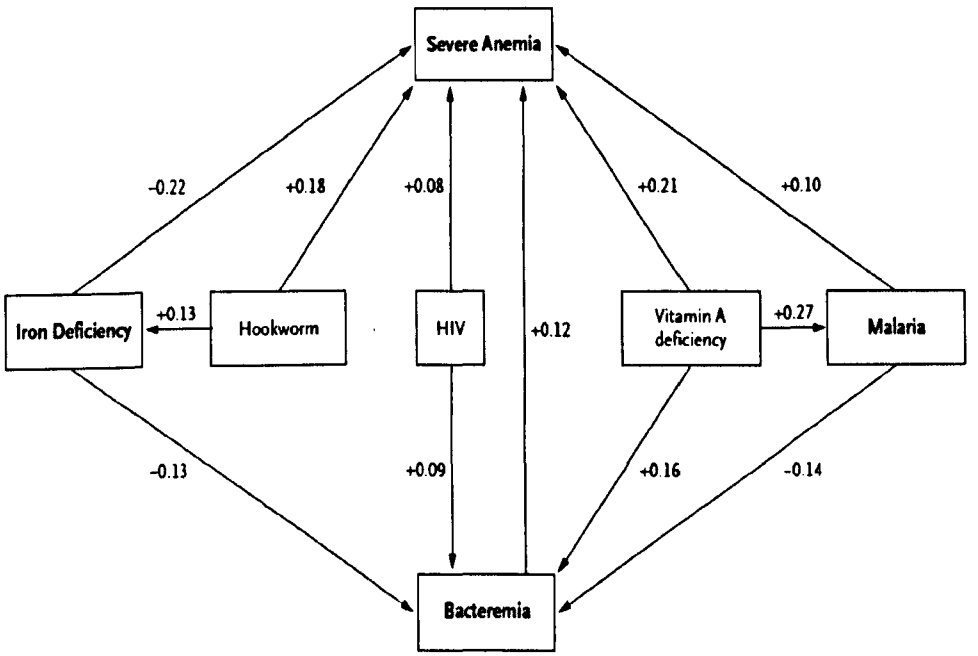


Figure 1.2. Exploratory model of factors associated with severe anaemia in Malawi. The sizes of the associations are indicated by the standardized regression coefficients (range, -1.0 to +1.0). Inverse (protective) associations are indicated by red lines. This model was created containing all possible associations between the displayed variables, after which all nonsignificant arrows (P≥0.05) were removed. The displayed variables were all adjusted for age; in addition, malaria was adjusted for previous use of antimalarial agents, and iron deficiency was adjusted for a history of transfusions (before the previous 4 weeks) or use of hematinic agents (within the previous 8 weeks). Replacement of severe anaemia by continuous hemoglobin levels and iron deficiency (ratio of soluble transferrin receptor to log ferritin, >5.6) The overall model fit was valid (root mean square area of approximation, 0.043; 95% confidence interval, 0.039 to 0.048). Diagram adapted from (Calis *et al.*, 2008).

Table 1.2. A summary of the importance of commonly accepted risk factors for NTS bacteraemia in Africa in children and adults

Risk Factor	*Grade of Evidence** children <3yrs	children >3yrs and adults
Environment		
Food and Water	B	B
Hospital-acquired infection	B	B
direct and indirect animal contact	C	C
transmission between humans	C	D
Host		
Age	A	D
HIV infection	A	A
Malnutrition	A	-
Sickle cell disease	-	C
Malarial anaemia	B	-
Schistosomiasis	-	D
Recent antimicrobial use	B	C

**Levels of evidence were estimated from African studies by (Morpeth et al., 2009) and were classified according to the grades of the Oxford Centre for Evidence-based Medicine (www.cebm.net/index.aspx?o=1001)*
- lack of data

1.3.4.4. Multidrug resistance in NTS

In the developing world and Africa in particular, multidrug resistance (MDR) to commonly available antibiotics remains a major challenge to the healthcare system. MDR NTS have been the cause of life threatening invasive disease in children in many African countries such as Zaire (now DRC) (Cheesbrough *et al.*, 1997; Green & Cheesbrough, 1993), Malawi (Graham, 2002; Milledge *et al.*, 2005), Rwanda (Lepage *et al.*, 1987), Nigeria (Adejuyigbe *et al.*, 2004), Tanzania (Vaagland *et al.*, 2004) and the Central African Republic (Kassa-Kelembho *et al.*, 2003).

Bachou and co-workers investigated the occurrence of bacteraemia in malnourished children in Uganda and found a strong correlation between previous use of antibiotics and MDR NTS bacteraemia. Malnourished children under five years of age were particularly at risk of developing bacteraemia, with the most common cause being NTS resistant to commonly available antibiotics such as ampicillin, chloramphenicol and co-trimoxazole (Bachou *et al.*, 2006). Second-line antibiotics such as ciprofloxacin and ceftriaxone are not readily available and are often too expensive; furthermore, resistance is also emerging to extended-spectrum cephalosporins. These findings are consistent with studies from Malawi (Graham *et al.*, 2000), Kenya (Kariuki *et al.*, 2005) and Ethiopia (Shimeles & Lulseged, 1994).

Chloramphenicol is the drug of choice for Malawian children; whereas *S.* Typhimurium remains susceptible the sudden appearance of chloramphenicol resistant *S.* Enteritidis is worrying. Interestingly, *S.* Enteritidis has become a more common cause of bacteraemia since it has developed resistance. Chloramphenicol resistance in Malawi increased from 0% in January 1999 to over 80% in December of the same year (Graham, 2002). During this period, the prevalence of *S.* Enteritidis as a blood culture isolate in children rose from 13 to 37% (Graham, 2002).

A twelve year surveillance study at a Kenyan referral hospital showed that *S.* Typhimurium and *S.* Enteritidis accounted for 70.8% of all NTS isolates and that this distribution remained stable over a decade. A large proportion of these NTS were multiply resistant to several commonly available antibiotics including ampicillin, streptomycin, cotrimoxazole, chloramphenicol and tetracycline (Kariuki *et al*, 2005), but remained susceptible to cefotaxime and ciprofloxacin.

Over the counter availability of antibiotics is common in Africa and reliable data on use of antibiotics without prescription is sparse; it is common practise for parents in urban settings to treat acute respiratory infections and childhood diarrhoea with over the counter antibiotics. The most commonly used antibiotics for treatment of NTS infections were cefuroxime or a combination of ampicillin and gentamicin. Even though it has been established that antimicrobial treatment of bacterial diarrhoea is not beneficial and may prolong NTS carriage, it is common practise to prescribe antibiotics in most cases of gastroenteritis, especially in children, which leads to immense antibiotic misuse. Between 1994 and 1997 48% of clinical NTS isolates were resistant to three or more antibiotics; between 1997 and 2000 the level of resistance rose to 64% of MDR isolates; between 2001 and 2006 72%, were MDR. (Kariuki *et al.*, 2006a; Kariuki *et al.*, 2006b) (personal communication Prof C.A. Hart). Gastrointestinal infection with MDR NTS has also been associated with a higher risk of developing invasive disease. The widespread and increasing presence of pentaresistant *S.* Typhimurium Definitive Phage Type (DT) 104 is also alarming (Kariuki *et al*, 2005). Since the 1990s *S.* Typhimurium DT104 has been recognized as the most common *S.* Typhimurium clone in many European countries, including Germany, Austria, the Netherlands and the United Kingdom (Prager *et al*, 1999).

1.3.4.5. Routes of transmission and Reservoirs of NTS infections

In industrialized countries, most antimicrobial resistant *Salmonella* infections are acquired through contaminated food of animal origin (White *et al.*, 2004). Animal husbandry routinely uses antimicrobial agents to treat and prevent infection as well as for growth promotion. Not surprisingly this practice promotes antimicrobial resistance in bacteria, which are then passed on to humans through the food chain (Graham *et al.*, 2007). However in African countries such as Kenya human-to-human transmission appears to play a much greater role than animal to human transmission. Using antimicrobial susceptibility profiling and plasmid and genomic DNA typing Kariuki and collegues showed that there was no significant association between NTS isolates from humans and those from animals living in close proximity, food or from the environment (Kariuki *et al.*, 2002). The study showed multiple resistance patterns to all the common antibiotics amongst human NTS isolates, while NTS from animals and the environment, including the homes of patients, were fully susceptible. Pulsed-field gel electrophoresis was used to demonstrate that the NTS genomic DNA differed between human and chicken isolates, although it was possible to group NTS from chickens, cows and pigs together. Contaminated water is the most likely vehicle for human to human transmission in communities with unhygienic facilities and poor sanitation. This makes spread of NTS similar to that of typhoid fever (Gasem *et al.*, 2001).

1.4. *Salmonella* Genomes

In 2001, McClelland *et al* sequenced the 4,857 kb chromosome and the 94 kb virulence plasmid (pSLT) of *Salmonella* Typhiumurium strain LT2. They compared the LT2 genome to eight previously sequenced genomes of closely related strains/species including *Salmonella* Typhi and Paratyphi, *E. coli* K12 and O157 and *Klebsiella pneumoniae*, to determine the presence of sequences sharing similarity with *Salmonella* Typhiumurium open reading frames (ORF). This comparison enabled the detection of 204 pseudogenes in *S.* Typhi and 39 pseudogenes in *S.* Typhiumurium LT2 (McClelland *et al.*, 2001). Pseudogenes are coding sequences (CDS) that have been inactivated through mutations including frameshifts, nonsense substitutions, rearrangements, insertions or deletions. It is likely that through a combination of adaptations and genetic drift, pseudogene selection is a marker of host-restricted pathogenic bacteria compared to host-generalist bacteria. A

combination of adaptation and genetic drift associated with population bottlenecks during or after adaptation appears to be responsible for loss of gene function (Parkhill *et al.*, 2001).

Therefore the difference in pseudogene numbers may explain the host specificity of *S.* Typhi compared to the broad host range of *S.* Typhiumurium.

Genes found exclusively in *Salmonella* (marked in green in Fig 1.3.) appear to be closely linked to its pathogenicity as they encode proteins involved in host invasion and effector proteins which are excreted by a type III secretion system (TTSS), some of which will be discussed in more detail in another section. These genes have homologues in more distantly related Salmonellae such as *S. bongori*, which indicates their importance. A subset of these genes, including some genes in the fimbrial cluster, are found only in the closely related serovars Typhi and Paratyphi, and may therefore be involved in the specialisation of this group to warm-blooded host species (McClelland *et al.*, 2001).

Microarray-based comparative genomic hybridisation has since been used to monitor the genetic content of *Salmonella* subspecies I. Porwollik and co-workers (2004) investigated the genomes of 79 strains of prevalent serovars of human clinical and veterinary origin. They found that intraserovar variation, in the case of Paratyphi B Pb 7 , Muenster MeA1 and Infantis In3, can be greater at times than genetic variation between strains of different serovars, such as Choleraesuis and Paratyphi C. They have taken the approach of grouping strains that share a distinct pattern of gene content into genovars (Porwollik *et al.*, 2004).

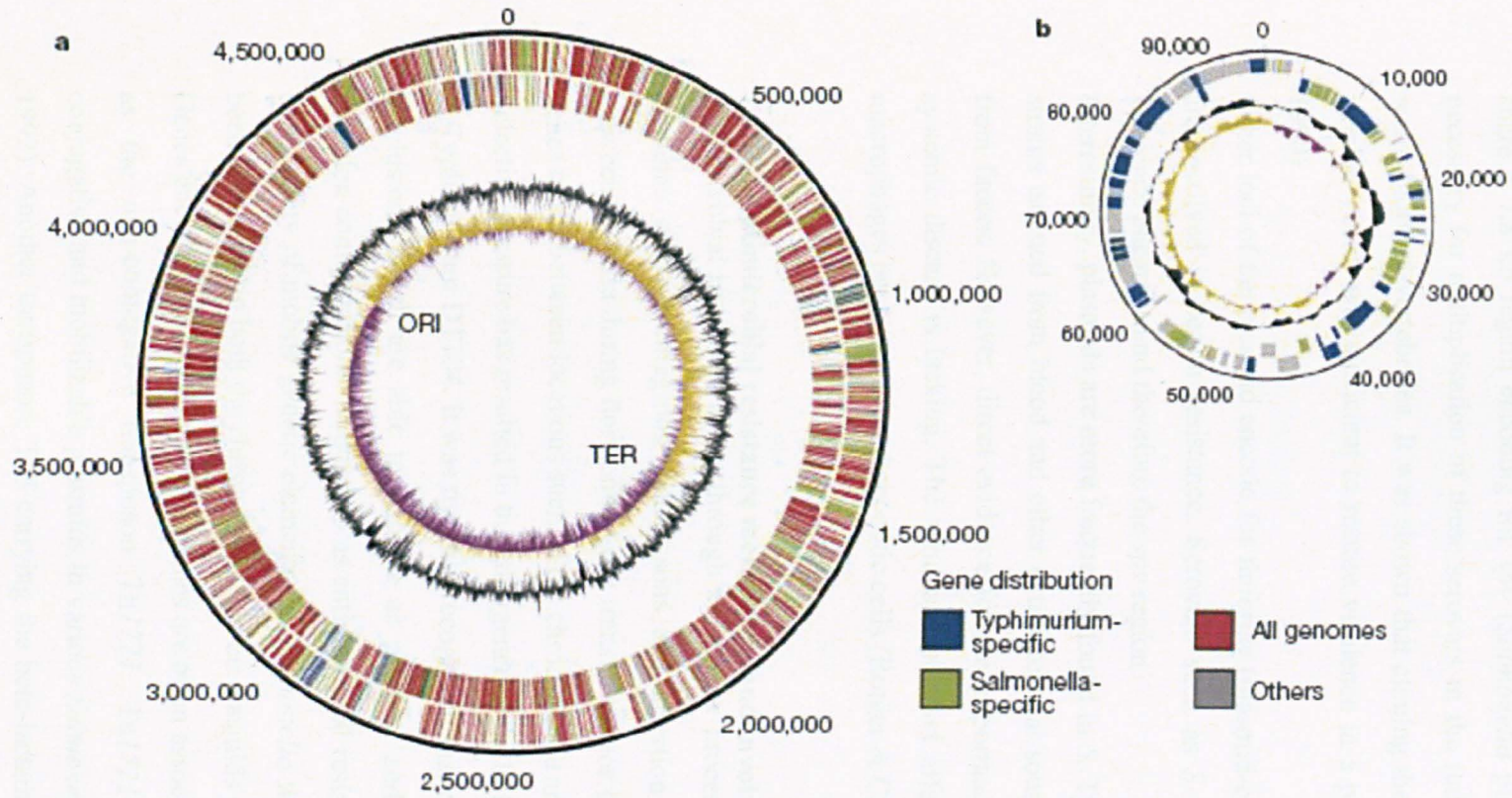


Figure 1.3. The *Salmonella* Typhimurium LT2 genome, **(a)** the chromosome, base pairs are indicated outside the outer circle. The two outer circles indicate the coding orientation, with the forward strand on the outside and the reverse strand on the inside. Red indicates close homologous in all genomes, while green indicates genes with close homologous in other *Salmonella* serovars, but not *E.coli* or *Klebsiella*. Blue indicates genes present only in *S*. Typhimurium LT2 and grey indicates other combinations. The black circle shows the GC content; the yellow/purple circle represents the GC bias. The position of the origin of replication (ORI) and the terminus (TER) are shown. **(b)** The plasmid pSLT, base pairs are indicated on the outside, the plasmid is not to scale, the same colouring scheme as for (a) applies (McClelland *et al*, 2001)

1.4.1. The *Salmonella* virulence plasmid

Salmonella serovars belonging to subspecies I carry a large virulence plasmid (Figure 1.3 b). Despite their low copy number (1-2 per chromosome) the plasmids are very stable. Their involvement in the enteric stage of infection is unclear. *Salmonella* virulence plasmids are heterogeneous in size (50-90 kb), but all of them share a 7.8 kb region encoding for *spv* (*Salmonella* plasmid virulence), which is necessary for multiplication of these serovars in the reticulo-endothelial system of warm-blooded vertebrates. It was shown that cloning the *spv* locus into a low copy number vector was sufficient to restore virulence in a plasmid-cured strain (Gulig, 1990).

Other loci of the plasmid encode for fimbriae (plasmid-encoded fimbriae = *pef*) and are involved in serum resistance. Serovars such as *S.* Typhi and Paratyphi lack virulence plasmids and therefore the *spv* region.

Interestingly, plasmids are more frequently found in *S.* Typhimurium and Enteritidis strains isolated from blood and other extraintestinal sources than in strains isolated from faeces. However, direct evidence for the importance of virulence plasmids in systemic disease is lacking. The virulence plasmid affects intracellular growth in macrophages but not in non-phagocytic cells (Rotger & Casadesus, 1999).

1.4.1.1. Antimicrobial resistance mechanism and involvement of plasmids

Antimicrobial resistance arises through a complex process, whereby mobile genetic elements such as integrons, transposons and insertion sequences are exchanged between bacteria during horizontal gene transfer and/or through translocation where genes move between locations such as the chromosome and a plasmid.

Selective pressure has resulted in the emergence of multidrug resistant strains such as *S.* Typhimurium DT104. It was recently recognized that resistance genes accumulate in clusters, which are able to relocate as physical and functional entities. These complex configurations, referred to as antimicrobial resistance islands, are made up of a variety of mobile genetic elements. In *Salmonella* these resistance islands have been identified on both the chromosome and on plasmids (Miriagou *et al.*, 2006).

Genes encoding resistance to tetracyclines are often associated with transposons such as the non-conjugative transposon Tn*1721*. Tn*1721* has been identified on conjugative and mobilizable plasmids in various *Salmonella* isolates (Allmeier *et al.*, 1992). Another transposon, Tn3 carrying the beta-lactam resistance gene TEM, has

also been identified in *Salmonella*. Some serovars carry a TN*3*ΔTn*1721* fusion element which may promote combined beta-lactam and tetracycline resistance (Pasquali *et al.*, 2005). The *strA-strB* genes, which mediate streptomycin resistance, are often located on plasmids as part of the transposon Tn*5393*. A particular derivative of Tn*5393* carrying the addition of the insertion sequence IS*1133* was first identified in plant pathogens but has found its way into several unrelated strains of *Salmonella enterica*. The insertion of IS*1133* results in increased expression of *strA-strB* (Miriagou *et al.*, 2006).

1.4.2. *Salmonella* Bacteriophages

In 1950 Boyd first recognized that *Salmonella* spp carried symbiotic bacteriophages (Boyd, 1950). Genes can be exchanged between bacteria through conjugation, transformation and transduction (Jain *et al.*, 2002). Bacteriophages that mediate transduction can be isolated from the environment including the intestine. Some bacteriophages can enter a lysogenic state and thereby contribute to the bacterial chromosome. Furthermore they can carry genes that are not essential for phage proliferation but alter the host pathogenicity (Campbell *et al.*, 1992). *S.* Typhimurium strains usually carry four to five prophages, while *S.* Typhi strains CT18 and Ty2 carry seven prophages (Thomson *et al.*, 2004). The *Salmonella* phages fall within five groups, namely P27-like, P2-like, lambdoid, P22-like and T7-like, and three outliers: ε15, KS7 and Felix O1. *Salmonella* only contains one representative of P-27-like (ST64B) and one of T7-like (SP6) phages. *Salmonella* subspecies I genomes contain three complete lambdoid-related prophages of the Siphoviridae family: Fels-1, Gifsy-1 and Gifsy-2 (Figueroa-Bossi *et al.*, 1997; Figueroa-Bossi *et al.*, 2001).

Fels-1 (41.7 kb) carries two potential virulence genes, *nanH* (neuroaminidas) and *sodC3* (superoxide dismutase) and integrates between host genes *ybjP* and STM0930 (Figueroa-Bossi *et al.*, 2001). Gifsy-1 (47.8 kb) integrates into the *lepA* gene that encodes ribosome-binding GTPase (Caldon *et al.*, 2001), Gifsy-1 possesses a number of virulence modulating genes including *gipA*, which is involved in colonization of the small intestine. Gifsy-2 (45.5 kb) is probably defective in LT2; it integrates between *pncB* (nicotinate Phosphoribosyl-transferase) and *pepN*. Two genes involved in host pathogenesis, *gtgA* and *sodC1* (periplasmic superoxide dismutase), are located on Gifsy-2. Deletion of *gtgA* and *sodC1* results in seven- and five-fold reduction in virulence respectively (Ho *et al.*, 2002). *Salmonella* carries three P2-like

phages: Fels-2, SopEΦ, PSP3, and five P22-like phages: ε34, ES18, P22, ST104, ST64T (see Table 1.3 for more details)

Table 1.3. A summary of *Salmonella* bacteriophages (Kropinski *et al.*, 2007)

phage groups	*Salmonella* phages	Genus	size (kb)	GC content (%)	integration site	virulence related genes
P22-like	ST104	Podoviridae	41.4	47.3	?	*gtr*
	ES18	Siphoviridae	46.9	48.6	?	?
	ST64T	Podoviridae	40.7	47.5	*attP*	?
	ε34	Podoviridae	43.0	47.3	*argU*	?
lambdoid	Fels-1	Siphoviridae	41.7	?	*ybjP* and *STM0930*	*nanH, sodC3*
	Gifsy-1	Siphoviridae	47.8	?	*lepA*	*gipA*
	Gifsy-2	Siphoviridae	45.5	?	*pncB* and *pepN*	*gtgA, sodC1*
P27-like	ST64B	Podoviridae	40.1	51.3	*serU*	*immC*
T7-like	SP6	Myoviridae	43.8	47.2	?	?
P2-like	PSP3	Myoviridae	30.6	52.8	?	?
	Fels-2	Myoviridae	33.7	52.5	*ssrA*	STM2730 (DAM methylase)
	SopEΦ	Myoviridae	34.7	51.3	?	*SopE*
outliers	KS7	Siphoviridae	40.8	?	?	?
	ε15	Podoviridae	39.7	50.8	?	?
	Felix O1	Myoviridae	86.2	39.0	?	?

1.5. Mechanism of invasion

1.5.1. Host cell invasion

Salmonella enterica is a facultative intracellular pathogen that can replicate inside epithelial cells and macrophages. The fact that the serovars *S.* Typhi and Paratyphi are exclusively human pathogens has led to the extensive use of *S.* Typhimurium in experimental models of disease, because of its ability to cause a typhoid fever-like illness in mice, making it the most widely studied serovar among the Salmonellae (Miller *et al.*, 1989).

In mice *Salmonellae* preferentially enter through M cells, which are specialized epithelial cells overlying lymphocyte-rich Peyer's patches (Donnenberg, 2000).

The main sites of replication during systemic infection are the spleen and the liver. *Salmonellae* will replicate within membrane-bound intracellular vacuoles inside macrophages.

1.5.2. *Salmonella* fimbriae

Fimbriae have been shown to be important in adhesion of pathogenic bacteria to the host tissue. The fimbriae of the Enterobacteriacae are broadly divided into mannose resistant and mannose sensitive on the basis whether mannose is able to inhibit adhesion of fimbriae to erythrocytes (haemagglutination) (Paranchych & Frost, 1988; Popiel & Turnbull, 1985; Salit & Gotschlich, 1977). At least four different types of fimbriae have been identified on *Salmonella* species: type 1, plasmid encoded, long polar and thin-aggregative (curli) fimbriae, all of which have been shown to mediate adhesion. Their role in invasion however remains to be elucidated. Fimbrial-mediated contact has been shown to induce the formation of appendage-like structure, called invasomes. Mutants lacking invasomes are able to initiate contact but are unable to invade the host epithelial cells (Ginocchio *et al.*, 1994). Following adhesion the microvilli on the surface of epithelial cells disappear and membrane ruffling occurs, whereby cytoskeletal changes inside the cell permit the cell membrane to stretch around and thereby engulf the bacterium into a large fluid-filled vacuole (macropinocytosis). This phagosome does not contain lytic enzymes and therefore the bacteria survive. Inside the phagosome, proliferation is accompanied by the formation of *Salmonella* induced filaments (Sifs), which attach to the phagosome and are thought to thereby provide the bacterium with nutrients (Knodler & Steele-Mortimer, 2003).

Salmonellae possess an abundance of different adhesive structures. Surprisingly little is known about their properties and role in virulence. Type 1 fimbriae (*fim*) and thin aggregative fimbriae (*agf*) seem to be present in most serovars, while the prevalence of others, such as the plasmid encoded (*pef*) and long polar fimbrial (*lpf*) operons, is limited to a few serovars. The *fim* and *agf* operons have also been studied in more detail in *S.* Typhimurium. Type 1 fimbriae are about 7 nm in diameter and have a hollow tubular structure made up of >1000 major subunit FimA protein arranged in a right hand helical turn. The minor subunit FimH is the receptor recognition element at the tip of the fimbriae. The FimH adhesin is responsible for *Salmonella* binding to HeLa, HEp-2 and mouse intestinal epithelial cells (Kisiela *et al.*, 2006), however FimH shows sequence variation in serovars Pullorum and Gallinarum, which prevents mannose-binding (Kisiela *et al.*, 2005)

The *pef* operon is located on the *S.* Typhimurium virulence plasmid (Friedrich *et al.*, 1993). Expression of *pef* is negatively controlled by histone-like protein (H-NS), the

stationary phase sigma factor (RpoS), and the presence of type 1 fimbrial biosynthesis genes (*fimAICDHF*), and expression can therefore not be detected under standard laboratory conditions (Humphries *et al.*, 2005; Humphries *et al.*, 2003). Expression of the *pef* operon in *E. coli* has shown a thin, flexible fibrillae, 2-5 nm in diameter, made up of PefA subunits (Baumler *et al.*, 1996a; Friedrich *et al.*, 1993).

The *agf/csg* operon (also known as SEF17) is the most ancient of the *Salmonella* fimbrial operons. It is likely that this fimbrial operon was present in the common ancestor of *E. coli* and *Salmonella*. *Agf* fimbriae are highly stable structures of a wiry appearance, made up of AgfA subunits (Collinson *et al.*, 1991; Collinson *et al.*, 1996).

Long polar fimbriae (*lpf*) have been implicated in murine cell invasion (Weening *et al.*, 2005) Figure 1.4 is a diagrammatic representation of the fimbrial operons, while Table 1.4 summarizes their most important properties.

Whole genome sequencing has identified 13 fimbrial operons in the serovar Typhimurium (Weening *et al.*, 2005). These include eight members of the γ-Fimbriae (*fim*, *bcf*, *sti*, *sth*, *lpf*, *saf*, *stc*, and *stb*), one member of the β-Fimbriae (*stj*), two members of the π-Fimbriae (*std* and *stf*), and one member of the κ-Fimbriae (*pef*). However, the binding specificity is known only for type 1 fimbriae, which are encoded by the *S.* Typhimurium *fim* operon.

Table 1.4. Summary of selected fimbriae and fimbrial genes

gene	Name	operon	location on Typhimurium chromosome	function	Reference
fim	type 1 fimbriae	*fimAICDHF*	centisome 15	attachment to variety of cell types	(Hancox *et al.*, 1997; Lindquist *et al.*, 1987)
pef	plasmid-encoded fimbriae	*pefABCDI*	pSLT	attachment to murine small intestine	(Baumler *et al.*, 1996a)
lpf	long polar fimbriae	*lpfABCDE*	centisome 80	attachment to M cells	(Baumler *et al.*, 1996b)
stf	*S.* Typhimurium fimbriae	*stfACDEFG*	centisome 5	may have a role in host range determination	(Emmerth *et al.*, 1999)
agf/csg	thin aggregative fimbriae (curli)	*agfBAC*	centisome 26	colony morphology, autoaggregation	(Collinson *et al.*, 1991)
saf	Salmonella atypical fimbriae	*safABCD*	centisome 7	unclear	(Folkesson *et al.*, 1999)
bcf	bovine colonization factor	*bcfABCDEFG*	?	shedding of *S.* Typhimurium and colonisation	(Weening *et al.*, 2005)

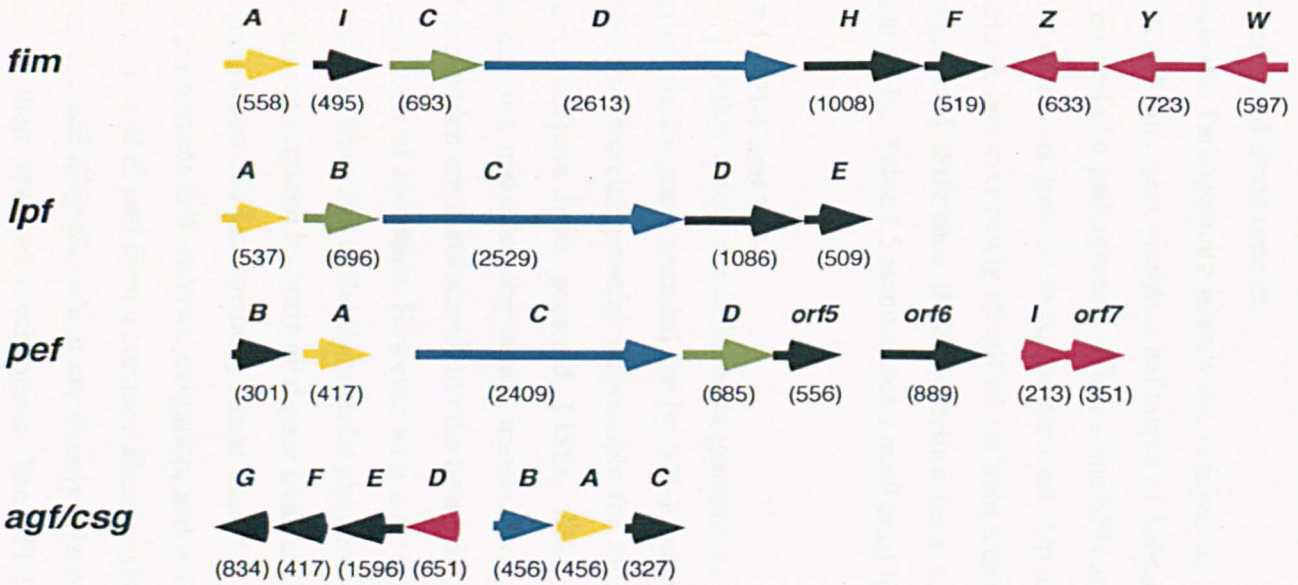


Figure 1.4. Organization of genes within four fimbrial operons of *S.* Typhimurium. Gene lengths in base pairs are indicated in brackets below each open reading frame. Genes encoding the major fimbrial subunits are indicated in yellow, genes encoding chaperones are indicated in green, genes encoding ushers are indicated in blue, genes encoding minor subunits are indicated in black, and genes encoding regulatory proteins are indicated in red (Darwin & Miller, 1999).

1.5.3. *Salmonella* Pathogenicity islands (SPIs)

Pathogenicity islands are present in a large number of pathogens. These mobile genetic elements have been described by Groisman as "quantum leaps in bacterial evolution" (Groisman & Ochman, 1996), enabling horizontal transfer of complex virulence functions between species. PAIs are often divergent in their base composition from the core genome and carry a number of virulence genes; they are found near an insertion site such as tRNA and are sometimes unstable, characteristics of which are DNA mobility elements such as transposases, bacteriophage genes, integrases and direct repeats.

Salmonella Pathogenicity islands are referred to as SPIs and are associated with a number of important virulence hallmarks of *Salmonella*, including host cell invasion and intracellular pathogenesis. While some SPIs are conserved across the *Salmonella* genus, others are specific to certain serovars. Up to 13 SPIs have been identified and described, however newly identified variable regions are now increasingly described as regions of difference (RODs), before their role in virulence has been proven conclusively. Table 1.5 summarizes a number of important SPIs.

1.5.3.1. SPI-1 and SPI-2

SPI-1 is stable within the *Salmonella* genome, as it is not located at a tRNA locus. Most of the 25 genes encoded for by SPI-I make up the Type 3 Secretion System (T3SS) and secreted proteins responsible for invasion of enterocytes. Many Gram negative bacteria have acquired T3SSs, which consist of conserved structural elements that make up the actual translocation apparatus, and multiple effector proteins, which are translocated into the host cell. SPI-I genes are further involved in the induction of apoptosis. However with a G+C content of 47%, the island differs significantly from that of the *Salmonella* chromosome (52%). This suggests that SPI-I was indeed acquired by horizontal gene transfer (Hensel, 2004).

The acquisition of a pathogenicity island encoding a type III secretion system (T3SS) was a main event in *S. enterica* evolution, and is thought to define the divergence of *Salmonella* and *E. coli* from a common ancestor (Baumler, 1997).

S. enterica and *Shigella,* which are closely related, share many similarities when it comes to their invasion mechanisms. The SPI-1 T3SS is similar to the *mxi/spa* invasion gene system of *Shigella* (Groisman & Ochman, 1993). Both organisms

stimulate extensive membrane ruffling in their host cells during entry and both require actin regulatory GTPases for entry.

The SPI-1 T3SS mediates contact-dependent translocation of a number of important effector proteins into the eukaryotic host cell. One subset of effectors is responsible for invasion and rearrangements of the cytoskeleton as well as extensive membrane ruffling. SipA, SipC, SopE and SopE2, SopB/SigD and SptP are collectively involved in bacterial uptake. Another subset is involved in enteropathogenesis, inflammation and symptoms associated with diarrhoea (Knodler & Steele-Mortimer, 2003).

In order to induce membrane ruffling, *Salmonella* effectors target the actin-regulatory GTPases Cdc42 and Rac1 (Hardt *et al.*, 1998). *Salmonella* outer protein E (SopE) is a guanine nucleotide exchange factor that is inserted into the host cell via the SPI-1 T3SS. SopE was part of the genome of a bacteriophage that infected strains *S. enterica* serovars Dublin, Typhi and some strains of Typhimurium (Mirold *et al.*, 1999). Once inside the cytoplasm SopE binds Rac and Cdc42, which results in the activation of signalling proteins and thereby the formation of membrane ruffles.

SptP (*Salmonella* protein tyrosine phosphatase) antagonizes the function of SopE. The composition of Spt is modular, with two distinct domains. The carboxyl terminus is homologous to eukaryotic tyrosine phosphatase, while the amino-terminal end acts as a GTPase Activating Protein (GAP) for Rho proteins. Spt stimulates the release of GTP, returning Cdc42 and Rac1 to their inactive form, and thereby returning the host cell to its original state. It is presumed that this benefits both the bacterium and host cell (Fu and Galan, 1999) (Figure 1.5.).

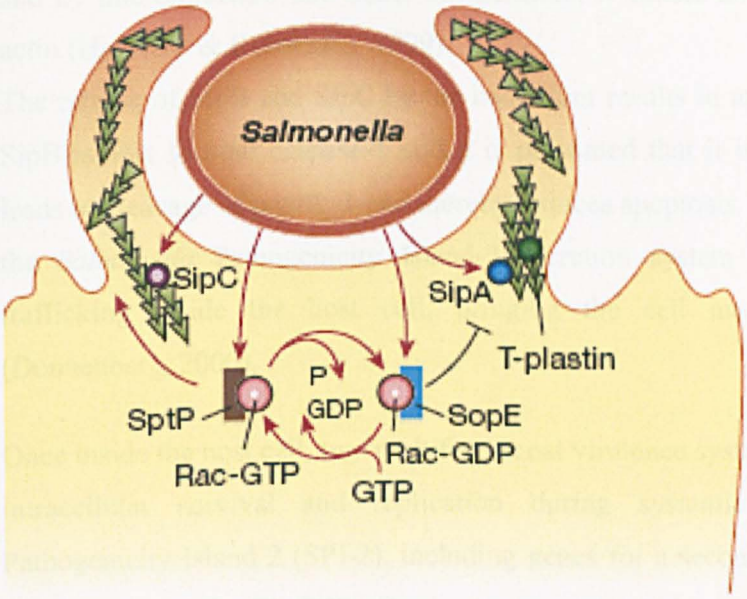


Figure 1.5. *Salmonella* targets the cytoskeleton of the host cell to induce its uptake. It requires the small actin-regulatory GTPases Rac and Cdc42 for its entry. SipA is able to directly bind actin and through actin binds the actin-bundling protein T-plastin. The SipC protein is indispensable for invasion as it induces an increase in filamentous actin. SopE is injected into the cell via the SPI-1 TTSS, where it binds to and activates Cdc42 and Rac, which results in the formation of membrane ruffles known as filopodia or lamillipodia. SptP has an antagonizing effect to SopE by stimulating Cdc42 and Rac towards their inactive state. (modified from Donnenberg, 2000)

The SPI-1-located *sip* (*Salmonella* Invasion Protein) operon encodes the SipB, C, D and A secreted proteins, which are essential for intestinal invasion. It appears that all four Sip proteins are involved in invasion. However whereas inactivation of SipB, C and D causes a profound effect on the invasive properties of *Salmonella* (Collazo & Galan, 1997a), mutations of SipA reduce its capacity to invade but the effect is overall more subtle. This observation is consistent with the discovery that compared to SipB, C and D, SipA does not have a role in the translocation of effector proteins into the eukaryotic cell. Meanwhile, SipA is able to directly interact with F-actin, modulating the bundling ability of fimbrin, enhancing the rearrangements associated with bacterial internalization into intestinal epithelial cells. The exact way by which this function contributes to the entry process is not understood (Zhou *et al.*, 1999). SipC, on the other hand, is indispensable for invasion. It was shown that the amino-terminal end of SipC induces actin bundling (crosslinking), while the carboxyl terminal end is involved in actin polymerisation. This has been shown both *in vitro*

and by microinjection into cells. Furthermore, it causes an increase in filamentous actin (Hayward & Koronakis, 1999).

The release of SipB and SipC by the bacterium results in apoptosis of the host cell. SipB is able to bind caspase-1 and it is presumed that it is able to activate it; this leads to cleavage of pro-IL-1 and thereby induces apoptosis. SipC is secreted through the *Salmonella* Pathogenicity Island-2 secretion system and will block vesicle trafficking inside the host cell, bringing the cell machinery to a standstill (Donnenberg, 2000).

Once inside the host cell, two multifunctional virulence systems are known to control intracellular survival and replication during systemic infection: *Salmonella* Pathogenicity Island 2 (SPI-2), including genes for a second T3SS and some of the components of the PhoP-PhoQ two component regulatory system (Holden, 2002). SPI-2 and associated genes are responsible for maintaining the *Salmonella* containing vacuole (SCV) within the host cell.

The SPI-2 T3SS is distinctly different from the SPI-1 T3SS. They are controlled by different environmental signals and each translocates a unique set of effector proteins. Not all of the effector proteins secreted by SPI-1 T3SS are encoded within SPI-1. The SPI-2 T3SS shares similarity with enteropathogenic *E.coli* (EPEC). SPI-1 genes are induced prior to internalization, and are down-regulated once the bacterium has entered the host cell. SPI-2 genes on the other hand are essential for intracellular survival. However, due to the lag-time during internalization it is likely that SPI-1 genes do contribute to intracellular pathogenesis. SPI-2 consists of at least two distinct elements. The 25kb encoding the SPI-2 T3SS is only present in *S. enterica* and has a GC content of 43%. A 15kb element is found in both *S. enterica* and *S. bongori* and its GC content is significantly higher (54%). This 15kb element encodes tetrathionate reductase (Ttr) which is involved in anaerobic respiration and is dispensable for systemic virulence (Hensel *et al.*, 1999). A comparison with T3SSs of various pathogens showed that the SPI-2 encoded T3SS is more closely related to EPEC T3SS than to the SPI-1 encoded T3SS. These findings suggest that these two T3SS have resulted from unrelated events of horizontal gene transfer and are not the result of duplication of a gene cluster (Foultier *et al.*, 2002). The SPI-2 T3SS translocates effectors across the vacuole membrane; it has been shown to be important in causing systemic infection in mice and replication within macrophages

(Hensel *et al.*, 1998; Ochman *et al.*, 1996). Gene expression of SPI-2 is regulated by the OmpR-EnvZ two component system. The transcriptional regulator OmpR will bind the promoter region of *ssrA*, which in turn is required for the expression of the SPI-2 genes (Beuzon *et al.*, 2000; Lee *et al.*, 2000). SPI-2 gene expression is strongly upregulated at 1 and 6 hours post macrophage entry (Cirillo *et al.*, 1998; Pfeifer *et al.*, 1999). Gene expression is pH dependent and is upregulated after the bacterium has entered the macrophage and magnesium and phosphate starvation has set in. The pH inside the SCVs drops from pH6.0 to between pH4.0 and pH5.0 in the first hour following the uptake of bacteria, making it likely that acidification triggers SPI-2 T3SS gene expression (Beuzon *et al.*, 1999; Deiwick & Hensel, 1999; Deiwick *et al.*, 1999; Rathman *et al.*, 1996).

The *spiC* gene product is an important inhibitor of a number of cell trafficking events, including phagosome-lysosome and phagosome-endosome fusion, as well as normal vesicular trafficking in the degradative pathway (Uchiya *et al.*, 1999). SpiC may also have a role in secretion/translocation, the details of which are still unclear (Holden, 2002).

In order to accommodate the growing number of bacterial cells, *Salmonella* must ensure the recruitment of membrane to the vacuole, but it is not understood how exactly this is accomplished; however it was observed by EM that *S.* Typhimurium cells are closely associated with the vacuole membrane. *Salmonella*–induced filaments (Sifs) are tubular structures thought to be involved in maintaining the *Salmonella* containing vacuole SCV by either promoting fusion of specific vesicles with the SCV or inhibiting membrane loss from the vacuole (Salcedo *et al.*, 2001). *Sif* is expressed several hours after the bacteria have been taken up by endothelial cells. *sifA* is the principle gene encoding Sif tubules and its expression is regulated by the SPI-2 *SsrAB* regulatory system and it appears to be translocated via the SPI-2 T3SS, however it is physically located elsewhere on the *Salmonella* chromosome (Beuzon *et al.*, 2000; Miao & Miller, 2000).

In order for the bacteria to survive the SCV, they must avoid being damaged by the host cell through NADPH oxidase. It has become clear that the SPI-2 T3SS has a role in preventing the respiratory burst but the exact effector protein responsible remains to be identified (Holden, 2002). A more recent study postulates that *Salmonella* may actually require oxidants for the activation of SPI-2 and that rather than preventing oxidative stress, it is able to modulate it, they were able to implicate

two possible effectors, SseF and SseG in phox iNOS trafficking (Suvarnapunya & Stein, 2005). The SPI-2 T3SS translocates a number of effector proteins into the SCV involved in the regulation of vacuole membrane dynamics, including the formation of an F-actin meshwork which maintains the integrity of the SCV. One of the effectors involved in F-actin meshwork formation is *steC* kinase (*Salmonella* translocator effector C) (Meresse *et al.*, 2001; Poh *et al.*, 2008). The *spvB* gene, part of the *spv* operon is situated on the *Salmonella* virulence plasmid. It encodes for an ADP-ribosylating enzyme that depolymerises actin. The process appears to occur independent of SPI-2. SpvB is essential for *Salmonella* pathogenesis in mice (Lesnick *et al.*, 2001).

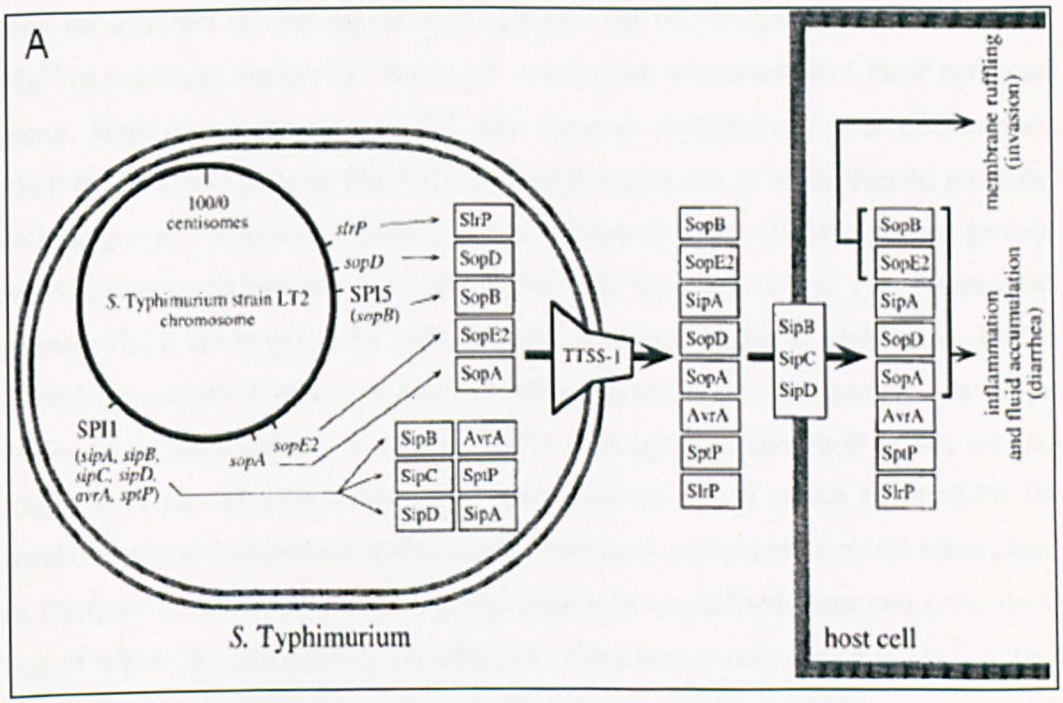


Figure 1.6. (A) Salmonella Pathogenicity Island Effector Proteins. The *Salmonella* chromosome is shown as a circle inside the bacterial cell, boxes represent the targets of the invasion-associated TTSS-1 and their role in causing diarrhoea. The TTSS gene products form a needle complex which spans the inner and outer membrane. Transport of TTSS-1 effector proteins into the host cell cytosol by the translocation complex formed by SipB, SipC and SipD is shown on the right. Positions of genes (*sop*A, *sop*E2, *slr*P and *sop*D) and pathogenicity islands (SPI-1 and SPI-5) on the physical map of the *Salmonella* Typhimurium chromosome are based on the complete genome sequence of strain LT2 (Zhang *et al.*, 2003) **(B)** An electronmicrograph of needle complexes isolated from *S.* Typhimurium (Kubori *et al.*, 1998)

The other multifactorial system governing intracellular survival of *Salmonella* is the PhoP/PhoQ two-component regulatory system, which governs virulence gene expression in *Salmonella* and other Gram-negative species such as *E. coli* and *P. aeruginosa*. It was first discovered in *S.* Typhiumurium as a regulator for the expression of non-specific acid phosphatase, but was later found to be the major regulon of *Salmonella* virulence (Groisman, 2001). PhoQ is a sensor histidine-kinase which phosphorylates PhoP, a response regulator, in response to environmental conditions. PhoQ features two transmembrane regions with a long C-terminal cytoplasmic tail carrying the histidine residue which is the site of autophosphorylation. The periplasmic domain harbours several acidic residues which may be involved in sensing divalent cations. The PhoQ sensor uses extracellular Mg^{2+} as a primary signal. Ca^{2+} and Mn^{2+} can repress transcription of PhoP activated genes. High concentrations of Mg^{2+} will promote dephosphorylation of phospho-PhoP by the PhoQ protein. PhoP/PhoQ control expression of more than 40 proteins, including over 15 outer membrane proteins, designated *pags* (PhoP-activated genes), which promote *Salmonella* survival within host tissues and *prg*, (PhoP-repressed genes), which are required for epithelial cell invasion (Miller & Mekalanos, 1990). Therefore, a major function of PhoP/PhoQ is remodelling of the bacterial envelope during host colonization (Ernst *et al.*, 2001). A suggestion that PhoP senses pH and responds to low pH has largely been dismissed as only a subset of PhoP/PhoQ-regulated genes is expressed under acidic conditions and expression still takes place in PhoQ or PhoP mutants. The *phoP/Q* operon is transcribed from two promoters, one of which is constitutively switched on. The other is only active if Mg^{2+} is low and the PhoP/PhoQ proteins are present (Socini and Groisman, 1996).

Transcription of *pags* is induced 3 to 4 hours post macrophage-internalisation. PhoQ, located inside the inner bacterial cell membrane, senses low Mg^{2+} and Ca^{2+} concentrations, to which it responds by phosphorylating PhoP, and activates *pags* expression (Garcia Vescovi *et al.*, 1996). Pags are involved in LPS modification and antimicrobial resistance (Groisman, 2001).

The PhoP/PhoQ regulons other important role is in avoidance of late endocytic compartments of the phagocytosis pathway (Garvis *et al.*, 2001).

Host cell death can occur via three different mechanisms, the first of which is SPI-1 dependent, while the second is dependent of SPI-2. During early stages of disease the SPI-1 effector SipB interacts with Caspase-1, an interleukin-1-beta-converting

enzyme. Caspases are cysteine proteases important in apoptosis. It was therefore suggested that SipB mediated cell death was apoptotic (Monack *et al.*, 2000), however evidence suggests that SipB causes necrosis, caspase-1 is dispensable during apoptosis. A hallmark of necrosis is membrane damage, with cytosol leaking out of the damaged cell, which in turn causes a strong proinflammatory response, as was observed for *Salmonella* infection of macrophages (Brennan & Cookson, 2000). The second, SPI-2 dependent process is apoptotic and delayed, the effectors have not been identified but are known to be *ompR* dependent (van der Velden *et al.*, 2000). The third type of host cell death has recently been described as pyroptosis, this mechanism shows hallmarks of both apoptosis and necrosis. Once inside the host cell cytosol, SipB will bind caspase-1, as it does in apoptosis, resulting in the catalytic cleavage and release of proinflammatory cytokines IL-1-ß and IL-18, resulting in a rapid proinflammatory cell death (Hersh *et al.*, 1999; Mariathasan *et al.*, 2004).

1.5.3.2. SPI-3

The *Salmonella* Pathogenicity Island 3 (SPI-3) harbours the *mgtC* gene that is required for intramacrophage survival by ensuring that proper levels of Mg^{2+} are available to the microorganism. *mgtC* mutants are defective for growth in a low Mg^{2+} medium. SPI-3 is conserved between *Salmonella* serovars Typhimurium and Typhi (Blanc-Potard *et al.*, 1999). There are, however extensive variations in the structure of SPI-3 ranging from deletions to insertions of whole additional gene clusters. Variations are mainly located near the insertion site *selC*, while the *mgtCB* genes are conserved. Interestingly, close relatives to *Salmonella,* such as *E. coli,* lack the *mgtC* gene and unlike *Salmonella* cannot grow in low Mg^{2+}. On the other hand very distantly related pathogens such as *Mycobacterium tuberculosis* harbour a *mgtC* homolog that is also required for growth in low Mg^{2+} concentrations, proliferation within macrophages and virulence. These findings suggest that acquisition of the *mgtC* gene was an important step in the development of *Salmonella* as an intracellular pathogen (Groisman, 2001).

It was found that genes towards the centre of SPI-3 were flanked by remnants of IS elements, indicating an insertion within SPI-3. *selC* appears to be a hot spot for integration of foreign DNA given the variability of SPI-3 in *Salmonella* and the presence of different PAIs at this locus in various other pathogens (Hensel, 2004).

Control by the PhoP/PhoQ regulon ensures that these genes are expressed at the right time and in the right place, (i.e. within the macrophages). It was hypothesised that *Salmonella* determines its location at least in part by examining the Mg^{2+} concentration: a low concentration indicates an intracellular environment while a high Mg^{2+} concentration is typical for an extracellular environment. It appears that all PhoP genes in *Salmonella* were acquired by horizontal gene-transfer (Groisman, 2001). shdA is a fibronectin binding protein located on SPI-2 that is exclusive to subspecies I, it is thought that shdA is important in the persistence of subspecies I in the intestine of warm blooded animals (Kingsley *et al.*, 2002)

1.5.3.3. SPI-4 and SPI-5

The sequences of SPI-4 in *S.* Typhi and *S.* Typhimurium LT2 are almost identical. SPI-4 carries 18 ORFs, one locus was initially implicated in murine macrophage survival, leading to SPI-4 being designated a pathogenicity island. SPI-4 and SPI-9 of *S.* Typhi share four genes with ~40% nucleotide identity. It is possible that only SPI-4 is functional in *S.* Typhimurium, while SPI-9 has a similar function in *S.* Typhi. In SPI-4, six of the 18 ORFs (originally designated STM4257-4262) are organized into the *siiA-F* operon (*Salmonella* intestinal infection). *siiE*, which encodes a 595 kDa secreted protein, of *S.* Typhi contains a stop codon and is therefore likely to be non-functional. (Parkhill *et al.*, 2001), *siiC*, *siiD*, *siiF* encode a type one secretion system, the function of *siiA* and *siiB* is currently unknown. A large conserved operon polarity repressor (*ops*) in located upstream of *siiA*. This motif is required for transcription elongation under control of the RfaH protein and is associated with several other virulence gene clusters in the Enterobacteriaceae. Using signature tagged mutagenesis Morgan and co-workers were able to demonstrate that SPI-4 is not required for enteropathogenic responses, it is required for intestinal but not systemic infection (Morgan *et al.*, 2004). siiE and siiF appear to play a minor role in the infection of calves but not chicken or pigs, with *S.* Typhimurium (Morgan *et al.*, 2007).

SPI-5 genes encode effector proteins for both of the T3SSs. The SopB effector protein is translocated by the SPI-1 encoded T3SS and its gene expression is under control of HilA the major transcriptional regulator, the gene for which is also located on SPI-1. The PipB effector is translocated by the SPI-2 encoded T3SS and under the control of its SsrAB two-component regulatory system. *sopB* encodes an inositol phosphatase, responsible for triggering fluid secretion, resulting in diarrhoea. The

gene is present in all *Salmonella* including *S.* bongori. In contrast *pipAB* is absent from *S. bongori* and *S. enterica* subspecies II. The base composition of these two elements also differs suggesting independent acquisition of the two genes (Knodler *et al.*, 2002).

1.5.3.4. SPI-6

SPI-6 also known as *Salmonella* Chromosomal Island (SCI) contains the *saf* gene cluster encoding fimbriae and the invasion encoding *pagN* gene, as well as several other genes with unknown functions. Deletion of the SPI-6 locus has no effect on *S.* Typhimurium's systemic pathogenicity (Folkesson, *et al* 2002). Parts of SPI-6 were also detected in *Salmonella* subspecies IIIb, IV and VII. Homologoes of SPI-6 have also been identified in enterohaemorrhagic *E. coli*, *Pseudomonas aeruginosa* and *Yersinia pestis* genome sequences, but the function of these homologues is unknown (Hensel, 2004).

1.5.3.5. *S.* Typhi pathogenicity islands SPI-7, SPI-8 and SPI-9

SPI-7 is specific to *Salmonella* serovars Typhi, Dublin and Paratyphi C. It is also known as the Major Pathogenicity Island (MPI) of *S.* Typhi (Zhang *et al.*, 1997). Arguably the most important viulence factor encoded for by SPI-7 is the Vi antigen, a capsular exopolysaccharide. Another potential virulence factor is the type IVB pilus encoded by the *pil* gene cluster. SPI-7's composition is rather complex, indicating independent acquisition of its genetic elements by horizontal gene transfer. It was demonstrated that parts of SPI-7 are also present in *P. aeruginosa*, in which the locus is referred to as PAGI-3, and the plant pathogen *Xanthomonas axonopodis*, suggesting that *Salmonella* acquired this locus through the contact with environmental bacteria. *S.* Typhi can lose its Vi capsule, suggesting that SPI-7 is not stable within the *S.* Typhi chromosome (Pickard *et al.*, 2003).

SPI-8 is another, small PAI found in *S.* Typhi. It encodes bacteriocin genes, however its function in virulence has yet to be demonstrated. SPI-9 is also present in *S.* Typhi, carrying putative virulence factors such as a Type 1 Secretion System (T1SS) and a large RTX-like toxin. However, at least in parts, this PAI has also been found in *S.* Typhimurium, *S. bongori* and other serovars. SPI-10 carries a cryptic bacteriophage, the role of which is unclear (Parkhill *et al.*, 2001). In a subset of serovars including Typhi and Enteritidis, SPI-10 also carries the *sef* fimbrial genes,

which are known virulence determinants and responsible for host specificity (Townsend *et al.*, 2001).

1.5.3.6. *Salmonella* Genomic Island 1 (SGI-1)

Multidrug resistant *Salmonella* such as the *S.* Typhimurium strain DT104 carry the SGI-1 island. This island is 43 kb in size and, in contrast to plasmid mediated resistance, is stable within the genome in the absence of selective pressure. SGI-1 confers pentaresistance: resistance against ampicillin, tetracycline, chloramphenicol, streptomycin and sulfonamides. These resistance gene clusters are in a region made up of two integrons. Variants of SGI-1 carrying other combinations of antibiotic resistance genes have been identified in various serovars, integrated into the genome at the same chromosomal location as SGI-1. These findings suggest horizontal transfer with site-specific recombination and that antibiotic resistance loci are easily adapted (Hensel, 2004).

1.5.3.7. High Pathogenicity Island (HPI)

HPI was initially identified in highly virulent *Yersinia enterocolitica* and *Y. pseudotuberculosis* and is present in a variety of Gram negative pathogens (Schubert *et al*, 2004). HPI carries the biosynthesis pathway for the siderophore yersiniabactin and cognate iron uptake system. The presence of HPI within a strain appears to determine its ability to cause septicaemic infections. HPI is absent from human-adapted subspecies I, but has been found in subspecies IIIa, IIIb and IV (Oelschlaeger *et al.*, 2003).

Table 1.5. Summary of important SPIs

Name	size (kb)	GC content (%)	Distribution	insertion site	role
SPI -1	39.8	47	*Salmonella* spp Conserved	*flhA-mutS*	T3SS, Fe^{2+} uptake, invasion of enterocytes, apoptosis
SPI-2	39.7	44.6	*S. enterica,* conserved	tRNA *valV*	T3SS, survival inside macrophages
SPI-3	17.3	39.8-49.3	*Salmonella* spp variable	tRNA *selC*	Mg^{2+} uptake, survival inside macrophages
SPI-4	23.4	44.8	*Salmonella* spp conserved	tRNA?	T1SS, survival inside macrophages
SPI-5	7.6	43.6	*Salmonella* spp variable	tRNA *serT*	T3SS effectors, enteropathogenicity
SPI-6 (SCI)	59	51.5	subsp. I, parts in IIIb, IV, VII	tRNA *aspV*	fimbriae
SPI-7 (MPI)	133	44-53	subsp. I serovars instable	tRNA *pheU*	Vi antigen, pilus assembly, *sopE*
SPI-8	6.8	38.1	*S.* Typhi	tRNA *pheV*	unknown
SPI-9	16.3	56.7	subsp. I serovars	prophage	putative toxin, unknown
SPI-10	32.8	46.6	subsp. I serovars	tRNA *leuX*	*sef* fimbriae
SGI-1	43	48.4	subsp. I serovars variable	*thadF-yidY*	5 antibiotic resistance genes
HPI	?	?	subsp. IIIa, IIIb, IV	tRNA *asnT* (*ychF*)	high affinity Fe^{2+} uptake

1.6. Suppression Subtractive Hybridisation (SSH)

Suppression Subtractive Hybridisation (SSH) was developed by Diatchenko and co-workers as a variation of suppression polymerase chain reaction (PCR). SSH combines two major events: normalisation and subtraction. The normalisation step equalizes the abundance of DNA within the target population, the subtraction step eliminates sequences common to both populations (Diatchenko *et al.*, 1996).

Diatchenko and co-workers initially used the technique for comparison of differentially expressed genes in eukaryotes, by reverse transcribing messenger RNA (mRNA) into complementary DNA (cDNA). However since then SSH has found its application in prokaryotic genetics as an elegant method of comparing bacterial genomes and to identify genes and sequences present in some bacteria but absent in other members of the species/genus/family (Winstanley, 2002). These genes may determine strain-specific characteristics such as drug resistance, bacterial surface structure or restriction modification. SSH is particularly applied in the identification

of pathogenicity islands (PAI) present in virulent strains but absent from non-pathogenic strains. The first bacterium used in a SSH study was *Helicobacter pylori* in 1998 (Akopyants *et al.*, 1998). The study initially identified the *cag* PAI, a type IV secretion system, by comparing a tester strain with a strain where *cag* had been deleted. They then proceeded to compare the genomic contents of two strains of *H. pylori*, one of which had been fully sequenced (Akopyants *et al.*, 1998).

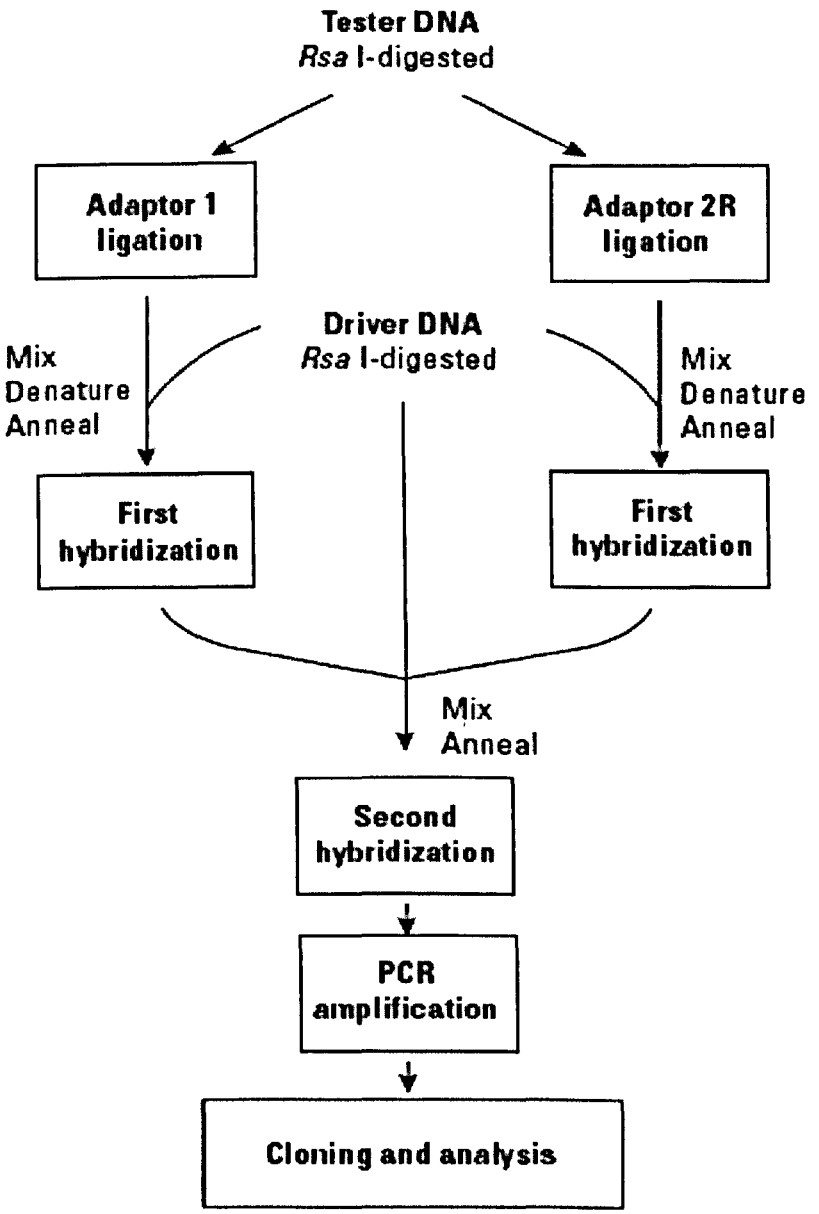


Figure 1.7. Schematic representation of PCR-Select suppression subtraction hybridisation. (Clontech PCRSelect™ Bacterial Genome Subtraction kit User Manual).

Figure 1.7 shows a schematic representation of a general SSH. The two strains are allocated into tester and driver. The driver strain is the control strain, whose genomic

contents will be subtracted from the tester strain, leaving only DNA sequences unique to the tester. The tester is digested with a restriction endonuclease such as *Rsa*I. It is subsequently divided into two equal quantities. A ligation reaction is carried out whereby a different type of adaptor is attached to the 5' ends of the sequences. Both portions are separately hybridised to the driver DNA which is in excess to ensure that all the sequences also present in the tester are bound. The first hybridisation results in the presence of tester-specific single stranded sequences amongst a pool of double-stranded sequences. A second hybridisation step is then carried out in which the two portions are mixed. Homologous single-stranded DNA will hybridise and sequences specific to the tester will carry two different adaptors at the ends. PCR primers are specifically designed to the adaptors and will therefore only amplify sequences carrying two different adaptors. Sequences which have previously remained single stranded will form secondary structures and become unavailable for amplification. PCR amplicons can be used to create a subtraction library, by cloning the sequences into a suitable vector. The procedure is not 100 per cent effective but >50% of clones should be tester specific (Winstanley, 2002).

1.7. Aims and Objectives

Bacterial infections continue to be a major cause of morbidity and mortality in African children. NTS are one of the commonest bloodstream isolates in these children. Evidence discussed above indicates that these bacteria differ from the gastroenteritis isolates found in the developed world in several ways, notably their mode of transmission, the risk factors of acquiring infection, the symptoms shown by patients and the antibiotic resistance profiles. It remains to be elucidated whether these isolates differ in their genetic make-up from the common European isolates. It is certain, that better characterisation of NTS will allow for more efficient identification and better disease management in developing countries.

The aim of this research was to use Suppression Subtractive Hybridisation in order to identify potential virulence factors exclusive to invasive African NTS and in turn determine their distribution among a panel of strains isolated from extra-intestinal NTS infection patients from sub-Saharan Africa. The panel of strains consists of bacteraemia isolates from Malawi, Kenya, Uganda and Zaire (DRC), as well as human diarrhoea and veterinary isolates from the UK.

A total of four subtractions were carried out on NTS strains of four different serovars, namely Heidelberg, Typhimurium, Enteritidis and Bovismorbificans, isolated from Malawian children with NTS bacteraemia. *Salmonella* Typhimurium and Enteritidis are by far the commonest isolates in sub-Sahran Africa and Heidelberg and Bovismorbificans feature prominently among the common-rare isolates.

Salmonella serovar Bovismorbificans has not previously been described in great detail and using next-generation sequencing techniques we set out to characterise a set of Malawian bacteraemia and UK veterinary isolates of this serovar.

1.8. Hypothesis

We propose that NTS bacteremia isolates from sub-Sahran Africa differ in their genetic content through loss or gain of genetic material, or loss of gene function from common gastroenteritis isolates.

CHAPTER 2: MATERIALS AND METHODS

2.1. Bacterial strains

African NTS isolates were obtained from existing collections from Uganda, Malawi, Kenya and the Democratic Republic of Congo (Zaire) maintained in our Department. UK isolates were obtained from human and animal faeces (Table 2.1 and 2.2). Other isolates were obtained from the Sanger Institute. For UK isolates, serovar designations were confirmed by the *Salmonella* Reference Laboratory, Colindale, London. For the Malawi isolates, serovars other than *S.* Typhimurium and Enteritidis were confirmed by the National *Salmonella* Reference Laboratory, Galway, Republic of Ireland.

All *Salmonella* strains were stored in 10% glycerol broth at both -20 and -80°C. Strains were subcultured by partially thawing the stored sample and inoculating a fresh agar plate with a 10µl aliquot. Strains were subcultured onto Columbia agar. These cultures can then be stored at 4°C for up to 8 weeks. *Salmonella* strains were cultured in Luria broth overnight at 37°C shaking at 200rpm. For a summary of all strains used throughout this study see Tables 2.1. and 2.2.

Table 2.1. Summary of African and UK *Salmonella* strain sets

Country of origin	Serovar	Number	Details and source
Uganda	Typhimurium	9	Adult diarrhoea isolates
	Enteritidis	13	Martin Okong (unpublished)
	Heidelberg	1	
	Stanleyville	1	
Malawi	Typhimurium	17	Children bacteraemia isolates
	Enteritidis	4	1998-2004
	Heidelberg	1	Melita Gordon *et al*, 2008
	Bovismorbificans	4	
	Bukavu	1	
	Sundsvall	1	
Kenya	Typhimurium	4	Adult bacteraemia isolates , 1994-
	Enteritidis	4	2003, Sam Kariuki *et al*, 2005
DRC (Zaire)	Typhimurium	7	Children bacteraemia isolates
	Enteritidis	5	Green *et al*, 1993
UK	Typhimurium	6	RLUH[1],
	Enteritidis	6	(unpublished) Adult diarrhoea isolates
UK Veterinary	Heidelberg	3	Veterinary isolates
	Bovismorbificans	4	Veterinary Faculty, Leahurst (unpublished)
sporadic	Heidelberg	3	Gut, Sanger Institute, (unpublished)
Zanzibar	Heidelberg	1	Gut, Sanger Institute, (unpublished)
Kenya	Heidelberg	1	Blood, Sanger Institute, (unpublished)
Tanzania	Heidelberg	1	Faeces, Sanger Institute, (unpublished)
Nigeria	Heidelberg	1	Gut, Sanger Institute, (unpublished)
Peru	Heidelberg	1	Gut, Sanger Institute, (unpublished)
Thailand	Heidelberg	1	Gut, Sanger Institute, (unpublished)
Malaysia	Heidelberg	1	Gut, Sanger Institute, (unpublished)

Table 2.2. Panel of invasive African NTS strains and UK control strains. Origin = U=Uganda, M=Malawi, K=Kenya. Genotyping is based on pulsed-field gel electrophoresis (PFGE) carried out by C. Broughton (Malawi strains), M. Okong (Uganda strains) and S. Javed (UK). The genotype is only specified were known. All the strains originated from children, with the exception of the Ugandan and UK control strains which are adult or veterinary isolates (ND= no data).

Strain	Serovar	Origin	Date of Isolation	Source A=Adult C=Child	site of infection	Phage type	Genotype
81	*S.* Enteritidis	U	1996-2003	A	Blood		1
4002	*S.* Enteritidis	U	1996-2003	A	Blood		2
93	*S.* Enteritidis	U	1996-2003	A	Blood		1A
4016	*S.* Enteritidis	U	1996-2003	A	Blood		1B
204	*S.* Enteritidis	U	1996-2003	A	Blood		1C
213	*S.* Enteritidis	U	1996-2003	A	Blood		1C
263	*S.* Enteritidis	U	1996-2003	A	Blood		1C
4068	*S.* Enteritidis	U	1996-2003	A	Blood		1E
298	*S.* Enteritidis	U	1996-2003	A	Blood		1F
300	*S.* Enteritidis	U	1996-2003	A	Blood		1C
791	*S.* Enteritidis	U	1996-2003	A	Blood		1
1181	*S.* Enteritidis	U	1996-2003	A	Blood		1B
1305	*S.* Enteritidis	U	1996-2003	A	Blood		1
146	*S.* Typhimurium	U	2002-2003	A	Blood		1
256	*S.* Stanleyville	U	1996-2003	A	Blood		2
666	*S.* Typhimurium	U	2002-2003	A	Blood		1
812	*S.* Typhimurium	U	2002-2003	A	Blood		1A
845	*S.* Heidelberg	U	1996-2003	A	Blood		3
868	*S.* Typhimurium	U	1996-2003	A	Blood		1A
4448	*S.* Typhimurium	U	1996-2003	A	Blood		1
884	*S.* Typhimurium	U	1996-2003	A	Blood		1
4234	*S.* Typhimurium	U	1996-2003	A	Blood		1A
1060	*S.* Typhimurium	U	1996-2003	A	Blood		1
1190	*S.* Typhimurium	U	2002-2003	A	Blood		1
D20941	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D20890	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D21685	*S.* Enteritidis	M	01.09.2004	C	Blood		1
D22076	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D22209	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D22219	*S.* Typhimurium	M	18.09.2004	C	Blood		1D/G
D22337	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D22404	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D22988	*S.* Typhimurium	M	18.09.2004	C	Blood		1D/G
D23002	*S.* Typhimurium	M	18.09.2004	C	Blood		1D/G
D23145	*S.* Enteritidis	M	01.09.2004	C	Blood		2
D23424	*S.* Typhimurium	M	03.09.2004	C	Blood		1D/G
D23674	*S.* Typhimurium	M	18.09.2004	C	Blood		1K
D23682	*S.* Enteritidis	M	01.09.2004	C	Blood		2
D23734	*S.* Heidelberg	M	17.09.2004	C	Blood		2.0
D24627	*S.* Typhimurium	M	21.09.2004	C	Blood		1D/G
D25352	*S.* Typhimurium	M	ND	C	Blood		1D/G
D25834	*S.* Typhimurium	M	15.09.2004	C	Blood		1D/G
D25840	*S.* Typhimurium	M	17.09.2004	C	Blood		1D/G
D25907	*S.* Typhimurium	M	20.09.2004	C	Blood		1D/G
D25991	*S.* Typhimurium	M	15.09.2004	C	Blood		1K
D26104	*S.* Typhimurium	M	17.09.2004	C	Blood		1.0
D8993.1	*S.* Enteritidis	M	ND	C	Blood		2
D8993.2	*S.* Enteritidis	M	ND	C	Blood		2A
1	*S.* Typhimurium	DRC	pre 1992	C	Blood		
2	*S.* Typhimurium	DRC	pre 1992	C	Blood		
3	*S.* Typhimurium	DRC	pre 1992	C	Blood		
4	*S.* Typhimurium	DRC	pre 1992	C	Blood		
5	*S.* Typhimurium	DRC	pre 1992	C	Blood		
1	*S.* Enteritidis	DRC	pre 1992	C	Blood		
2	*S.* Enteritidis	DRC	pre 1992	C	Blood		
3	*S.* Enteritidis	DRC	pre 1992	C	Blood		

Strain	Serovar	Origin	Date of Isolation	Source A=Adult C=Child	site of infection	Phage type	Genotype
4	*S.* Enteritidis	DRC	pre 1992	C	Blood		
5	*S.* Enteritidis	DRC	pre 1992	C	Blood		
6	*S.* Enteritidis	DRC	pre 1992	C	Blood		
6.2	*S.* Enteritidis	DRC	pre 1992	C	Blood		
9663	*S.* Typhimurium	K	2002-2005	A	Blood		
9664	*S.* Typhimurium	K	2002-2005	A	Blood		
9665	*S.* Typhimurium	K	2002-2005	A	Blood		
9812	*S.* Typhimurium	K	2002-2005	A	Blood		
417	*S.* Enteritidis	K	2002-2005	A	Blood		
506	*S.* Enteritidis	K	2002-2005	A	Blood		
705	*S.* Enteritidis	K	2002-2005	A	Blood		
767	*S.* Enteritidis	K	2002-2005	A	Blood		
811	*S.* Bovismorbificans	M	31.05.1996	C	Blood	PT2	
D993	*S.* Bovismorbificans	M	31.03.1996	C	Blood	PT2	
3476	*S.* Bovismorbificans	M	18.02.1997	C	Blood	PT2	
3064	*S.* Bukavu	M	ND	C	Blood	PT2	
3114	*S.* Bovismorbificans	M	20.01.1997	C	Blood	PT2	
2291	*S.* Bovismorbificans	M	30.10.1996	C	Blood	PT2	
170	*S.*Bovismorbificans	M	15.03.1996	C	Blood		
3064	*S.*Bovismorbificans	M	04.01.1997	C	Blood		
3160	*S.*Bovismorbificans	M	26.01.1997	C	Blood		
3180	*S.*Bovismorbificans	M	26.01.1997	C	Blood		
2819	*S.*Bovismorbificans	M	19.02.1996	C	Blood		
5104	*S.*Bovismorbificans	M	02.07.1997	C	Blood		
6571	*S.*Bovismorbificans	M	05.12.1997	C	Blood		
A7360	*S.*Bovismorbificans	M	ND	A	Blood		
D4891	*S.*Bovismorbificans	M	31.01.2000	C	Blood		
D4896	*S.*Bovismorbificans	M	ND	C	Blood		
D1253	*S.*Bovismorbificans	M	27.04.1999	C	Blood		
D2877	*S.*Bovismorbificans	M	14.09.1999	C	Blood		
A5893	*S.*Bovismorbificans	M	31.08.1999	A	Blood		
D9582	*S.*Bovismorbificans	M	08.01.2001	C	Blood		
A10387	*S.*Bovismorbificans	M	24.12.2000	A	Blood		
D4551	*S.*Bovismorbificans	M	09.01.2000	C	Blood		
7822	*S.*Bovismorbificans	M	ND	C	Blood		
A9112	*S.*Bovismorbificans	M	ND	A	Blood		
A8737	*S.*Bovismorbificans	M	ND	A	Blood		
12855	*S.*Bovismorbificans	M	ND	C	Blood		
A1104	*S.*Bovismorbificans	M	ND	A	Blood		
A1668	*S.*Bovismorbificans	M	ND	A	Blood		
A9362	*S.*Bovismorbificans	M	ND	A	Blood		
A1608	*S.*Bovismorbificans	M	ND	A	Blood		
A15523	*S.*Bovismorbificans	M	ND	A	Blood		
A16934	*S.*Bovismorbificans	M	ND	A	Blood		
A16982	*S.*Bovismorbificans	M	ND	A	Blood		
D15823	*S.*Bovismorbificans	M	ND	C	Blood		
A15363	*S.*Bovismorbificans	M	ND	A	Blood		
A18317	*S.*Bovismorbificans	M	ND	A	Blood		
A19308	*S.*Bovismorbificans	M	ND	A	Blood		
D18327	*S.*Bovismorbificans	M	ND	C	Blood		
D19094	*S.*Bovismorbificans	M	ND	C	Blood		
D19206	*S.*Bovismorbificans	M	ND	C	Blood		
A22622	*S.*Bovismorbificans	M	ND	A	Blood		
A22921	*S.*Bovismorbificans	M	ND	A	Blood		
A22920	*S.*Bovismorbificans	M	ND	A	Blood		
D19953	*S.*Bovismorbificans	M	ND	C	Blood		
A23856	*S.*Bovismorbificans	M	ND	A	Blood		
A24091	*S.*Bovismorbificans	M	ND	A	Blood		
A30633	*S.*Bovismorbificans	M	20.08.2004	A	Blood		
A31126	*S.*Bovismorbificans	M	19.09.2004	A	Blood		
D28312	*S.*Bovismorbificans	M	25.11.2004	C	Blood		
D28655	*S.*Bovismorbificans	M	13.12.2004	C	Blood		
D9375	*S.*Bovismorbificans	M	15.12.2000	C	Blood		
D22402	*S.* Sundsvall	M	18.09.2004	C	Blood		
37002	*S.* Typhimurium	UK	ND	A	Diarrhea	DT193	

Strain	Serovar	Origin	Date of Isolation	Source A=Adult C=Child	site of infection	Phage type	Genotype
37069	S. Typhimurium	UK	ND	A	Diarrhea	DT104	
47014	S. Typhimurium	UK	ND	A	Diarrhea	DT104	
57021	S. Typhimurium	UK	ND	A	Diarrhea	DT104	
57193	S. Typhimurium	UK	ND	A	Diarrhea	DT193	
57315	S. Typhimurium	UK	ND	A	Diarrhea	DT8	
57115	S. Enteritidis	UK	ND	A	Diarrhea	PT4	
57126	S. Enteritidis	UK	ND	A	Diarrhea	PT21	
57190	S. Enteritidis	UK	ND	A	Diarrhea	PT8	
57255	S. Enteritidis	UK	ND	A	Diarrhea	PT4	
57285	S. Enteritidis	UK	ND	A	Diarrhea	PT1	
67155	S. Enteritidis	UK	ND	A	Diarrhea	PT6	
20031619	S.Heidelberg	Zanzibar	ND		gut		HeiX1
20040049	S.Heidelberg	K	ND		blood		HeiX1
20040778	S.Heidelberg	Malaysia	ND		gut		nd
20041283	S.Heidelberg	Peru	ND		gut		HeiX1
20050631	S.Heidelberg	sporadic	ND		gut		HeiX9
20051845	S.Heidelberg	Thailand	ND		gut		HeiX5
20060448	S.Heidelberg	Nigeria	ND		gut		HeiX3
20070231	S.Heidelberg	sporadic	ND		gut		HeiX10
20070302	S.Heidelberg	Tanzania	ND		faeces		HeiX1
20070502	S.Heidelberg	sporadic	ND		gut		HeiX10
KMS1977	S.Heidelberg	UK	ND		Pig		
55/366	S.Heidelberg	UK	ND		Pig		
17705	S. Heidelberg	UK	ND		Pig		
544086/79	S. Bovismorbificans	UK	ND		Pig		
518927/76	S. Bovismorbificans	UK	ND		pig		
5840/80	S. Bovismorbificans	UK	ND		Pig		
NCTC5754	S. Bovismorbificans	UK	ND		pig		

2.2.Culture Media

Luria Broth medium / LB agar plates consisted of the following

1l	deionised H_2O
10g	tryptone
5g	yeast extract
5g	NaCl
1%	agar

The ingredients were mixed using a magnetic stirrer until dissolved and the liquid was clear. In order to produce Luria (L) -agar, 1% of agar (Sigma, Steinheim, Germany) was added, the broth/agar was divided into Duran glass bottles and autoclaved.

The media supplements ampicillin (Sigma) (100 μg ml^{-1}), 5-bromo-4-chloro-3-indoyl-β-D-galactoside (X-Gal) (Melford Laboratories, Ipswich, UK) (80μg ml^{-1}) and isopropylthio-β-D-galactoside (IPTG) (Melford Laboratories) (100μg ml^{-1}) were stored as stock solutions. LB plates containing 100 μg/ml ampicillin, 100 μg/ml of X-Gal and 200 μg/ml IPTG were made for Blue/White colony screening. 400 ml of agar were melted slowly in a microwave on the medium setting and then allowed to cool in order not to damage the ampicillin. This was achieved by placing the bottle in a waterbath set at 40-50°C, which ensures that the agar will not begin to set. After the agar had cooled down, ampicillin, X-Gal and IPTG were added. 400ml of agar are sufficient for about 35 plates. The plates were stored at 4°C until further use for up to 4 weeks.

LB ampicillin plates were produced in the same way as the above, using the same concentration of ampicillin. LB ampicillin agar was also used to produce 96-well culture plates, used for sequencing. A multichannel pippettor was set to 200 μl, which is sufficient to fill about 4 rows of wells (~50μl in each well); it is important to move on swiftly to prevent the agar from setting inside the tips. Again plates were stored at 4°C for up to 4 weeks.

2.3. DNA extraction methods

2.3.1. Preparation of crude DNA for PCR amplification

1 to 3 colonies were picked up from the plate using a sterile loop and were suspended in 50 µl of sterile distilled water (SDW). The suspension was heated to 99°C for 5 min in a thermal cycler (Labnet Multi Gene II, Labnet International, Inc, Oakham, Rutland, UK) and centrifuged at 13,000 rcf for 5 min. These preparations cannot be stored and were always used immediately.

2.3.2. Chelex-100 suspension method

One or more pure colonies were suspended in 200 µl of a 5% Chelex-100 solution (Chelex-100 Molecular Biology Grade Resin, Bio-Rad Laboratories, Hemel Hempstead, UK). The suspension was mixed vigorously before and after adding the colonies. The suspension was then boiled in a waterbath for 5-10 min, mixed again and centrifuged at 13000 rpm for 1 min. 150 µl of the supernatant were carefully removed and placed in a fresh tube. The DNA preparation can be stored at -20°C until further use. Chelex suspensions were mainly used for PCR Screening Assays.

2.3.3. Genomic DNA extraction and DNA purification

Pure DNA for Subtraction Hybridisation, Dot Blot analysis and sequencing was isolated using the Wizard® Genomic DNA Purification Kit (A1120, Promega, Madison, USA) as described in the manufacturer's instructions with a few exceptions: instead of one, three ml of culture were used. To improve protein precipitation 3 extra steps from the troubleshooting section were incorporated into the protocol: the sample is chilled on ice for five minutes instead of just cooling to room temperature, it is then vortexed for 20 s and centrifuged for 3 min at 13000 rpm before proceeding with the standard protocol. In step 12 the centrifugation time was increased to 5 min (instead of 3 min), in step 13 the amount of isopropanol was reduced to 500µl to account for a reduction in volume of the supernatant. The same applied to step 9 where 500µl ethanol were used. The Centrifugation time was increased to 15 min in step 10, and in step 12 the rehydration step consisted of both 1 h at 65°C and overnight at 4°C.

2.3.4. Microcentrifuge Spin Columns

Vector inserts amplified by PCR were further purified using Microcentrifuge spin columns, following the manufacturer's instructions (MicroSpin S-400 Columns, GE

Healthcare, Little Chalfont, Buckinghamshire, UK,) before being sent for sequencing with the vector primers (M13F 5'-gttttcccagtcacgac-3', M13R 5'-caggaaacagctatga-3'), (Lark Technologies/ Cogenics, Takeley, Essex, UK).

2.3.5. Plasmid Extraction

Plasmid Extraction was performed on *S.* Typhimurium strain D26104 using the QIAprep Spin Miniprepkit (Qiagen, Crawley, UK) following the manufacturer's instructions:

In brief: *Salmonella* strains were cultured overnight in LB medium at 37°C, 200 rpm. RNase was added to Buffer P1. Bacterial cells were pelleted by centrifugation and then resuspended in 250 µl Buffer P1 and transferred to a microcentrifuge tube. 250 µl of Buffer P2 were added and mixed thoroughly by inverting the tube 4–6 times. 350 µl of Buffer N3 were added and mixed immediately and thoroughly by inverting the tube 4–6 times. The sample was centrifuged for 10 min at 13,000 rpm. The supernatants are then applied to the QIAprep spin column by decanting or pipetting and the column was centrifuged for 60 sec. The flow-through was then discarded and the column washed by adding 0.5 ml Buffer PB and centrifuging for 30–60 sec.

A final wash of the column by adding 0.75 ml Buffer PE and centrifuging for 30–60 s was performed. The column was centrifuged for an additional 1 min to remove residual wash buffer. The column was then placed into a clean 1.5 ml microcentrifuge tube. 50 µl Buffer EB (10 mM Tris·Cl, pH 8.5) were then added to the center of each column, which was left to stand for 1 min, and centrifuged for 1 min to elute the DNA.

2.4. Polymerase Chain Reaction (PCR) DNA amplification

2.4.1. Standard PCR protocol

PCR reactions were carried out in either 25 or 50 µl volumes, depending on the application. Each reaction contained 1.25 U of Taq Polymerase (GoTaq, Promega), 300 nM of each oligonucleotide primer, 1x Taq buffer, 2.5 mM $MgCL_2$ and 100 µM Nucleotides (dATP, dCTP, dGTP and dTTP) (Promega). 1µl of template DNA was used in each reaction.

All the primers were supplied by Sigma-Genosys (Pampisford, Cambridgeshire, UK) and stored in a stock solution of 1 mM at -20°C. Table 2.3 summarizes the

composition of a typical PCR master mixture, 24 µl (49µl) of Master Mix were aliquoted into 0.2 ml Eppendorf tubes and 1µl of DNA was added.

The standard PCR amplification program consisted of an initial denaturation step at 95°C for 5 min, followed by 30 cycles of denaturation at 95°C for 1 min, annealing at a suitable annealing temperature (as determined by temperature gradient PCR) for 1 min, and primer extension at 72°C for 2 min. A final extension step of 10 min at 72°C can be added. The following PCR machines have been used: Eppendorf Mastercycler Thermal Cycler (Eppendorf, Cambridge, UK), Labnet Multigene II (Labnet International, Inc), Geneamp 9600 thermal cycler (PerkinElmer, Beaconsfield, UK

Table 2.3. Standard PCR Master Mixture

PCR Master Mix	1 rxn (µl)	1 rxn (µl)
10X Taq Buffer	2.5	5.0
dNTP mix*	0.5	1.0
Primer 1 (1mM)	0.75	1.5
Primer 2R (1mM)	0.75	1.5
$MgCl_2$ (2.5 mM)	2.5	5.0
Taq polymerase (1.25 U)	0.25	0.5
SDW	17.25	34.5
Total Volume	24.0	49.0

*10µl of each dATP, dTTP, dCTP and dGTP (100mM each) were added to 60µl of SDW, rxn=reaction

2.4.2. Temperature Gradient PCR

Gradient PCR was carried out to determine the optimum annealing temperature for the designed primers. The Master Mix (Table 2.4.) was aliquoted into ten separate 0.5 ml Eppendorf tubes (24 µl each) and mixed with 1 µl of genomic tester DNA (Chelex suspensions or Wizard genomic DNA preparation). The reaction mix was subjected to 30 cycles of gradient PCR using the Eppendorf Mastercycler thermal cycler, with a temperature range from 48.1 to 66.9°C (Table 2.5.).

Table 2.4. PCR Master Mix for Gradient PCR

PCR Master Mix	1 rxn (µl)	11 rxn (µl)
10X Taq Buffer	2.5	27.5
dNTP mix (10mM, of each dATP, dTTP, dGTP, dCTP)	0.5	5.5
Primer XF (10µM)	0.75	8.25
Primer XR (10µM)	0.75	8.25
$MgCl_2$ (25µM)	2.5	27.5
Taq polymerase (1.25U)	0.25	2.75
SDW	11.75	129.5
Total Volume	29.0	319.0

Table 2.5. Annealing temperature range of Gradient PCR (Eppendorf Mastercycler gradient)

	temperature range (°C)									
columns	1	2	3	4	5	6	7	8	9	10
annealing temperature (°C)	48.1	48.8	50.3	52.3	54.8	57.4	60.2	62.8	65.1	66.9

2.4.3. Digoxigenin (DIG) labelling by PCR amplification

Digoxigenin-labelled PCR probes for genomic DNA blots were obtained using the standard 50 µl PCR protocol (Table 2.3.), but including 60 µM of digoxigenin-11-dUTP (Roche Diagnostics, Mannheim, Germany), adjusting the amount of SDW accordingly. A standard PCR reaction using the same primers but no digoxygenin was carried out alongside to produce a control amplicon. Both PCR products were subjected to electrophoresis on a 1.5% agarose gel.

2.4.4. Oligonucleotide primers

Oligonucleotide primers used for PCR assays to determine the distribution of SSH sequences and for other PCR amplifications are shown in Table 2.6. Primers were designed from identified sequences using the GenoSys oligomail program. Additional primers targeting novel prophage regions (Kingsley *et al*, 2009) are shown in Table 2.7.

Table 2.6. Summary of the oligonucleotide primers used, showing the primer sequence, amplicon sizes and annealing temperature.

Primer name	Primer sequence	Amplicon size (bp)	Annealing temp. (°C)*	Putative function of target
D23734-1F	ACGTTGCACAATCCGGAT	204	50	hypothetical
D23734-1R	CCACCGAAGAAGGAGCAA			
D23734-14F	GGCATCAAATTCACCAGC	301	52	autotransporter
D23734-14R	AGCGCTGGATAACGGTTC			
D23734-A1F	AAGTCGCAACGGCTGAGA	363	52	fimbrial protein StkD
D23734-A1R	GTAGTGGTGAAGTTCGTA			
D23734-A5F	CTGTGTCGCTATGTTTGC	187	52	fimbrial protein TcfA
D23734-A5R	CTCCTTCTCCTGTGGCATA			
D23734-B2F	CGAGGGTAAGTTAAAGGT	371	52	Type 2 restriction enzyme
D23734-B2R	AGTTTTCGCGCGCAGTCT			
D23734-B9F	TGCCATCCCAGATAACAA	300	50	conserved hypothetical
D23734-B9R	TTGTCTCATGTTGTATCC			
D23734-B10F	TTTACCGTTTGTGATCCA	332	50	lipoprotein
D23734-B10R	AGCAATCCTGGACAACTA			
D23734-D2F	AATGCATCGACAACCGTA	384	50	fhuA
D23734-D2R	GATAGTCATTACCGGTGC			
D23734-D4F	GCAGATCCACTAATTGAAG	417	52	TinR
D23734-D4R	GTTCTGCTTTCCTATGAC			
D23734-G3F	CTCACTGTTGCGCTACCCTTA	179	50	Fimbrial subunit
D23734-G3R	CAGTGGTTTTGCGATTGG			
D26104-11F	CCAGCTCTCATCTTCCCA	419	45-58	Tum phage 186
D26104-11R	GTGAGCACGTTATGATTGA			
D26104-ST1F	TGCCTTAACCTGTGCGTA	315	53-65	EaA phage 22
D26104-ST1R	GATAAGACGTTCGGCGAT			
D26104-ST2F	GCAGGCCAAAATCGTTCA	757	45-65	TnP phage P7

Primer name	Primer sequence	Amplicon size (bp)	Annealing temp. (°C)*	Putative function of target
D26104-ST2R	GTTACTAAATGCCCGTCA			
D26104-B6F D26104-B6R	TGTATCGCGTGGTGTTGA TAACGATGGCTGCGCGACA	514	59-65	Gp10 phage ST104
D26104-H11F D26104-H11R	TGCCATTGTGACAGCTCT CGCGGCTCATTTTGTTCA	649	51-55	G protein phage 186
D26104-9F D26104-9R	CGGGAGAAGGCTGAAACA GCGAGAGCTTCATTCACA	689	45-65	Colicin ColE1
D26104-C6F D26104-C6R	TCGCCCTTATTCCCTTTT CTCAACCAAGTCATTCTG	284	53-58	TEM-ESBL
D26104-C11F D26104-C11R	CAAGTCACAGAAAAGCAT GTTCATCCATAGTTGCCT	501	45-65	TEM-160
D26104-D7F D26104-D7R	CAATGCTTAATCAGTGA CGGTATCATTGCAGCACT	135	45-58	hypothetical protein
D26104-E11F D26104-E11R	AGTGGATGGCGGCCTGAA CAAATGCGGGACAACGTA	471	61-66	AadA1
D21685-3F D21685-3R	ACGAGCCGATGCCAGCCTT TACACTCCGCTATCGCTA	176	45-58	Rom-like protein
D21685-SE4F D21685-SE4R	TCTTCCACGGATAACCTT GAAAATGCCAGCCAGCCA	452	45-58	outer membrane protein
D21685-LasAF D21685-LasAR	GCCGTAGAATTATTAGCACA TGCGCGCAATACGGGATT	683	45-65	hypothetical
3114-F10F 3114-F10R	CCGCATAGGCATACACCA TGTCGATTGCCCGCGTCT	423	66-70	Gifsy-1 terminase Gifsy-1
3114-B6F 3114-B6R	TGGTGGCGTTCGTGGCAT ATATCGGGTGCGCTGGTT	556	66-68	tail fibre protein Phage P27
3114-A9F	GGGCAAGGCACATGGTA	561	56-64	phage terminase GpA

Primer name	Primer sequence	Amplicon size (bp)	Annealing temp. (°C)*	Putative function of target
3114-A9R	TTACCTGCGTGAACCGAT			
3114-G4F 3114-G4R	AGCCTCAAGCACCCGCAA CTGTATGCGGAAAATGA	186	50-61	hypothetical Stx2
3114-C12F 3114-C12R	GAAACTGAATCGGCCTACA GCTAATCTCCACCACACT	226	50-66	hypothetical
3114-11F 3114-11R	ACATTCACGCCCGCAGCCT GCGATTATTGGAAGCAGT	107	61-65	putative phosphotransferase
3114-18F 3114-18R	CTACGGCTATATCGGGGA AAAGCGGGAAGCGTCAAA	544	45-64	Peptidase, ClpP

*PCR primers worked equally well over this range of temperatures and the annealing temperature was varied to allow for the use of the same PCR machine.

Table 2.7. *S.* Typhimurium novel prophage primers (Kingsley *et al*, 2009); the annealing temperature for all primer pairs was 55°C.

Prophage	Primer sequence	Annealing site
BTP1 A-F BTP1 A-R	gcacgcagaaattacatgc tgacctgaggcacacgtg	anneal to sites spanning the left hand insertion site of BTP1
BTP1 B-F BTP1 B-R	ttccagtccatgaagcagc ctgaggtggaaggtcaggg	anneal to sites internal to BTP1
BTP2 D-F BTP2 D-R	gcagtacgaaccgtacccgatacag cagtccgtcacgccatgctcaaaact	anneal to sites internal to BTP2
BTP2 E-F BTP2 E-R	cagcagatgaatgggtttg tttaagtggacactatcatcgc	anneal to sites spanning sseI gene in BTP2, Gifsy and prophage 2 of DT104 prophages
DT104 F-F DT104F-R	actgcccaatacatcactcc cttgttgagccaggacaag	anneal to sites internal to prophage 3 of DT104
BTP3 I-F BTP3 I-R	gcggctggtagtctgagc ggtgcggcagcttatcag	anneal to sites internal to BTP3
BTP4 J-F BTP4 J-R	taataccgcgctcacgctcaagatc ctataacagtaccgggaactgttcg	anneal to sites spanning the left hand insertion site of BTP4
BTP5 O-F BTP5 O-R	ttcaactgtcgatcggtg gacgggcttagagcttgag	anneal to sites internal to BTP5
BTP6 Q-F BTP6 Q-R	cctgccacattaaccctg ggtcttctctattgttcaggtgg	anneal to sites internal to BTP6

2.5. Agarose Gel Electrophoresis

In order to estimate the amount of DNA isolated, typically 5 µl of the genomic DNA preparation were subjected to electrophoresis on a 1% agarose gel (1 g agarose (Melford) in 100 ml 0.5X TBE Buffer (Hybaid PS25 Electro 4) or 1X TBE Buffer (EmbiTec RunOne Electrophoresis Cell, San Diego, USA and Appleton Woods MultiSub Choice, Birmingham, UK)[10X TBE Buffer contains 108 g Tris (hydroxymethyl) methylamine (T/P630/60, Fisher Scientific, Loughborough, UK) , 9.3g EDTA (BDH, Poole, UK) and 55g Boric acid (B/3800/60, Fisher Scientific, Loughborough, UK) per litre]. A size marker (1kb Plus DNA Ladder, Invitrogen, Paisley, UK) was used on every gel throughout the project (Figure 2.1.).

1% agarose gel electrophoresis was used throughout the project for the visualization of PCR products. 1.5% agarose gels were used for purified DNA and 2% agarose gels were used for the visualization of primary and secondary PCR products during suppression subtractive hybridisation. 2 µl of ethidium bromide (10µg/ml, Ethidium Bromide Tablets, E-2515, Sigma, Steinheim Germany) was added to 100 ml of agarose prior to pouring the gel. Unless stated otherwise, 5 µl of sample were mixed with 1µl of 6x loading dye (0.25% bromophenol blue, 0.25% xylene cyanol FF and 30% glycerol, 6x DNA Loading Dye, R0611, Fermentas, St. Leon-Rot, Germany) and the mixture was loaded into the wells of the gel.

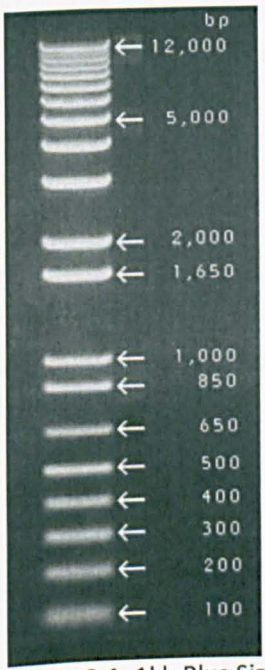


Figure 2.1. 1kb Plus Size Ladder (10787-018, Invitrogen)

2.6.Suppression subtractive hybridisation (SSH)

2.6.1. Restriction enzyme digestion

Genomic DNA isolated from strains G639, NCTC13349, D21685, D26104, D23734 and 3114 using the Wizard genomic DNA Purification Kit (A1120, Promega, Madison, USA) was digested with *Rsa*I (ER1122, Fermentas, St Leon-Rot, Germany) a restriction endonuclease which cuts double stranded DNA at 5'-G T^A C-3' producing blunt ends. 43.5 µl of genomic DNA were digested with 1.5 µl *Rsa*I in the presence of 5.0 µl 10X SuRE/Cut Buffer A (Roche Diagnostics) and incubated at 37°C for 5-16 h.

2.6.2 SSH

SSH was carried out using the PCR Select™ Bacterial Genome Subtraction Kit (Clontech, Mountain View, USA) and following the supplier's instructions. In the hybridisation steps *S.* Typhimurium LT2 (G639) was used as a driver for the tester strains *S.* Typhimurium (D26104), *S.* Heidelberg (D23734) and *S.* Bovismorbificans (3114). *S.* Enteritidis NCTC13349 was used as a driver for the tester *S.* Enteritidis (D21685).

Gel electrophoresis was carried out on a 2% agarose gel to compare primary and secondary PCR products of SSH.

2.6.3. Construction of subtraction libraries

The PCR amplicons derived from the SSH were cloned into the pGEM-T vector (Invitrogen) following the manufacturer's instructions. XL10-Gold® Kan^r Ultracompetent cells (Stratagene, Amsterdam, The Netherlands) were transformed with pGEM®-T following the supplier's transformation protocol (Catalog #200317). Varying volumes (50-200 µl) of transformation mix were plated onto LB plates containing ampicillin, X-Gal and IPTG for subsequent Blue/White screening. X-Gal is a colourless modified galactose sugar, which acts as a substrate for beta-galactosidase, the gene product of *lacZ*, and turns blue when broken down. Following disruption of the *lacZ* gene, cells containing an insert sequence are unable to produce beta-galactosidase and will remain white, while cells with no insert will turn blue.

Following overnight incubation, white colonies were picked up from the plates using sterile pipette tips and transferred onto a grid, drawn on a fresh Ampicillin/IPTG/X-gal plate and again incubated overnight at 37°C

2.6.4. 96-well plate cultures for plasmid extraction and sequencing

A 96-well plate was prepared using L-agar also containing ampicillin (100µg/ml) and inoculated with 96 separate subtracted clones. Each well was inoculated with a separate colony using a pipette tip. The colonies were subjected to plasmid extraction and sequencing using a vector primer (Lark Technologies).

2.6.5. Sequence Data Analysis

Subtraction sequences were edited using Chromas (Version 1.45, 32-bit) and BioEdit (v 7.0.9), and were used to search the NCBI BLASTN microbes sequence database for *S.* Typhimurium LT2, and the Wellcome Trust Sanger Sequence Search engine (http://www.sanger.ac.uk/Projects/Microbes/) for *S.* Enteritidis PT4 in order to exclude unsubtracted sequences. The NCBI BLASTX Protein Database (http://www.ncbi.nih.gov/ BLAST/) was used to determine the putative function of the isolated sequences.

2.7. DNA –DNA hybridisation

2.7.1 Dot blot analysis

2.7.1.1. Preparation of membranes

Multiple membranes were prepared at the same time. For four membranes 20 µl of high quality, concentrated genomic DNA was mixed with 2µl ($1/10^{th}$ of the volume) of 1M NaOH in a 1.5 ml Eppendorf tube, vortexed and subsequently centrifuged at 13000 rpm for 10 s. 5 µl of this denatured DNA was dotted onto Hybond-N-membrane (Amersham Biosciences/GE Healthcare) cut to size and placed on 3MM Whatman paper. The DNA was added slowly, thereby producing a small dot. This was repeated with all the DNA samples, taking care to spread them out evenly, producing a grid. The order of the samples was noted on a sheet of paper and the membrane was marked in a way to indicate the orientation of the grid. The DNA was then allowed to air dry and adhere to the membrane. The membranes were then placed onto Whatman paper soaked in 0.5 x SSC (20x Stock solution: sodium chloride 175.32 g/l, sodium citrate 88.23 g/l, pH to 7.0 (BDH)) for 2 min. After 2

min the membranes were placed onto fresh 3 mm Whatman paper and allowed to air dry for a few minutes before being covered in more filter paper and placed into an oven for 2 h at 80°C. Membranes can then be stored inside an envelope at 4°C until further use.

2.7.1.2. Pre-hybridisation

Nylon membranes to be hybridised were submerged in 2 x SSC with a slightly bigger piece of mesh placed on top. The membrane with the mesh on the inside were rolled and together with some 2 x SSC placed inside a glass hybridisation tube, allowing the membrane to unravel and adhere to the inside of the tube. Multiple membranes can be placed inside the same tube when separated by a mesh. The 2x SSC was then replaced by 50 ml of pre-hybridisation solution (6 x SSC and 2% blocking reagent, Roche Diagnostics) The membranes were prehybridised at 68°C for 3 h, with constant rotation in a Robbins Scientific Model 2000 Hybridisation Incubator.

2.7.1.3. DNA-DNA hybridisation

After 3 h prehybridisation, 30 ml of the pre-hybridisation solution were removed. The DIG labelled probe, produced by PCR amplification, was heated at 99°C for 10 min in a thermal cycler to denature the DNA and subsequently added to the remaining 20 ml of pre-hybridisation solution. The bottles were returned to incubation rotating at 68°C overnight. The following day the probe solution was removed. It can be stored in a sterile universal at -20°C and can be reused.

2.7.1.4. Detection of Hybridisation

Washes

In order to detect specific attachment of the probe, several washes were performed in order to remove any non-specific binding of the probe to the membrane. 50 ml of pre-heated (68°C) washing buffer were used in each wash. All the washes were performed rotating at 68°C

The following washes were carried out:
2 x 15min in 6 x SSC, 0.1% SDS
2 x 15min in 2 x SSC, 0.1% SDS
2 x 15min in 0.2 x SSC, 0.1% SDS
2 x 5min in 0.1 x SSC, 0.1% SDS

Detection of DIG

Hybridisation was detected by using anti-DIG-alkaline phosphatase (AP) Fab fragments (Roche Diagnostics) and the chemiluminescent substrate CDP-star/CSDP (Roche Diagnostics). The AP conjugated antibody binds to the hybridisation probe and the antibody-probe hybrids are visualized by alkaline phosphatase substrates which are converted to chemiluminescent products. The membrane was removed from the tube and all subsequent steps were carried out at room temperature on a shaker.Buffer 1 (0.1M Tris pH7.5, 0.15M NaCl pH to 7.5 (can be made to 10X strength and then diluted) acts as a bases for all the buffers required.

The membranes were rinsed for two rounds in washing buffer (0.3% (v/v) Tween 20 in buffer 1) before incubation in 80 ml buffer 2 (1% (w/v) blocking reagent in buffer 1 (requires heating to obtain proper suspension)) for 30 min. After 30 min buffer 2 was replaced by anti-DIG-AP diluted 1:10000 in 20 ml of buffer 2. The membranes were again incubated for 30 min, in the antibody solution. Antibody binding non-specifically was removed through a series of 4 x 8 min washes in washing buffer. The membranes were equilibrated in 50 ml of buffer 3 0.1M Tris pH9.5, 0.1M NaCl, 50mM $MgCl_2$ (12.11g Tris, 5.84g NaCl, 10.16g $MgCl_2$ per litre) before incubation with the substrate CSDP (CDP*). This incubation step lasted 3 min taking care that the membrane was covered evenly with the diluted substrate. The membrane was then removed and wrapped in clingfilm. To detect the chemiluminescence the membranes were exposed to X-ray film (Kodak, Rochester, NY, USA) for up to 25 min, depending on the substrate (CSDP or CDP*). It is advisable to start with very short exposure times (i.e. 10 s.). The film was then placed into a developing bath (GBX developer and replenisher, Kodak) with shaking, followed by fixation (GBX fixer and replenisher, Kodak) for approximately 2 min each. Finally the film was rinsed in water and allowed to air dry (Details of all the solutions are given at the end of section 2.7).

2.7.2. Southern Blotting

Plasmid DNA isolated from strain D26104 using the Wizard the Quiagen Mini Prep Kit (Quiagen) was digested with *Rsa*I.

DNA samples were loaded on to a 1% agarose gel. The agarose gel was photographed alongside a ruler as a reference so that distances can be related back to the original gel at the end of the procedure. The gel was then submerged in 0.25 M

HCl (nicking / depurination) for 15 min and subsequently rinsed in SDW. The gel was then submerged in 0.5 M NaOH / 1.5 M NaCl (denaturing) for 2 x 15 min, then rinsed in SDW. The gel was finally submerged in 0.5 M Tris / 3 M NaCl pH 7.4 for neutralising for 2 x 30 min. The blot was then assembled as shown in Figure 2.2. The assembled blot was left overnight to enable transfer of DNA to the nylon membrane. After overnight blotting, the nylon membrane was rinsed in 2 x SSC, allowed to air dry and then baked for 2 hr at 80°C to fix ssDNA to the membrane.

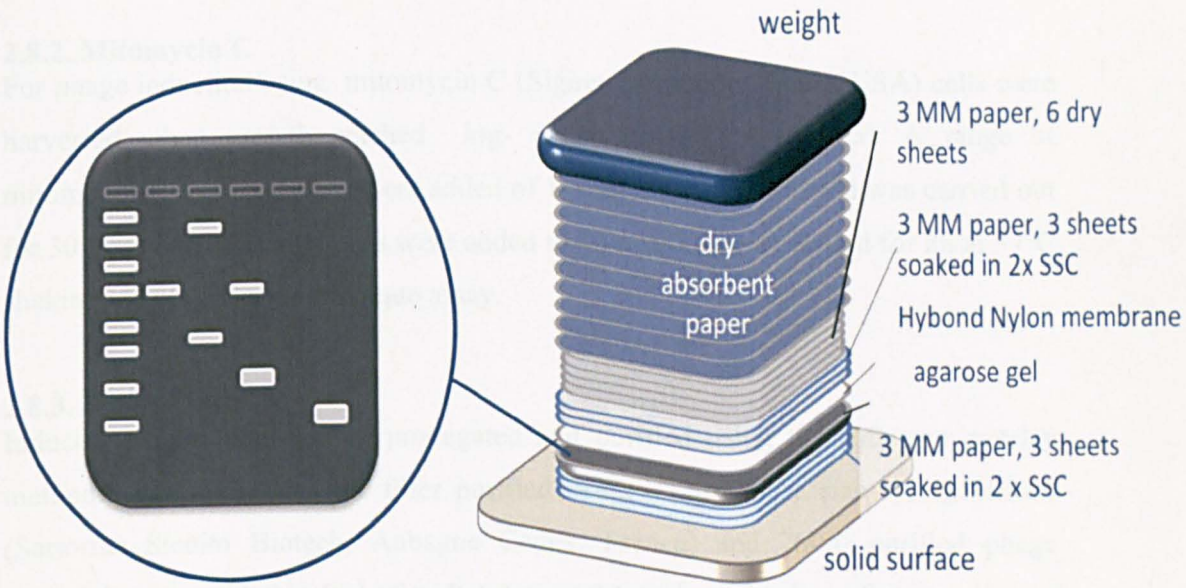


Figure 2.2. Schematic of Assembly of a Southern Blot

Prehybridisation, DNA-DNA hybridisation and detection were carried out as described in Dot Blot above.

2.8. Phage induction and Plaque assays

2.8.1. UV induction

Phage induction was performed on *S. enterica* Typhimurium D26104 and Bovismorbificans 3114. For phage induction using UV radiation, cultures were harvested when growth reached an OD600 = 0.4-0.5. Cultures were transferred to Petri dishes and exposed to long-wavelength UV light for 30-60 s. 1ml of treated cells were added to 10ml LB incubated for 2h at 37C shaking at 200rpm prior to plaque assay.

2.8.2. Mitomycin C

For phage induction using mitomycin C (Sigma-Aldrich, St-Louis, USA) cells were harvested when growth reached log- phase (OD600 = 0.4-0.5). A range of mitomycin C concentrations were added of 1-3 ug/ml and incubation was carried out for 30 min. 1ml of treated cells were added to 10ml LB and incubated for 2h at 37°C shaking at 200rpm prior to plaque assay.

2.8.3. Plaque assays

Induced phages were further propagated and purified using the soft agar overlay method: 1ml of culture was filter purified using < 2μm pore size syringe filters (Sartorius Stedim Biotech, Aubagne Cedex, France) and 200μl purified phage suspension was added to 5ml of molten top-agar together with 1ml of suspensions of recipient strains *S.* Typhimurium LT2 or D23580. The top-agar was then poured onto plates containing L-agar and incubated overnight at 37°C. Phage was isolated from plaques by inserting a straight wire into the plaque and suspending the phage in SDW.

2.8.4 DNase I treatment

In order to remove any unwanted genomic DNA, phage suspensions were subjected to DNase I treatment. In a sterile 1.5 ml Eppendorf tube the following were added: 25μl phage suspension, 3.5μl buffer, 3.5μl DNase I (X Units) (Roche Diagnostics) and 3μl SDW. The mixture was incubated at room temperature for 15 min, after which 1μl of 0.5M EDTA (Sigma) was added and a further incubation at 65°C for 10 min was carried out. The sample can be stored at -20°C.

2.9. Preparation of cDNA for reverse transcription (RT) PCR amplification

A subset of 6 *S.* Heidelberg strains was used to determine the expression of SSH sequences.

Cells were grown overnight in Luria Broth at 37°C shaking at 200rpm. Cells were collected by centrifugation of 1ml of overnight culture at 13000 rpm for 2 min. RNA extraction was performed using the RiboPure^{TM} Bacteria whole RNA isolation kit (Ambion, Applied Biosystems) following the manufacturers instructions, except that two rounds of the recommended DNase I treatment of the RNA samples were performed. The RNA was converted to cDNA using the SuperScript II RT kit (Invitrogen). For each RNA preparation, a control reaction lacking the reverse transcriptase was also prepared. We confirmed that expression of the flagellin gene (*fliC*) was detectable in all cDNA samples but not in any of the controls.

2.10. Multilocus Sequence Typing (MLST)

BigDye sequencing reactions were carried out following the manufacturer's instructions (Applied Biosystems http://www.ibt.lt/sc/files/BDTv3.1_Protocol_04337035.pdf) Standard PCR reactions (see 2.4.1.) were performed cycling at 94°C for 1 min, 55°C for 1min and 72°C for 2 min. Both PCR and Sequencing reactions were performed on an Eppendorf Mastercycler gradient. Primer sequences summarized in Table 2.8. were obtained from the MLST website (http://mlst.ucc.ie/mlst/dbs/Senterica/documents/primersEnterica_html.) Samples were send to Oxford University Sequencing Facility, Zoology Department for Gel Electrophoresis.

Table 2.8 MLST primers (www.mlst.net)

Amplification		Product Size (bp)
thrA-F	5'-GTCACGGTGATCGATCCGGT-3'	852
thrA-R	5'-CACGATATTGATATTAGCCCG-3'	
purE-F	5'-ATGTCTTCCCGCAATAATCC-3'	510
purE-R	5'-TCATAGCGTCCCCCGCGGATC-3'	
sucA-F	5'-AGCACCGAAGAGAAACGCTG-3'	643
sucA-R	5'-GGTTGTTGATAACGATACGTAC-3'	
hisD-F	5'-GAAACGTTCCATTCCGCGCAGAC-3'	894
hisD-R	5'-CTGAACGGTCATCCGTTTCTG-3'	
aroC-F	5'-CCTGGCACCTCGCGCTATAC-3'	826
aroC-R	5'-CCACACACGGATCGTGGCG-3'	
hemD-F	5'-ATGAGTATTCTGATCACCCG-3'	666
hemD-R	5'-GAAGCGTTAGTGAGCCGTCTGCG-3'	
dnaN-F	5'-ATGAAATTTACCGTTGAACGTGA-3'	833
dnaN-R	5'-AATTTCTCATTCGAGAGGATTGC-3'	
Sequencing		
thrA-sF	5'-ATCCCGGCCGATCACATGAT-3'	
thrA-sR	5'-CTCCAGCAGCCCCTCTTTCAG-3'	
purE-sF	5'-CGCATTATTCCGGCGCGTGT-3'	
pure-sR	5'-CGCGGATCGGGATTTTCCAG-3'	
sucA-sF	5'-AGCACCGAAGAGAAACGCTG-3'	
sucA-sR	5'-GGTTGTTGATAACGATACGTAC-3'	
hisD-sF	5'-GTCGGTCTGTATATTCCCGG-3'	
hisD-sR	5'-GGTAATCGCATCCACCAAATC-3'	
aroC-sF	5'-GGCACCAGTATTGGCCTGCT-3'	
aroC-sR	5'-CATATGCGCCACAATGTGTTG-3'	
hemD-sF	5'-GTGGCCTGGAGTTTTCCACT-3'	
hemD-sR	5'-GACCAATAGCCGACAGCGTAG-3'	
dnaN-sF	5'-CCGATTCTCGGTAACCTGCT-3'	
dnaN-sR	5'-CCATCCACCAGCTTCGAGGT-3'	

2.11. Cell invasion Assays

Salmonella isolates were grown in 10ml L broth at 37°C in an orbital shaking incubator at 150rpm overnight. The invasiveness of the strains into nonphagocytic host cells was determined using Rabbit kidney cells (RKC). RKC were seeded to 10^6 cells/ml in 24-well plates, the growth media (DMEM - Dulbecco's Modified Eagles Media, Invitrogen, Carlsbad, USA) was removed and fresh media containing no antibiotic was added and cells were incubated for 2 hrs. Bacteria were added at a multiplicity of infection (MOI) of 10 and incubated for 1hr at 37 °C. The growth media was removed and 1ml of media containing gentamicin sulphate (100μg/ml) was added to each well, cells were than incubated for a further 1 hr at 37°C. Cells

were washed twice with Phosphate Buffered Saline (PBS) (Life Technologies). RKC cell lysis was achieved by adding 1 ml of PBS with 0.5% triton-x (Sigma-Aldrich) to each well. In a 96-well plate serial dilution of 200µl of sample was performed in triplicate for each bacterial strain, In replicates of three, 20µl of each dilution was added to L-agar plates and incubated overnight at 35°C. Differences in invasiveness were assessed by variance analysis using the SPSS statistical package.

2.12. Genome sequencing

2.12.1. 454 genome sequencing

S. Bovismorbificans strain 3114 was sequenced at the Advanced Genome Facility, University of Liverpool, using the Roche 454 Genome Sequencer FLX (GS-FLX) following the manufacturer's instructions (Roche 454 Life Science, Branford, CT, USA). In brief, each sample was made into both a paired-end and fragment library using the standard FLX chemistry for 454. Fragment libraries were prepared by fragmentation, attachment of adapter sequences, refinement of the ends and selection of adapted molecules. Paired-end libraries were produced by hydroshear shearing, circularisation, addition of adapters and selection as for the fragment library. Both libraries were amplified by emPCR and fragment-containing beads recovered and enriched. Sequencing primers were added and each library was deposited onto a quarter of a PicoTiterPlate plate and sequenced. 454 sequencing was performed by Dr M. Hughes, Biological Sciences, University of Liverpool. Figure 2.3. shows a schematic representation of 454 sequencing.

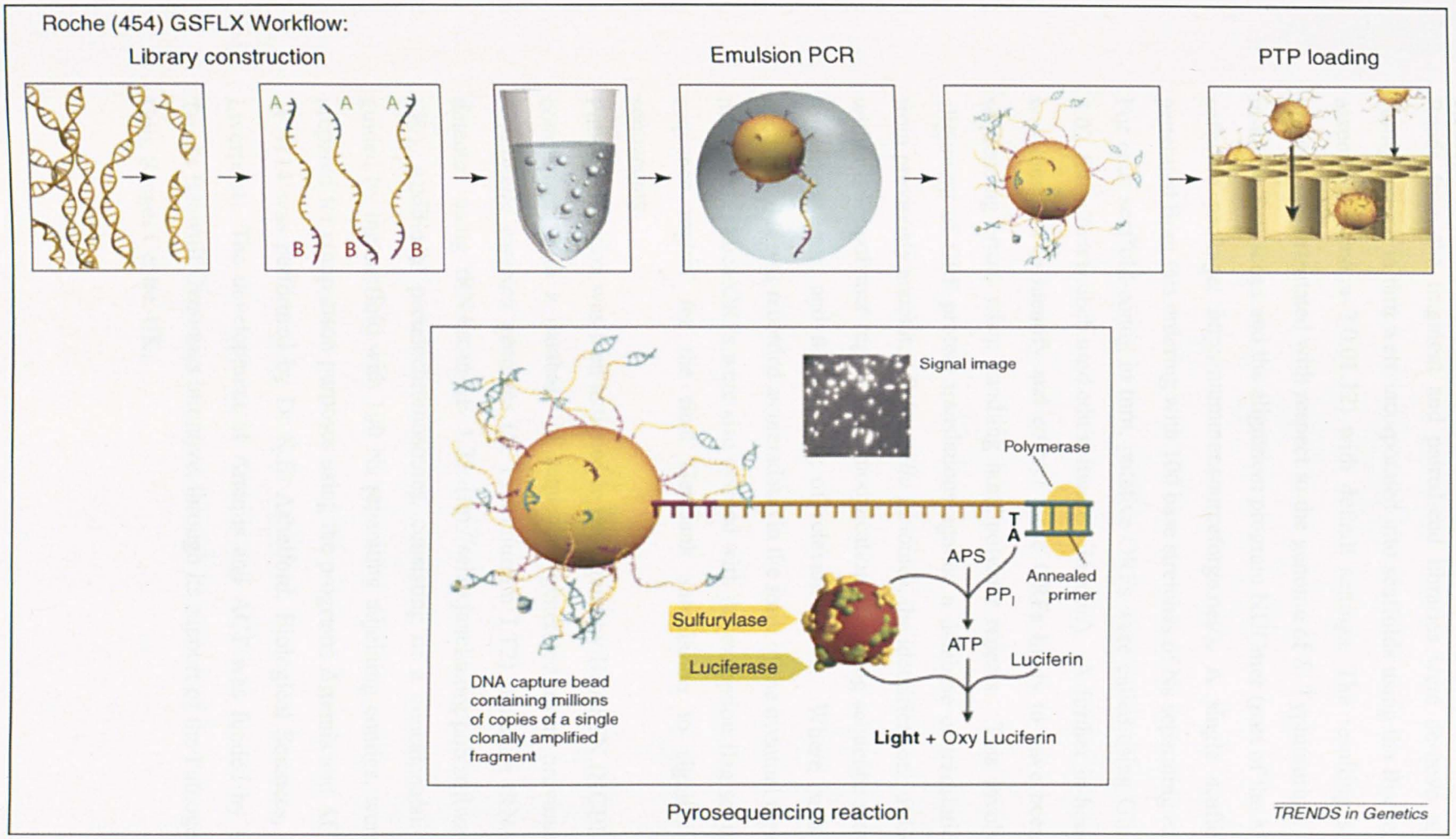


Figure 2.3 Diagrammatic representation of the Roche 454 genome sequencing process (Mardis, 2008)

2.12.1.2. Genome assembly and annotation

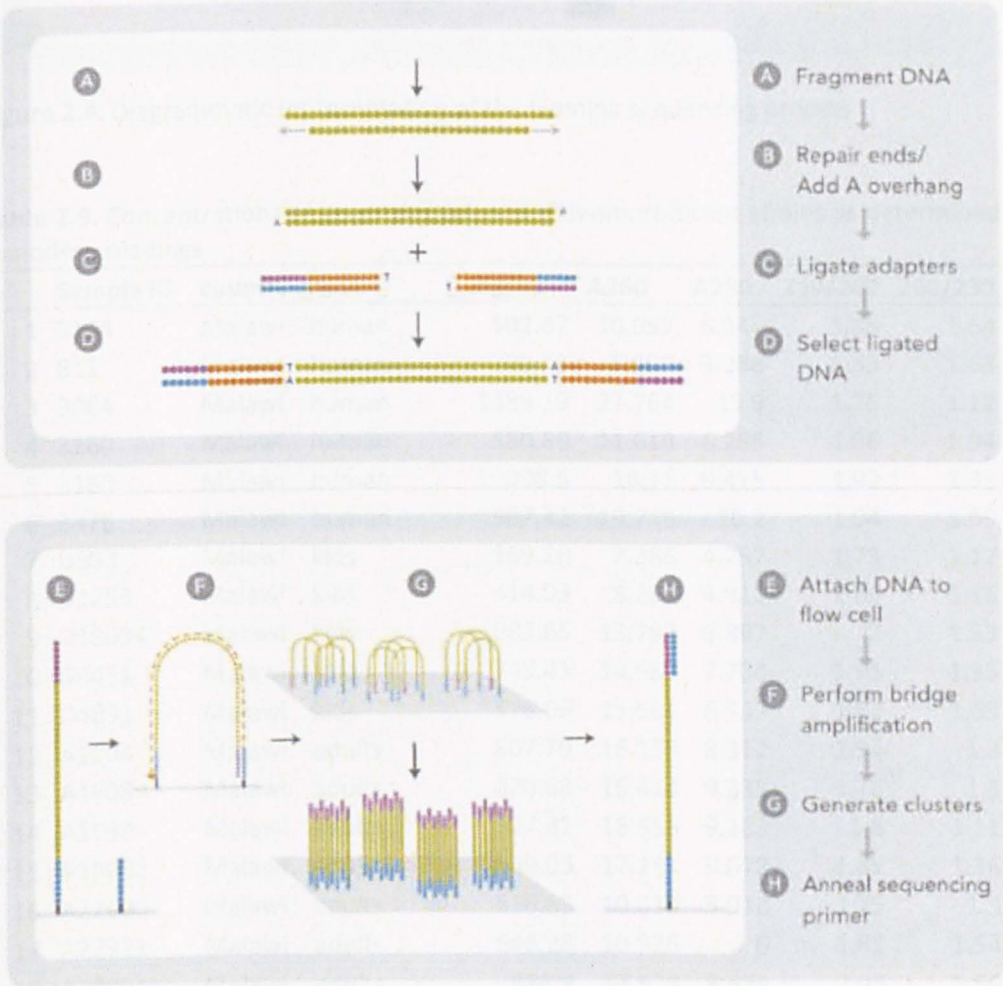
Reads from the fragment and paired-end libraries were de-novo assembled into contigs which in turn were incorporated into scaffolds using the Roche 454 Newbler assembler (version 2.0.01.12) with default settings. The resulting scaffolds were ordered and orientated with respect to the genome of *S.* Typhimurium LT2 using an in-house Perl script and the alignment program NUCmer (part of the MUMmer 3.20 software package; http://mummer.sourceforge.net/). A single scaffold was then generated from this ordering with 100 base stretches of Ns separating each scaffold.

For each scaffold-contig in turn, putative ORFs were called using Glimmer version 3.02 (http://www.cbcb.umd.edu/software/glimmer/). A further in-house Perl script was then run to identify and correct those ORFs likely to have been split due to sequencing errors when handling homopolymer repeats. This involved BLASTP alignment of ORF protein translations against a database of translations generated from previously annotated *Salmonella* genomes, the identifications of likely INDELS within homopolymer regions, the modification of coding sequence feature positions to correct error, and the merging of relevant ORFs. Where such modification occurred, this was recorded as metadata (in the form of the eventual GenBank feature note field). Such ORFs were also marked with the exception flag set to 'low-quality sequence region' for the final GenBank submission to signify poor quality sequencing.

Putative function was then assigned to each gene by BLASTN (NCBI Blast 2.2.17) comparison with a database of sequences generated from previously annotated *Salmonella enterica* genomes (*S.* Typhimurium LT2). Putative tRNA genes were detected using tRNAscan-SE 1.23 (ftp://selab.janelia.org/pub/software/tRNAscan-SE/). Additional pseudochromosomes, consisting of a concatenatation of contigs guided by the scaffold with 100 Ns separating adjoining contigs, were additionally prepared for comparison purposes using the programs Artemis and ACT. Assembly of 3114 was performed by Dr K.E. Ashelford, Biological Sciences, University of Liverpool). The development of Artemis and ACT was funded by the Wellcome Trust's Beowulf Genomics initiative, through its support of the Pathogen Sequencing Unit, Sanger Centre, UK.

2.12.2. Solexa genome sequencing

Genomic DNA was prepared using the Wizard Genomic DNA Kit (Promega) according to the manufacturer's instructions. 25 *S.* Bovismorbificans strains (Table 2.9.) were sequenced using the Illumina/Solexa Genome Analyzer System according to the manufacturer's specifications. Sequencing and assembly was carried out by Theresa Feltwell, Ruth Guilderthorp and personnel at the Wellcome Trust Sanger Institute. In all cases single-end reads were generated. Figure 2.4. summarizes the process of Illumina sequencing



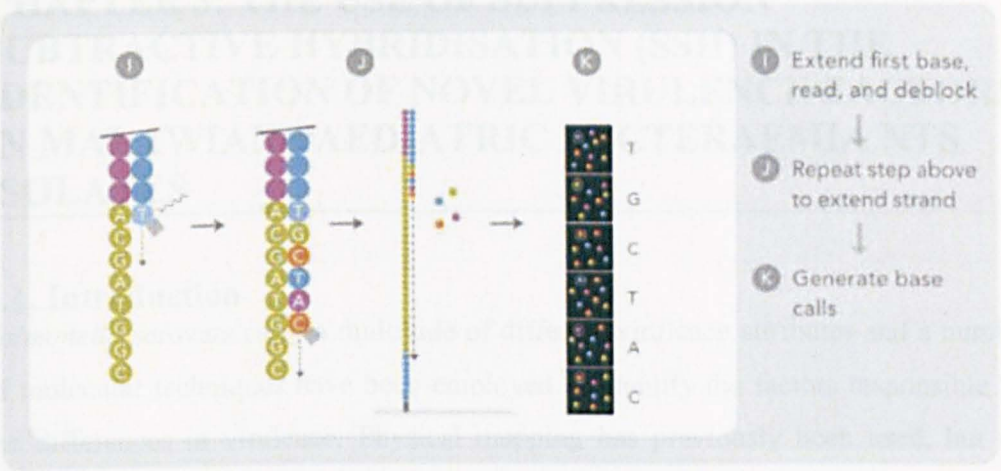


Figure 2.4. Diagrammatic representation of the Illumina sequencing process

Table 2.9. Concentrations of genomic DNA of *S.* Bovismorbificans strains as determined by Nanodrop readings

	Sample ID	country	host	ng/ul	A260	A280	260/280	260/230
1	3114	Malawi	human	502.87	10.057	5.344	1.88	1.64
2	811	Malawi	human	300.43	6.009	3.288	1.83	1.63
3	3064	Malawi	human	1388.19	27.764	15.9	1.75	1.12
4	3160	Malawi	human	580.89	11.618	6.255	1.86	1.94
5	3180	Malawi	human	908.5	18.17	9.475	1.92	1.35
6	3476	Malawi	human	987.42	19.748	10.2	1.94	1.59
7	D993	Malawi	kids	369.28	7.386	4.257	1.73	1.17
8	D1253	Malawi	kids	414.03	8.281	4.911	1.69	1.16
9	D19094	Malawi	kids	689.65	13.793	6.897	2	1.53
10	D4451	Malawi	kids	749.43	14.989	7.704	1.95	1.35
11	D4891	Malawi	kids	578.06	11.561	6.337	1.82	1.69
12	A1104	Malawi	adults	807.79	16.156	8.317	1.94	1.3
13	A1608	Malawi	adults	820.68	16.414	9.338	1.76	1.3
14	A1668	Malawi	adults	827.81	16.556	9.183	1.8	1.11
15	A16982	Malawi	adults	859.05	17.181	9.072	1.89	1.36
16	A22920	Malawi	adults	516.63	10.333	5.918	1.75	1.3
17	A22921	Malawi	adults	546.28	10.926	6	1.82	1.54
18	A24091	Malawi	adults	878.9	17.578	8.673	2.03	1.59
19	A31126	Malawi	adults	503.27	10.065	5.7	1.77	1.39
20	A5893	Malawi	adults	1203.39	24.068	13.26	1.82	1.19
21	A8737	Malawi	adults	1111.79	22.236	11.2	1.98	1.46
22	A93621	Malawi	adults	473.35	9.467	5.182	1.83	1.51
23	2766/08	UK	vet. - pig	560.63	11.213	6.215	1.8	1.4
24	4992/08	UK	vet. - alpaca	1010.49	20.21	11.46	1.76	1.03
25	518927/76	UK	vet. - pig	938.02	18.76	9.395	2	1.7
26	6533/08	UK	vet. - pig	1002.58	20.052	10.1	1.98	1.51

CHAPTER 3: THE USE OF SUPPRESSION SUBTRACTIVE HYBRIDISATION (SSH) IN THE IDENTIFICATION OF NOVEL VIRULENCE FACTORS IN MALAWIAN PAEDIATRIC BACTERAEMIA NTS ISOLATES

3.1. Introduction

Salmonella serovars carry a multitude of different virulence attributes and a number of molecular techniques have been employed to identify the factors responsible for the differences in virulence. Physical mapping has previously been used, but the major drawback of this technique is that it only identifies large insertions or deletions (Rode *et al.*, 1999). *Salmonella enterica* includes a large number of serovars that cause disease in animals and humans. Previous hybridisation experiments have shown that *S. enterica* serovars share >90% of DNA content, and *Salmonella* shares 80-85% with the closest relative *E .coli* (Boyd *et al.*, 1996; Crosa *et al.*, 1973). The sequence difference between *Salmonella* serovars therefore accounts for 500-600 kb. Pair-wise comparison of *Salmonella* genomes showed the presence of insertions and deletions; most of these unique DNA regions range in size from 1 to 50kb (Edwards *et al.*, 2002). This chapter will focus on using Suppression Subtractive Hybridisation to investigate the accessory genome of NTS bacteraemia isolates from Malawi by targeting these unique regions.

3.1.1. Subtractive Hybridisation (SH)

The concept of using subtractive hybridization to isolate deleted sequences was developed by Bautz and Reilly, who used DNA from a bacteriophage T4 deletion mutant to isolate mRNAs from the deleted region (Bautz & Reilly, 1966).

In 1990 Straus and Ausubel introduced subtraction hybridisation as a time saving alternative technique to genome walking and cloning for identifying specific genetic sequences linked to a phenotype. The goal of genomic subtraction is to isolate wild type DNA that is absent from a deletion mutant. DNA of the deletion mutant (the driver) is needed in excess; the DNA is sheared and biotinylated, before denaturation in the presence of a small amount of *Sau*3A-digested wild type (tester) DNA. A schematic representation is shown in Figure 3.1 The DNA mixture is allowed to hybridise and the majority of wild-type DNA will hybridise with complementary biotinylated driver DNA strands, however wild type strands that have no

corresponding driver DNA will not hybridise. In the next step biotinylated DNA is removed by incubating the DNA mixture with avidin-coated polystyrene beads, while unbound DNA is collected. The unbound DNA that is now enriched for the presence of sequences absent from the mutant is now mixed with a fresh batch of biotinylated deletion mutant DNA. Again the mixture is denatured, renatured and depleted of biotinylated sequences. Several rounds of this subtraction are carried out and the remaining DNA is amplified using PCR; this is made possible by ligating adaptors with *Sau*3A-compatible ends to the *Sau*3A-digested wild type sequences. In a final round of polymerisation, radioactive nucleotides are incorporated into the amplified DNA, the resulting labelled DNA can then be used as a probe to screen a wild type DNA library. In order to demonstrate the validity of their test Straus and Ausubel used a yeast mutant with a 5kb deletion which accounts for $1/4000^{th}$ of the yeast genome, and subtracted it against the wild-type version of TD33.3. They required 4-5 rounds of subtraction in order to identify the right clone. (Straus & Ausubel, 1990).

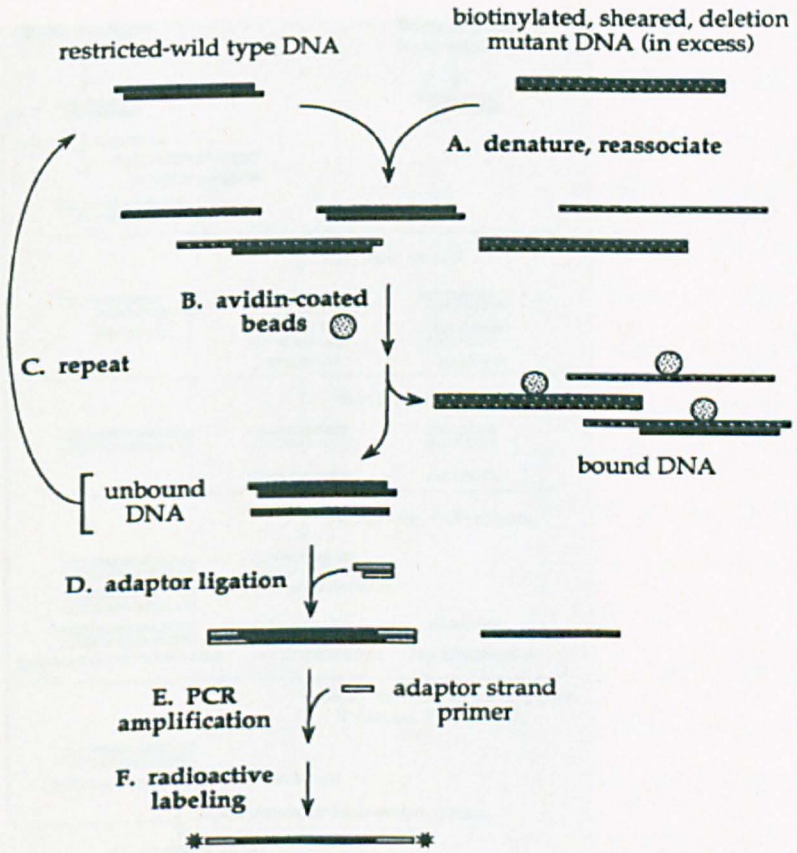


Figure 3.1. Schematic representation of subtractive hybridisation (SH) (Straus & Ausubel, 1990)

3.1.2. Representational Difference Analysis (RDA)

A modification of subtractive hybridisation termed representational difference analysis (RDA) was developed by Lisitsyn *et al* in 1993. RDA addressed the need for enrichment of the sample in order to isolate smaller differences between two DNA populations. Enrichment is achieved by kinetic amplification using oligonucleotide adaptors. DNA is cut with infrequently cutting restriction enzymes and adaptors are ligated to the fragments and amplified by PCR, which will enrich the sample with sequences below 1kb in size. The major advantage is that only a small amount of starting material is needed (less than 1µg). Once amplicons of both tester and driver DNA was made the adaptors were removed by cleaving and only tester sequences were ligated to new adaptors. A single hybridisation-amplification step now selects for tester specific sequences without the need for physical separation. During this second round of amplification only tester specific sequences that will have re-annealed will have an adaptor at each 5'end, to which the PCR oligonucleotide primers are targeted, a schematic of the hybridisation and amplification steps is shown in Figure 3.2 (Lisitsyn *et al.*, 1993).

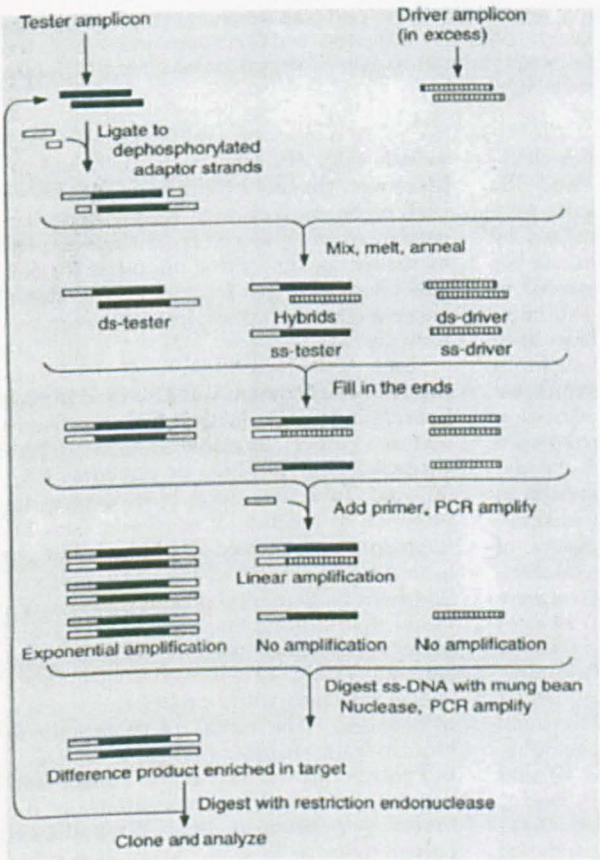


Figure 3.2. Schematic diagram of RDA , illustrating the hybridisation and amplification steps (Lisitsyn *et al.*, 1993)

3.1.3. Suppression Subtractive Hybridisation (SSH)

Suppression Subtractive Hybridisation (SSH) was developed by Diatchenko as a variation of suppression polymerase chain reaction (PCR) (Luk'ianov *et al.*, 1996), which uses adaptor ligated to the target DNA fragment as PCR primer target thereby making knowledge of the actual sequence unnecessary. SSH is the logical amalgamation of suppression PCR and subtractive hybridisation. SSH therefore combines two major events: normalisation and subtraction. The normalisation step equalizes the abundance of DNA within the target population; the subtraction step eliminates sequences common to both populations. Figure 3.3 shows a schematic representation of a general SSH. The two strains are allocated into tester and driver. The driver strain is the control strain, whose genomic contents will be subtracted from the tester strain, leaving only DNA sequences unique to the tester. The tester DNA is digested with a restriction endonuclease such as *Rsa*I. It is subsequently divided into two equal quantities. A ligation reaction is carried out whereby a different type of adaptor is attached to the 5' ends of the sequences. Both portions are separately hybridised to the driver DNA which is in excess to ensure that all the sequences also present in the tester are bound. The first hybridisation results in the presence of tester-specific single stranded sequences amongst a pool of double-stranded sequences. A second hybridisation step is then carried out in which the two portions are mixed. Homologous single-stranded DNA will hybridise and sequences specific to the tester will carry two different adaptors at the ends. PCR primers are specifically designed to the adaptors and will therefore only amplify sequences carrying two different adaptors. Sequences that have previously remained single-stranded will form secondary structures and become unavailable for amplification. PCR amplicons can be used to create a subtraction library, by cloning the sequences into a suitable vector. The procedure is not 100 per cent effective but >50% of clones should be tester specific (Diatchenko *et al.*, 1996 reviewed by Winstanley, 2002).

Diatchenko and co-workers initially used the technique for comparison of differentially expressed genes in eukaryotes, by reverse transcribing messenger RNA into complementary DNA (cDNA). Since then SSH has found its application in prokaryotic genetics as an elegant method of comparing bacterial genomes and to identify genes and sequences present in some bacteria but absent in other members of the species/genus/family (Winstanley, 2002). These genes may determine strain-specific characteristics such as drug resistance, bacterial surface structure or

restriction modification. SSH is particularly applied in the identification of pathogenicity islands (PAI) present in virulent strains but absent from non-pathogenic strains. The first bacterium used in a SSH study was *Helicobacter pylori* in 1998. The study initially identified the *cag* PAI, a type IV secretion system, by comparing a tester strain with a strain where *cag* had been deleted. They then went on to compare the genomic contents of two strains of *H. pylori*, one of which has been fully sequenced (Akopyants *et al.*, 1998). Since then a large number of bacterial genomes have been studied using SSH including *Salmonella* Typhi and Typhimurium. Table 3.1 summarises a number of *Salmonella* SSH studies.

Although 2nd and 3rd generation sequencing have led to considerably reduced costs for whole genome sequencing as an alternative to SSH, it is likely that SSH will still be useful as a cheap technique to identify strain- or serovar- specific sequences for microarray studies (Bae *et al.*, 2005; Kang *et al.*, 2006).

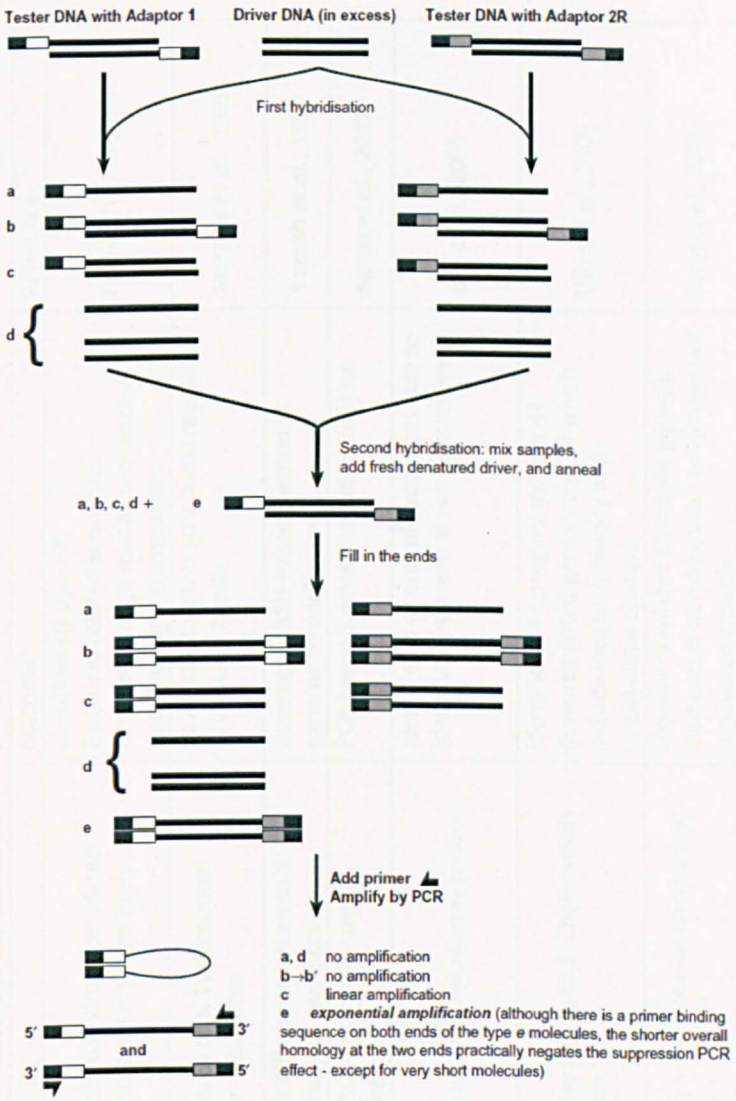


Figure 3.3. Schematic representation of PCR-Select suppression subtraction hybridisation. (Clontech PCRSelect™ Bacterial Genome Subtraction kit User Manual). 'Fill in the ends' refers to filling in the missing strands of adaptor sequences by pre-incubation at 72°C in the presence of *Taq* polymerase and nucleotides.

Table 3.1. The use of SSH on a number of *Salmonella* isolates in different experiments

Tester	Driver	Year	aim	outcome	Reference
S. Typhimurium	*S.* Typhi Ty2	1999	Typhimurium virulence related sequences absent from Typhi	identified *stf* operon F-plasmid related sequences lambdoid phage related sequences putative sugar transporters	Emmerth *et al.*, 1999
S. Typhimurium	*S.* Typhi	1999	expression of *S.* Typhimurium specific sequences	novel putative transcriptional regulator of the LysR family	Morrow *et al.*, 1999
S. Typhimurium E. coli	E. coli *S.* Typhimurium	1999	genetic difference between *S.* Typhimurium and E.coli	developing SSH-based method as genome screening tool	Bogush *et al.*, 1999
S. Enteritidis	*S.* Dublin	2001	identify *S.* Enteritidis specific sequences	PCR based S. Enteritidis identification	Agron *et al.*, 2001
various *S.* Choleraesuis subsp. *S.* Cholerasuis subsp Choleraesuis	*S.* Choleraesuis subsp Choleraesuis *S.* Typhimurium	2005	construction of microarray probes	identified a cluster of sequences able to identify *Salmonella* at sub-species level	Bae *et al.*, 2005
cDNA infected	cDNA uninfected	2006	response in pigs to *S.* Choleraesuis infection	identified 44 upregultated and 44 downregulated genes, roles of which include innate immunity and cytoskeletal changes	Uthe *et al.*, 2006
S. Typhimurium DT104 *S.* Newport *S.* Typhimurium DT160	*S.* Typhimurium LT2	2006	identify sequences for microarray probes	sequences related to mobile genetic elements that may confer advantages of survival or growth	Kang *et al.*, 2006
S. Dublin	*S.* Gallinarum	2008	Identify genetic bases of different virulence of Dublin and Gallinarum	Identified 21kb SDI-1 (*S. enterica* serovar Dublin Island 1), 24 mobile genetic element related sequences	Pullinger *et al.*, 2008

3.1.4. Clinical Information of Malawian NTS chosen for SSH

In this study, SSH was carried out on four Malawian paediatric bacteraemia isolates, that originated from HIV negative patients. The case-fatality rate of children with NTS bacteraemia in Malawi was around 24% and over 50% for children with NTS meningitis. HIV prevalence of bacteraemic children lies between 25 and 35% another risk factor is severe malnutrition (Gordon *et al.*, 2008). A rapid emergence of multi drug resistance has been observed in Malawi over the study period of ten years, which is defined as resistance to chloramphenicol, ampicillin and co-trimoxazole. Between 80-100% of *S.* Typhimurium isolates were resistant to chloramphenicol and ampicillin prior to January 1998, with resistance to co-trimoxazole emerging in 2001. Multi-drug resistance in *S.* Entertidis was >90% in 1998 and started to decrease between 2001 and 2004 (Gordon *et al.*, 2008).

S. Typhimurium D26104, *S.* Heidelberg D23734 and *S.* Enteritidis D21685 were all isolated in 2004, while *S.* Bovismorbificans 3114 was isolated in 1997. D26104 was found to be resistant to amoxycillin, trimethoprim, sulphamethoxazole, cefuroxime, rifampicine and streptomycin, D21685 is resistant to rifampicin, while D23734 and 3114 are both resistant to sulphomethoxazole, cefuroxime and rifampicin, the patient from which D23734 was isolated died (W. Dove, personal communication). For this chapter *S.* Typhimurium D26104, *S.* Heidelberg D23734 and *S.* Bovismorbificans 3114 were subtracted against *S.* Typhimurium LT2. *S.* Enteritidis D21685 was subtracted against *S.* Enteritidis PT4 NCTC13349.

3.1.5. Aims of this chapter

- The aim is to identify novel genetic sequences specific to NTS isolates of common serovars from sub-Saharan Africa using SSH

3.2. Results

3.2.1. SSH reactions

In total, four suppression subtractive hybridisations were carried out to the specifications of the CLONTECH PCR-Select™ Bacterial Genome Subtraction Kit, using four different *Salmonella* serovars as tester strains. Tester and driver strain genomic DNA was isolated using the Wizard® Genomic DNA Purification Kit (Promega). Genomic DNA was subsequently digested with the restriction

endonuclease *Rsa*I, which cuts double stranded DNA at 5'-G T^A C-3' producing shorter fragments with blunt ends, which are required for adaptor ligation. The size range of these fragments is optimal for SSH. Phenol-chloroform extraction and ethanol precipitation was carried out to remove impurities; some loss of DNA will occur at this step (Figure 3.4).

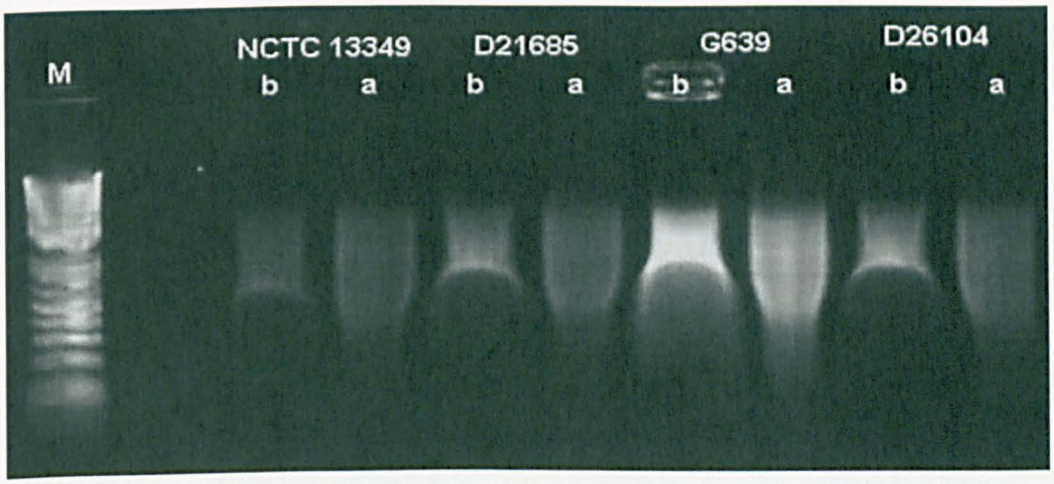


Figure 3.4 Genomic DNA of Driver strains *S.* Enteritidis PT4 (NCTC13349) and *S.* Typhimurium LT2 (G639) and tester strains *S.* Enteritidis D21685 and *S.* Typhimurium D26104, obtained through Wizard genomic DNA extraction (Promega) was digested with restriction enzyme *Rsa*I and further purified through ethanol precipitation and phenol extraction in preparation for SSH reaction. A 2% agarose gel was used to show the concentration of genomic DNA cut with *Rsa*I before (b) and after (a) phenol extraction / ethanol precipitation. M= 1kb-plus DNA ladder (Invitrogen)

Malawian paediatric isolates *S.* Heidelberg D23734, Typhimurium D26104 and Bovismorbificans 3114 were used as testers and subtracted against the driver strain *S.* Typhimurium LT2 (G639). *S.* Enteritidis D21685 was subtracted against the driver *S.* Enteritidis PT4 strain NCTC13349.

SSH was carried out using the PCR Select™ Genomic DNA Subtraction Kit (Clontech). During the primary hybridisation step of SSH, two different adaptors are ligated to two portions of tester DNA before driver DNA is added. In the second hybridisation the two portions are mixed and more driver DNA is added. The unsubtracted tester control (UTC) is produced by mixing adaptor 1 and adaptor 2R-ligated tester before the driver is added. The mixture is subjected to two rounds of PCR amplification, primary and secondary. Primary PCR is tester specific and will only amplify double-stranded DNA with different adaptor sequences on each end. A sample from the primary PCR was set aside to analyze alongside the secondary PCR

products. Secondary PCR is a type of nested PCR, which will further enrich tester specific sequences. Fig 3.2.3. shows primary and secondary PCR for the *S.* Typhimurium D26104 and *S.* Enteritidis D21685 subtractions. It was found in previous experiments that the quantity of DNA after the recommended 25 cycles of secondary PCR is low and therefore an additional 5 cycles of secondary PCR were carried out.

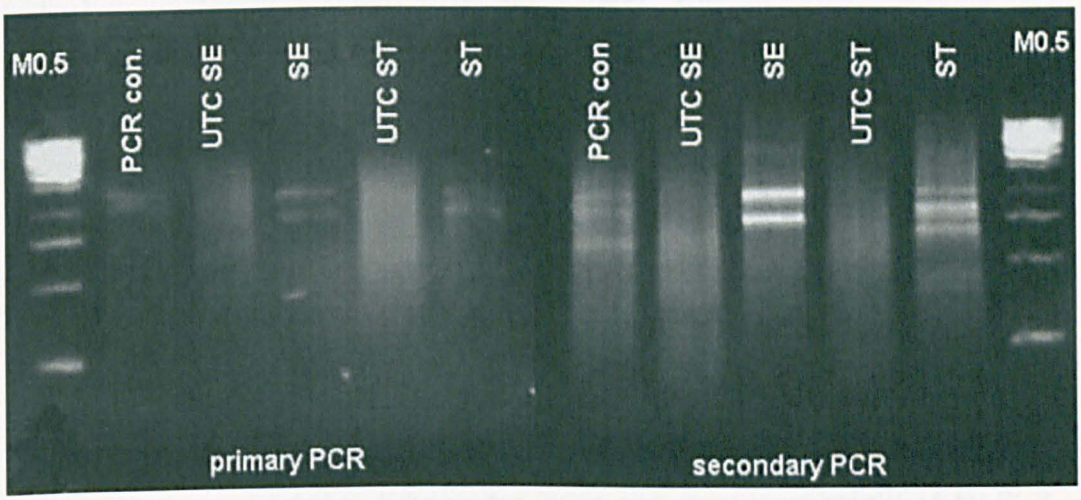


Figure 3.5. Primary and secondary PCR of *S.* Typhimurium D26104 (ST) and *S.* Enteritidis D21685 (SE) subtractions. The PCR control (PCR con.) is a control provided with the kit. UTC (Unsubtracted Tester Control) was produced by mixing adaptor 1 and adaptor 2R ligated tester DNA.PCR products were separated on a 2% agarose gel. Marker (M0.5) = 0.5 µl Invitrogen 1kb-plus DNA marker

Subtracted sequences of *S.* Heidelberg D23734, *S.* Typhimuirum D26104, *S.* Enteritidis D21685 and *S.* Bovismorbificans 3114 were cloned into the pGEM-T vector (Figure 3.5.). The insert will disrupt the *lacZ* α-peptide. The vector also carries an ampicillin resistance gene that enables selection for cells containing the plasmid against cells which did not take up a plasmid. *E.coli* XL10-Gold® Kan^r Ultracompetent cells were transformed with the clones and plated onto L agar containing ampicillin, X-Gal and IPTG.

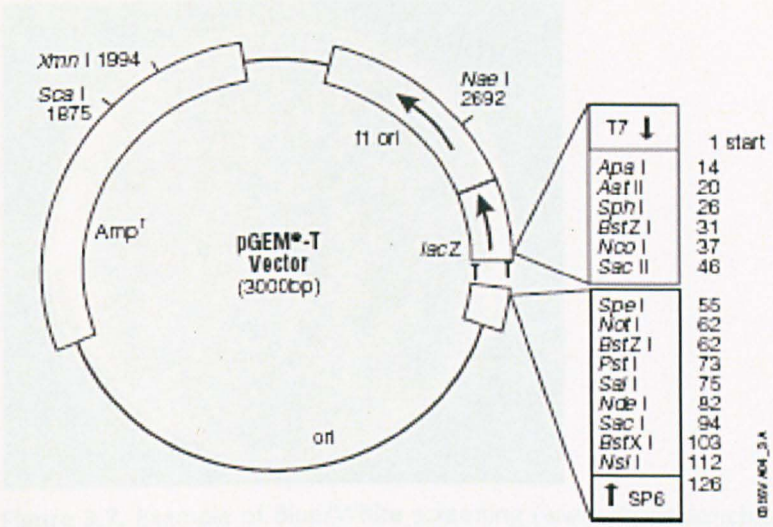


Figure 3.6. pGEM-T® Vector Map and sequence reference points (reproduced from Promega manual)

White colonies were subcultured and selected for plasmid insert sequencing (Figure 3.7.). Between 10 and 20 colonies were selected for PCR amplification, whereby the crude DNA was extracted and the vector insert sequences were amplified by PCR using M13F/R primers. Figure 3.8. shows an example of PCR amplification of insert sequences for the *S.* Enteritidis subtraction, PCR products found suitable for sequencing were subsequently purified using microcentrifuge spin columns and sent for nucleotide sequencing (Lark/Cogenics/Beckmann) to determine whether a significant proportion of the subtracted clones were tester-specific. On the example shown, due to the appearance of two products, product 13 was not used. If the majority of sequences were found to be tester-specific, a further 96 colonies were subcultured onto a microtitre plate containing ampicillin L agar and sent directly for plasmid extraction and sequencing (Lark/Cogenics/Beckmann).

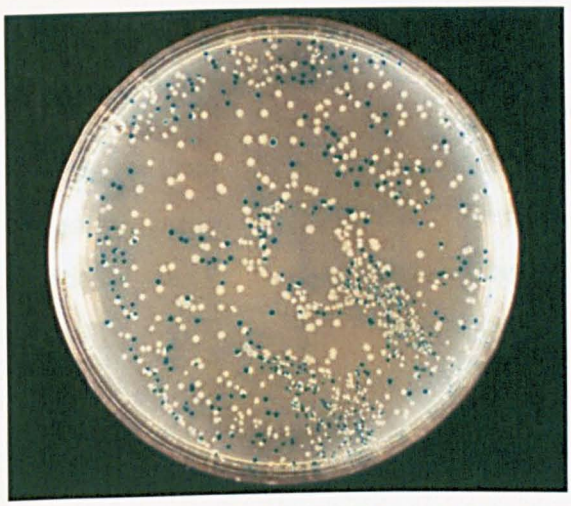


Figure 3.7. Example of Blue/White screening (www.sigmaaldrich.com/catalog) *E.coli* XL10-Gold® Kan^r Ultracompetent cells were transformed with the clones and plated onto L agar containing ampicillin, X-Gal and IPTG. White colonies containing the insertion sequence were selected for M13 PCR.

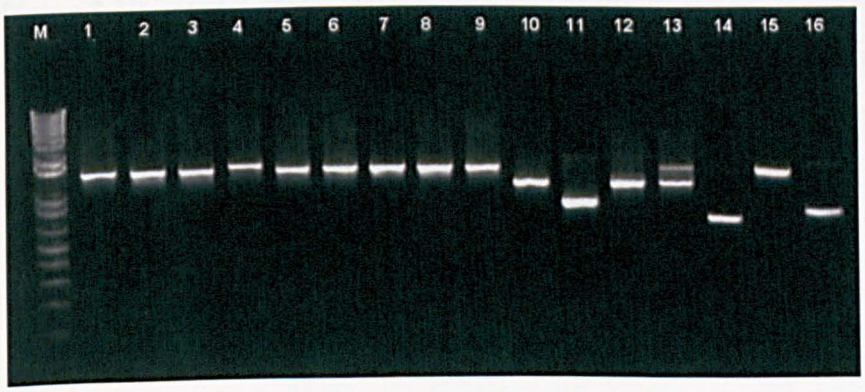


Figure 3.8. *S.* Enteritidis clones, vector inserts were amplified with M13 primers and the PCR products separated on a 1% agaorse gel, M = 1kb-plus DNA ladder (Invitrogen. Each clone was given a number (1-16)

3.2.2. BLASTX

The Basic Local Alignment Search Tool (BLAST) is a program that compares nucleotide and amino acid sequences to sequence databases and calculates the statistical significance of this match. BLASTX searches the protein database using a 6-frame translated nucleotide sequence, which allows for the detection of protein homologies, even with common sequencing errors such as frameshifts or replacements (Altschul *et al.*, 1990) (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

SSH sequences were subjected to BLASTX searches, in Tables 3.3.1 – 3.3.4 Sequences with >90% identity with *S.* Typhimurium LT2 and *S.* Enteritidis NCTC13349, respectively were classified as non-subtracted and are excluded from Table 3.3.1.-3.3.4, which show the best BLASTX matches according to function.

3.2.3. Summary of all SSHs

An overview of the success rate for the *S.* Heidelberg, Typhimurium, Enteritidis and Bovismorbificans subtractions is shown in Table 3.2. *S.* Typhimurium was the most successful subtraction with 77% of sequences being genuinely subtracted, *S.* Enteritidis was the least successful subtraction with a return of only 17% genuine subtracted sequences. With the exception of 3 sequences all of sequences were over 200bp in size and 64 sequences across the Table 3.2. were more than 500bp in size. A small number of sequences, especially from the. *S.* Heidelberg subtraction, matched *S.* Typhi or Paratyphi sequences more closely than NTS. The majority of *S.* Heidelberg sequences matched other *S.* Heidelberg sequences in the database (Figure 3.9). A high number of *S.* Typhimurium and *S.* Bovismorbificans sequences were bacteriophage related, with over 40% of sequences each (Figure 3.9.).

Figure 3.10. gives a more detailed overview of the BLASTX identities of genuinely subtracted sequences (repeats of genuine sequences were excluded from this figure).

Table 3.2. Success rate and outcome of four SSH reactions. A varying number of sequences were of poor quality and were excluded from the analysis

	SSH-D26104	SSH-D1685	SSH-D23734	SSH-3114
Tester serovar	**Typhimurium**	**Enteritidis**	**Heidelberg**	**Bovismorbificans**
Driver	**STM LT2(G639)**	**S Ent (NCTC13349)**	**STM LT2 (G639)**	**STM LT2 (G639)**
SSH success rate				
clones sequenced	**140**	**119**	**114**	**106**
unsubtracted (matching the driver)	**21 (15%)**	**83 (70%)**	**61 (54%)**	**61 (58%)**
genuinely subtracted	**119 (85%)**	**17 (14%)**	**52 (46%)**	**55 (52%)**
repeats of genuine sequences	**92 (77%)**	**10 (59%)**	**11 (21%)**	**3 (5%)**
SSH sequence length				
<200bp	**0**	**1**	**1**	**1**
200-500bp	**10**	**3**	**25**	**22**
>500bp	**17**	**3**	**15**	**29**
Best BLASTX match by serovar/species				
S. Typhi/Paratyphi	**1**	**0**	**4**	**3**
S. Cholerasuis	**1**	**1**	**0**	**7**
other NTS (incl. same serovar as tester)	**6**	**3**	**37**	**17**
E.coli	**6**	**2**	**0**	**2**
Shigella spp	**1**	**1**	**0**	**1**
bacteriophage and other	**12**	**0**	**0**	**22**
Number of SSH sequence chosen for distribution study	**10**	**2**	**14**	**7**

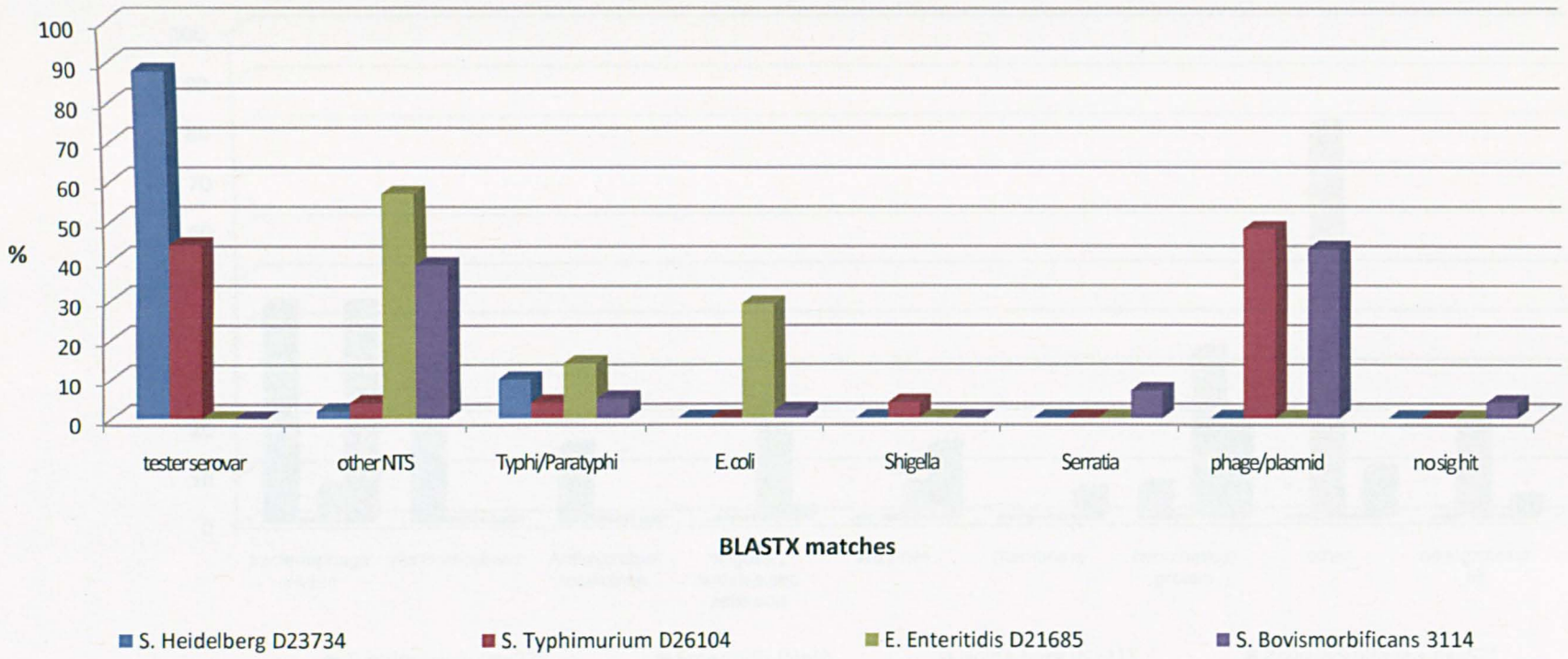


Figure 3.9. origin of BLASTX matches of SSH sequences in percent (%)

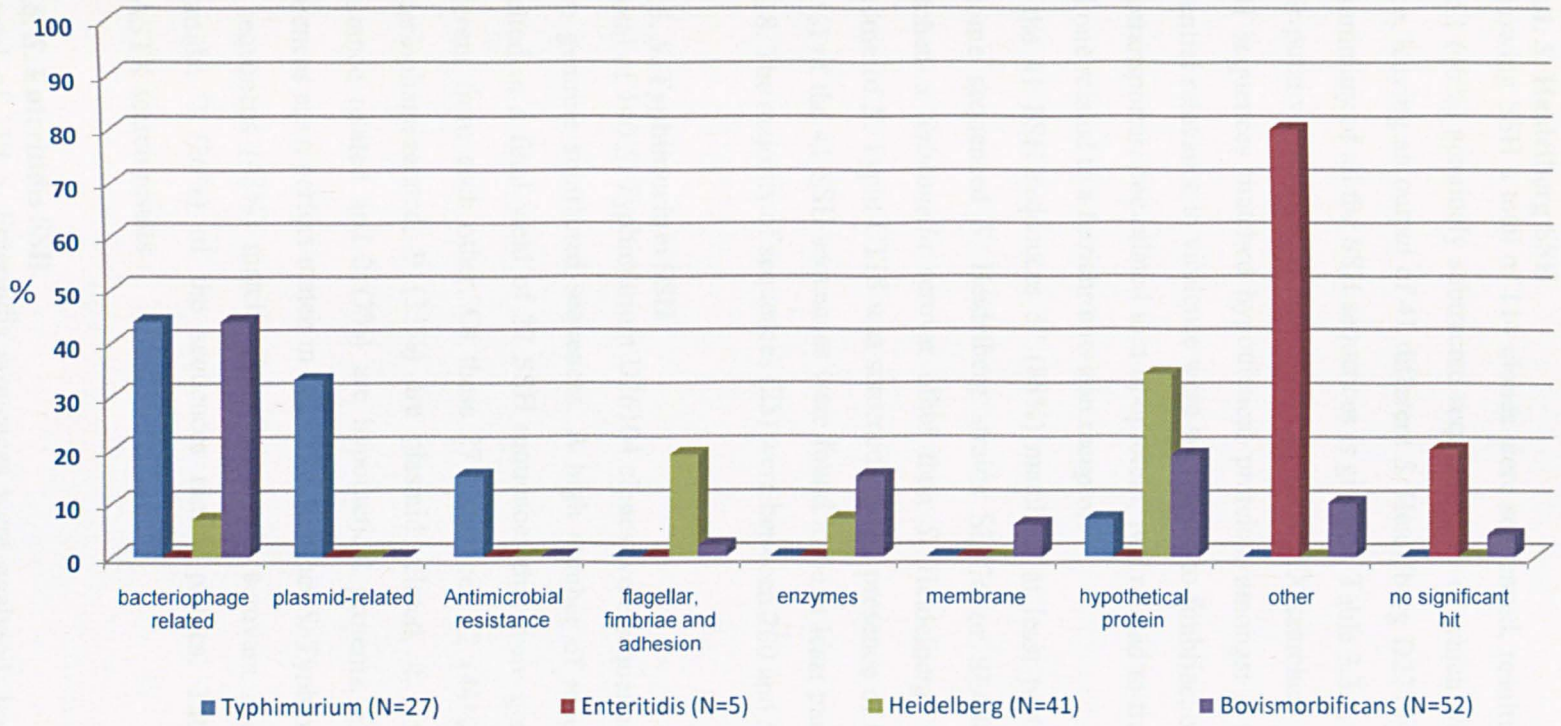


Figure 3.10. SSH sequence BLASTX identities in percent (%)

3.2.4. *S.* Heidelberg SSH

Following SSH a total of 114 clones were sequenced, resulting in the identification of 52 (46%) genuinely subtracted sequences, 11 of which were repeated more than once, leaving an output of 41 different *S.* Heidelberg D23734 subtracted sequences. A summary of all the SSH sequences is given in Table 3.3.1, organized according to their putative function as determined by BLASTX searches. Although many of the SSH sequences matched hypothetical proteins, amongst a number of matches of potential relevance to virulence were six related to fimbriae, one related to a putative autotransporter, one related to a lipoprotein, two related to transcriptional regulators and one related to a ferrichrome-iron receptor.

Of the 41 SSH sequences 37 (88%) matched, at least partly, one or both of the genome sequenced *S.* Heidelberg strains SL476 or SL486, and the other four matched a *Salmonella* serovar other than *S.* Heidelberg. Using BLASTN, the genome of *S.* Typhi CT18 was screened for the presence of the SSH sequences. 19 (46%) of the 41 SSH sequences were found to be at least partly present in *S.* Typhi CT18. The majority of sequences (25) were between 200 and 500 bp in size.

3.2.5. *S.* Typhimurium SSH

A total of 140 *S.* Typhimurium D26104 clones were sequenced of which 119 (85%) were genuine subtracted sequences. A high number of repeated sequences (77%) resulted in a final yield of 27 SSH sequences that were genuinely subtracted and different from each other. Of these 27 sequences 12 (44%) were transposons or bacteriophage-related, 9 (33%) are plasmid related, 4 (15%) are antimicrobial resistance related and 2 (7%) are hypothetical proteins. Eight (30%) of the 27 sequences are a perfect match in BLASTX to other *S.* Typhimurium in the database, 12 sequences (43%) match other *Salmonella* serovars and/or *E.coli* and *E.coli* plasmids, 7 (26%) of the sequences match phages. Table 3.3.2. summarizes BLASTX search results.

3.2.6. *S.* Enteritidis SSH

A total of 119 *S.* Enteritidis sequences were analysed; however the majority of sequences were unsubtracted (70%), matching the driver strain. A high percentage of the 119 sequences showed poor sequencing results. 17 genuine subtraction sequences were identified with 10 repeats, leaving only 7 different SSH sequences, 2 matched ROM-like proteins and 2 matched LasA, while one matched a putative outer

membrane protein, one dehydrogenase subunit and a hypothetical protein of *Shigella flexneri* (Table 3.3.3.).

3.2.7. *S.* Bovismorbificans SSH

The *S.* Bovismobificans subtraction resulted in 106 sequenced clones; 61 (58%) sequences matched the driver, while 55 (52%) were genuinely subtracted sequences with 3 (5%) repeated sequences, therefore SSH resulted in 52 different SSH-sequences. 56% of sequences were larger than 500 bp in size SSH. 21 (40%) of sequences were bacteriophage related, while 7 (13%) putatively encoded for enzymes, 5 (10%) were related to serotyping, and 3 (6%) sequences were membrane related. One sequence putatively encoded a putative fimbrial usher subunit, while another sequence was related to a restriction enzyme. Ten (19%) sequences matched hypothetical proteins, while two (4%) sequences showed no significant match in the database. The remaining sequences not matching bacteriophages matched other NTS, including *S.* Typhimurium, whereas 3 sequences matched *S.* Typhi and Paratyphi. Table 3.3.4 summarizes these SSH results, and gives the Accession numbers of the closest BLASTX matches.

From the genome sequence of 3114 presented in Chapter 5 it was possible to determine regions of difference in 3114 compared to S. Typhimurium LT2 through ACT comparisons (Table 5.4), which amassed to approximately 280.8 kb in difference. SSH of *S.* Bovismorbificans 3114 was not exhaustive, with approximately 5% of repeats (Table 3.2.). SSH identified approximately 27.3 kb or 9.7 % of difference between *S.* Bovismorbificans 3114 and *S.* Typhimurium LT2.

Table 3.3.1 Summary of the BLASTX searches for *S.* Heidelberg D23734 subtracted sequences, length refers to the length of the BLASTX match, sequences underlined were chosen for PCR screening; matches.

SSH Sequence	Length (bp)	Best BLASTX match / comments [accession number]	ID (%)	Length (AA)1	E-value	*S.* Heidelberg match	*S.* Typhi CT18 match
		Fimbriae-related					
D23734-A1	513	putative fimbrial protein, StkD; 100% identity with proteins from serovars Heidelberg, Hadar, Virchow; 97-98% identity with serovars Paratyphi, Kentucky; novel *sta*-like *fim* gene cluster in *S.* Paratyphi	100	134	$7e^{-67}$	ACF66768	STY0204
D23734-A5	521	putative fimbrial protein, TcfA; 100% identity with proteins from serovars Heidelberg, Schwarzengrund, Choleraesuis, Paratyphi, Typhi and Virchow	100	74	$2e^{-36}$	ACF70186	STY0345
D23734-E1	466	putative fimbrial protein TcfD; 98-100% identity with serovars Heidelberg, Paratyphi A and Typhi / weaker matches to other serovars	100	155	$1e^{-38}$	ACF66730	STY0348
D23734-B10	>711	probable lipoprotein; 91-92% identity with proteins from serovars Newport [ABF15025], Saintpaul and Weltevreden; lower (43%) to some other serovars, including Typhi	92	142	$1e^{-72}$	None	STY0332
		periplasmic fimbrial chaperone protein, SafB; 96-98% identity with protein from serovars Newport [ABF15026], Saintpaul, Weltevreden and Typhi	98	51	$1e^{-19}$	None	STY0335
D23734-G3	410	fimbrial subunit; 98% identity with proteins from serovars Heidelberg, Saintpaul; 87-88% identity with several other serovars (only 69% identity with Typhi)	98	78	$2e^{-35}$	ACF69962	STY3090
		hypothetical protein; 100% identity with protein from serovar Heidelberg; 97% identity with serovars Paratyphi A, Newport, Dublin, Cholerasuis, Hadar and Virchow	100	39	$1e^{-13}$	ACF68346	None
D23734-H7	>387	periplasmic fimbrial chaperone PapD; 98-100% identity with proteins from serovars Heidelberg, Choleraesuis, Typhi, Paratyphi A and others.	100	78	$2e^{-30}$	ACF69718	STY3087
		putative minor fimbrial subunit SteD; 95-100% identity with proteins from serovars Heidelberg, Choleraesuis, Typhi, Paratyphi A and others.	100	43	$2e^{-17}$	ACF69250	STY3088
		Other adhesion-related /					
D23734-14	399	putative autotransporter/pertactin; 100% identity with proteins from serovars Heidelberg Paratyphi A, Saintpaul; 97-99% identity with proteins from serovars Paratyphi B and Dublin (*yaiU* in *S.* Paratyphi A)	100	133	$4e^{-70}$	ACF65919	STY0405
		Flagellin-related					
D23734-B11	354	phase 1 flagellin; several serovars (including Heidelberg)	99	118	$6e^{-46}$	ACF69656	None
		Restriction enzyme-related					

D23734-D7	>422	type II restriction enzyme, methylase subunit (serovar Heidelberg)	93	163	$2e^{-85}$	ACF69659	None
D23734-D12	250	type II restriction enzyme, methylase subunit; 96% similarity with serovar Heidelberg	96	83	$6e^{-41}$	ACF69659	None
D23734-B2	374	type II restriction enzyme, methylase subunit, 100% identity with serovar Heidelberg	100	124	$2e^{-68}$	ACF69659	None
		Rhs protein / integrase / phage-related					
D23734-A9	531	phage integrase; 99% identity with serovar Heidelberg	99	125	$7e^{-67}$	ACF66750	STY0946
D23734-C6	>560	Rhs-family protein; 95% identity with Heidelberg and other serovars conserved hypothetical protein; 100% similarity with Heidelberg and other serovars	95 100	109 51	$5e^{-39}$ $8e^{-21}$	ACF69679 ACF68015	None None
D23734-F10	439	Rhs-family protein; 100% identitity with serovar Heidelberg, 99% identity with serovar Kentucky	100	145	$2e^{-82}$	ACF69679	STY0324
		Others					
D23734-4	526	2-nitropropane dioxygenase NPD; 100% identity with serovar Heidelberg, 98% identity with serovars Agona and Kentucky	100	174	$6e^{-94}$	ACF69599	None
D23734-10	554	putative 3-oxoacyl-(acyl-carrier-protein) reductase; 100% identity with serovar Heidelberg; 87% identitiy with *E.coli*	100	147	$7e^{-70}$	ACF67915	None
D23734-C1	394	class II aldolase, tagatose bisphosphate family; 100% identity with serovar Heidelberg, 97% identity with serovar Newport	100	95	$1e^{-46}$	ACF67959	STY3435
D23734-E2	508	ribose-5-phosphate isomerase B, 100% identitiy with serovar Heidelberg, 91% identitiy with *E.coli*	100	144	$2e^{-80}$	ACF70230	None
D23734-E6	>454	ribose-5-phosphate isomerase, N-terminal sequence, 100% identity serovar Heidelberg	100	22	$3e^{-04}$	ACF70230	None
D23734-E12	234	putative permease, 98% identity with serovars Heidelberg, Newport, Virchow, Hadar, Saintpaul, Agona and Paratyphi B	98	78	$2e^{-38}$	ACF66826	STY3861
D23734-D8	>676	zinc-binding domain of primase-helicase family; 89% identity with serovar Heidelberg	89	265	$2e^{-115}$	ACF68985	STY4832
D23734-20	552	DEAD/DEAH box helicase domain protein; 100% identity with serovar Heidelberg	100	183	$5e^{-100}$	ACF69276	None
D2374-D9	355	arsenical resistance operon repressor ; 100% identity with serovar Heidelberg site-specific recombinase, phage integrase family; 100% identity with serovar Heidelberg	100 100	59 30	$2e^{-27}$ $4e^{-10}$	ACF67656 ACF66087	None None
D23734-18	283	PTS system component, 98% identity with serovar Heidelberg / no significant *Salmonella* or *E. coli* matches, 96% identity with *Shigella dysenteriae*	98	93	$8e^{-31}$	ACF68718	None
D23734-D2	402	ferrichrome-iron receptor fhuA; 100% identity with serovars Heidelberg, Agona, Kentucky, Paratyphi B	100	133	$2e^{-49}$	ACF68652	None
D23734-D4	722	transcriptional regulator, TinR, 100% identitiy with serovars Heidelberg and	100	151	$1e^{-79}$	ACF66366	STY0349

		Schwarzengrund, 98-99% identity with Choleraesuis, Typhi and Paratyphi A					
D23734-9	691	HNH nuclease, 100% identity with serovar Heidelberg	100	212	$7e^{-123}$	ACF66326	STY4853
		Hypothetical proteins					
D23734-B8	433	hypothetical protein, 100% similarity with serovar Paratyphi B [ABX65485], weaker matches to other serovars hypothetical protein, 97% identity with serovars Heidelberg, Paratyphi B, Saintpaul, and Virchow	100 97	69 39	$4e^{-33}$ $2e^{-16}$	None ACF68852	STY0040 None
D23734-B9	549	conserved hypothetical protein, 100% identity with serovar Heidelberg	100	111	$9e^{-62}$	ACF66018	None
D23734-1	311	hypothetical protein (*S.* Typhi str. CT18) / matches *S.* Typhi and *S.* Paratyphi only	100	84	$9e^{-62}$	None	STY4253
D23734-2	720	hypothetical protein, 100% identity with serovar Heidelberg	100	117	$1e^{-61}$	ACF70374	STY3343
D23734-13	269	hypothetical protein SPAB_04978, 98% similarity with serovar Paratyphi B [ABX70270]	98	89	$6e^{-33}$	None	None
D23734-B4	>724	hypothetical protein, matches serovar Heidelberg (also partial overlap with LT2)	100	60	$5e^{-26}$	ACF69552	None
D23734-D1	433	conserved hypothetical protein, 100% identity with serovars Heidelberg and Virchow	100	102	$1e^{-54}$	ACF66085	STY0350
D23734-H12	190	conserved hypothetical protein, 100% identity with serovar Heidelberg and some other serovars including Virchow, Newport and Choleraesuis; weaker matches with other *Salmonella* serovars, including LT2	100	63	$9e^{-32}$	EDZ22919	None
D23734-5	434	conserved hypothetical protein, 100% identity with serovar Heidelberg	100	117	$1e^{-61}$	ACF68341	None
D23734-11	296	hypothetical protein, serovar Virchow str. SL491 [EDZ01229]' shorter match with serovar Heidelberg	89	73	$5e^{-30}$	ACF66488	None
D23734-19	260	hypothetical protein, 100% identity with serovar Heidelberg	100	86	$2e^{-41}$	ACF66306	None
D23734-F3	291	hypothetical protein, 100% identity with serovar Heidelberg	100	96	$1e^{-47}$	ACF66306	None
D23734-G1	421	hypothetical protein, 100% identity with serovar Heidelberg	100	78	$1e^{-34}$	ACF68029	None
D23734-G2	480	hypothetical protein SPAB_02188, serovar Paratyphi B str. SPB7 [ABX67571]	95	97	$1e^{-44}$	None	None

Table 3.3.2. SSH BLASTX match for *S.* Typhimurium D26104, length refers to the length of the BLASTX match, sequences underlined were chosen for PCR screening; matches. BLAST matches to Typhimurium refer to strains other than LT2.

Sequence ID	Length (bp)*	Best BLASTX match [accession number] / comments	Accesion number	%ID	Length (AA)	E-value	*S.* TM match	D23580 match
		Transposon / bacteriophage-related						
D26104-5	787	Transposase (Tn*3*) 94% match with *S.* Typhimurium, *E.coli*, *Klebsiella* and others	CAG35429	94	221	$1e^{-126}$	CAG25429	No
D26104-B9	302	Transposase resolvase (*S.* Typhimurium and Infantis)	CAG25428	100	100	$7e^{-48}$	CAG25428	No
D26104-11	808	Tum protein (*Enterobacteria* phage 186); Orf97 (*Enterobacteria* phage 186)	AAC34187 AAC34191	91 93	146 65	$2e^{-70}$ 33^{-16}	-	Yes
D26104-ST1	371	EaA (*Enterobacteria* phage P22)	AF527608	97	123	$7e^{-66}$	-	Yes
D26104-ST2	788	Tnp 99%-100% match *S.* Typhimurium, *E. coli*, *Klebsiella* and others	CAG25429	99	247	$8e^{-141}$	CAG25429	No
D26104-ST3	>753	TnpA 95% match *S.* Typhimurium, *E.coli*, *Klebsiella* and others	CAL44961	95	293	$5e^{-148}$	CAL44961	No
D26104-ST11	408	IS66 family transposase orfB 97% match *E. coli*, lower matches with *Salmonella* and *Shigella*	ACB19446	97	135	$3e^{-63}$	-	No
D26104-B6	582	gp10 (*S.* Typhimurium bacteriophage SE1)	AAY46510	97	193	$4e^{-96}$	-	Yes
D26104-H11	>667	gp24 protein (*Enterobacteria* phage PsP3)	AAN08386	94	225	$9e^{-113}$	-	Yes
D26104-F9	468	gp36 (*Enterobacteria* phage PsP3)	AAN08398	100	161	$2e^{-91}$	-	Yes
D26104-B4	847	Orf97 (*Enterobacteria* phage 186); gp37 (*Enterobacteria* phage PsP3)	AAC34191 AAN08401	96 93	79 63	$4e^{-44}$ $2e^{-16}$	-	Yes
D26104-E2	504	C3, regulatory protein; similar to ST104 C3 (*Salmonella* phage SE1)	AAY46470	97	41	$2e^{-15}$	-	Yes
		Plasmid-related						
D26104-9	>873	Colicin E1 protein cea, plasmid ColE1	AAB59121	100	242	$3e^{-114}$	-	No
D26104-ST8	725	entry exclusion protein 2: (Plasmid ColE1) Exc2	AAB59140	91	107	$2e^{-43}$	-	No
D26104-ST9	392	entry exclusion protein 1 (*Shigella sonnei* Ss046, also matches Exc1 from ColE1)	ABO87515	96	32	$4e^{-10}$	-	No
D26104-A4	564	Chain A, Crystal Structure Of The Channel-Forming Domain Of Colicin E1	AAA59410	100	88	$1e^{-42}$	-	No
D26104-A5	415	Mobilization protein 3 (plasmid ColE1)	AAB59134	95	117	$5e^{-45}$	-	No
D26104-A10	609	DNA transfer protein (*S.* Paratyphi A str. ATCC 9150)	AAV78271	94	155	$3e^{-69}$	ABF14990	Yes
D26104-B10	778	MobA protein (*E. coli* E24377A); MobB protein (*E. coli*)	ABV21205 CAF28884	96 93	256 117	$4e^{-87}$ $9e^{-35}$	-	No
D26104-C2	243	colicin protein (*E. coli*)	AAA59406	100	81	$4e^{-11}$	-	No

D26104-C3	725	entry exclusion protein 2: (Plasmid ColE1); entry exclusion protein 1 (ColE1)	AAB59140 AAB59142	91 93	107 107	$2e^{-43}$ $1e^{-37}$	-	No
		Antimicrobial resistance						
D26104-C6	615	TEM extended-spectrum beta-lactamase (*E. coli, S.* Typhimurium and others)	ABI20744	100	102	$2e^{-51}$	CAG25427	No
D26104-C11	>771	TEM extended-spectrum beta-lactamase (*E. coli, S.* Typhimurium and others)	ABI20745	100	183	$2e^{-90}$	AAM28884	No
D26104-D7	>247	TEM extended-spectrum beta-lactamase (*E.coli, S.* Typhimurium and others)	ABI20745	100	50	$1e^{-20}$	CAG25427	No
D26104-E11	483	aminoglycoside adenylyltransferase (*S.* enterica serovar Choleraesuis, Typhimurium and others)	ACF76696	100	160	$1e^{-86}$	CAD42867	No
		Hypothetical / unknown proteins						
D26104-C1	793	hypothetical protein SARI_02653 [*S.* enterica subsp. Arizonae and *S.* Typhimurium) Orf25 protein (Bacteriophage P22-pbi)	ABX22510 AAM81445	91 100	35 24	$1e^{-11}$ $2e^{-06}$	ABF14973	Yes
D26104-B11	404	hypothetical protein SCH_019 (*S.* Choleraesuis str. SC-B67), hypothetical protein CGSHiII_09998(*Haemophilus influenzae* PittII), probable RNAI modulator (*S.* Typhimurium)	AAS76303 EDK11269	96	30	$5e^{-09}$	AAR02633	No

Table 3.3.3. SSH BLASTX matches *S.* Enteritidis D21685, length refers to the length of the BLASTX match, sequences underlined were chosen for PCR screening; matches.

Sequence ID	Length (bp)	Best BLASTX match/ comments	Accession number	%ID	Length (AA)	E-value
D21685-1	445	D-amino acid dehydrogenase-Alanine racemase fusion protein small subunit (*S.* Gallinarum, LT2 and other NTS); Partial match with NCTC13349 and D23580	CAR37191	91	104	$5e^{-37}$
D21685-3	501	Rom-like protein (*E. coli*); Partial match with LT2 (99% ID over 111 bp) partial D23580 match (95% over 122 bp) crotonobetainyl-CoA:carnitine CoA-transferase (Citrobacter sp. 30_2, *S.* Typhi)	AAF71175 EEH94981	95 100	64 26	$2e^{-16}$ $7e^{-8}$
D21685-7	>706	Rom-like protein (*E. coli*); Partial match with LT2 (99% ID over 111 bp) (possibly same as 3)	AAF71175	95	64	$2e^{-16}$
D21685-8	666	hypothetical protein SFxv_5157 (Shigella flexneri 2002017)	ADA77052	98	57	$2e^{-25}$
D21685-F7	494	LasA (*S.* Virchow); partial	ACH87664	91	147	$5e^{-56}$
D21685-SE8	>720	LasA (*S.*Virchow); partial	ACH87664	92	102	$4e^{-45}$
D21685-SE4	466	putative outer membrane lipoprotein (*S.* Choleraesuis str. SC-B67); matches D23580 98%	AAX64324	100	155	$4e^{-85}$

Table 3.3.4. SSH BLASTX matches *S.* Bovismorbificans 3114, length refers to the length of the BLASTX match, sequences underlined were chosen for PCR screening; matches.

Sequence ID	**Size (bp)**	**Best BLASTX match/comments**	**Accession number**	**%ID**	**Length (AA)[1]**	**E-value**
		Bacteriophage related				
3114-A2	583	Gifsy-1 prophage head-tail preconnector gp4 [Phage Gifsy-1, *S.* Cholerasuis]	AAX65123	84	150	$9e^{-64}$
3114-F4	181	Gifsy-1 prophage VmtT, minor tail-like protein [Phage Gifsy-1, *S.* Cholerasuis, *S.* Typhimurium LT2]	AAX65134	67	59	$5e^{-08}$
3114-F10	433	Gifsy-1 prophage terminase large chaing gp2 [Phage Gifsy-1, *S.* Cholerasuis, *S.* Typhimurium LT2]	AAX65121	79	144	$2e^{-34}$
3114-D10	247	Protein gp55 precursor [Phage Gifsy-1]	AAX65118	93	81	$1e^{-20}$
3114-G11	536	Protein gp55 precursor [Phage Gifsy-1]; the rest of the sequence matches *S.* Typhimurium LT2	AAX65118	92	66	$4e^{-19}$
3114-H1	705	Putative prophage membrane protein [*S.* Cholerasuis, Phage Gifsy-1]; the rest of the sequence matches *S.* Typhimurium LT2	AAX65116	100	58	$3e^{-22}$
3114-G2	584	Putative phage encoded exported protein [*S.* Enteritidis str. P125109]; Hypothetical protein [Bacteriophage P27]	CAR32975 CAC83550	100 82	33 35	$8e^{-10}$ $1e^{-09}$
3114-H2	517	P27p48, hypothetical protein [Enterobacteria phage P27]	CAC83566	97	159	$3e^{-55}$

3114-A1	368	hypothetical protein P27p52 [*Enterobacteria* phage phiP27]	CAC83570	92	122	$6e^{-54}$
3114-B6	579	putative tail fibre protein [Bacteriophage P27]	CAC83574	59	201	$5e^{-49}$
3114-5	>685	phage portal protein, lambda family [*Serratia proteamaculans* 568]	EAV29393	76	271	$2e^{-80}$
3114-15	688	phage tail tape measure protein, TP901 family [*Serratia proteamaculans* 568]	ABV44003	68	214	$2e^{-64}$
3114-A9	585	phage terminase GpA [*Serratia proteamaculans* 568]	ABV44014	80	194	$7e^{-67}$
3114-2	710	Putative phage terminase, large subunit [*Escherichia coli* 101-1]	EDX38951	99	100	$2e^{-46}$
3114-6	675	Conserved hypothetical protein [*Escherichia coli* 83972] ; Hypothetical protein [*Escherichia coli* CFT073]	EEJ45413 AAN79880	98 96	101 97	$4e^{-51}$ $1e^{-39}$
3114-C2	603	Eaa protein, phage-related [*S.* Choleraesuis str. SC-B67]	AAX64232	91	80	$1e^{-26}$
3114-D7	318	Phage tape-measure protein [*S.* Enteritidis str.P125109]	CAR33510	56	96	$6e^{-11}$
3114-E9	537	Mutator family transposase [*Acinetobacter johnsonii* SH046]	EEY95040	72	40	$3e^{-27}$
3114-G4	655	hypothetical protein [Stx converting bacteriophages, *Escherichia coli* O157:H7]	BAB87919	57	135	$8e^{-35}$
3114-H10	566	hypothetical bacteriophage protein [*Shigella* spp.]	AAN43573	93	101	$7e^{-40}$
3114-3	291	no significant match, but matches phage-related proteins below the levels of significance				
3114-F6	348	Phage-holin analog protein [*S.* Heidelberg str. SL486]	EDZ23960	100	40	$1e^{-16}$
3114-D3	422	Predicted phage protein [*S.* Gallinarum str. 287/91]	CAR37076	100	15	4.8
		Serotype-related				
3114-8	379	rhamnosyl transferase, O-antigen-related [*S.* Cholerasuis]	CAA43911	100	125	$1e^{-44}$
3114-D4	309	O-antigen polymerase [*S.* Cholerasuis]	CAA43912	100	89	$1e^{-42}$
3114-E1	516	O-antigen polymerase [*S.* Choleraesuis]	CAA43912	100	106	$7e^{-73}$
3114-B8	371	lipopolysaccharide 1,2-N-acetylglucosaminetransferase [*S.* Typhimurium]	ABF15000	95	111	$6e^{-30}$
3114-A5	399	Outer membrane autotransporter barrel domain protein [*S.* Heidelberg str. SL486]	EDZ27152	99	112	$1e^{-29}$
		Fimbriae-related				
3114-C12	>771	putative outer membrane fimbrial usher protein [*S.* Typhimurium str. 14028S and other Salmonellae]	ACY90066	47	76	$2e^{-11}$
		Restriction-modification-related				
3114-B7	332	Type I restriction-modification system, M subunit [*S.* Hadar str RI_05P066 and other *S.* enterica]	EDZ35297	100	110	$2e^{-18}$
		Membrane proteins				

3114-A12	804	putative membrane protein [*S.* Enteritidis and other *S.* enterica]	CAR35549	99	267	$5e^{-121}$
3114-G6	320	putative membrane protein [*S.* Cholerasuis and other *S.* enterica]; the rest of the sequence matches *S.* Typhimurium LT2	AAX68279	97	43	$1e^{-16}$
3114-F8	791	lipoprotein [*S.* Gallinarum str. 287/91]	CAR36211	98	147	$3e^{-78}$
		Enzymes				
3114-17	>759	exodeoxyribonuclease VIII-like protein [*S.* Typhimurium]	AAS57774	97	249	$6e^{-141}$
3114-18	673	peptidase S14, ClpP [*Serratia proteamaculans* 568]	ABV44012	65	208	$2e^{-66}$
3114-19	493	peptidase S14, ClpP [*Serratia proteamaculans* 568]	ABV44012	54	122	$2e^{-22}$
3114-11	381	Putative membrane protein [*S.* Enteritidis str. P125109]	CAR35549	98	57	$1e^{-16}$
3114-D2	479	ThiF, thiamine biosynthesis protein [*S.* Paratyphi str. SPB7 and other *S.* enterica] ; the rest of the sequence matches *S.* Typhimurium LT2	ABX70433	100	45	$2e^{-08}$
3114-D12	>753	exodeoxyribonuclease VIII-like protein [*S.* Typhimurium]	AAS57774	98	226	$1e^{-107}$
3114-H5	307	Hypothetical protein SPAB_03274 [*S.* Paratyphi B str. SPB7]; the rest of the sequence matches *S.* Typhimurium LT2	ABX68635	100	26	$4e^{-05}$
		Hypothetical proteins				
3114-12	588	Hypothetical protein [*S.* Agona str. SL483]; Partial match with LT2	ACH50155	99	143	$2e^{-59}$
3114-C1	468	Aec79 [*S.*Virchow str. SL491]	EDZ01563	100	138	$2e^{-44}$
3114-C4	357	Major capsid protein E [*Serratia proteamaculans* 568]	ABV44010	70	119	$1e^{-26}$
3114-C10	296	conserved hypothetical protein [*Serratia proteamaculans* 568]	ABV44017	73	84	$6e^{-16}$
3114-E7	588	hypothetical protein SPC_2947 [*S.* Paratyphi C str. RKS4594] ; the rest of the sequence matches *S.* Typhimurium LT2	ACN47040	98	123	$1e^{-47}$
3114-H8	>427	putative periplasmic protein [*S.* Choleraesuis str. SC-B67]	AAX68101	100	121	$4e^{-39}$
3114-A08	>817	Hypothetical protein SNSL254_A2002 [*S.* Newport str. SL254]	ACF64594	97	42	$6e^{-16}$
3114-B9	527	Hypothetical protein Ec53638_4012 [*E.coli* 53638]	EDU66696	76	25	0.016
3114-E11	628	Hypothetical protein SNSL254_A2646 [*S.* Newport str. SL254]	ACF62532	98	195	$6e^{-106}$
3114-F11	628	Hypothetical protein SNSL254_A2646 [*S.* Newport str. SL254]	ACF62532	100	195	$3e^{-109}$
		No significant hits				
3114-G3	>802	No significant hits				
3114-H6	483	No significant hit				

3.3. Discussion

3.3.1. Effectiveness of SSH

SSH of *S.* Heidelberg, Typhimurium, Bovismorbificans and Enteritidis yielded a range of results. The most successful SSH reaction with 119 (85%) genuine SSH sequences was *S.* Typhimurium; the least successful was *S.* Enteritidis, with just 17 (14%) genuine SSH sequences. Both *S.* Heidelberg and *S.* Bovismorbificans subtractions gave good results with a yield of 36% and 52% genuinely subtracted SSH sequences respectively. We previously used SSH to identify virulence related characteristics of Crohn's disease and colon cancer *E. coli* isolates, where we identified 77% genuinely subtracted sequences (Bronowski *et al.*, 2008); another study investigating *Pseudomonas aeruginosa* CF isolates showed a success of 73.5% genuine SSH sequences (Smart *et al.*, 2006). Agron and co-workers isolated 192 clones in total, 100 (52%) of which were genuinely subtracted (Agron *et al.*, 2001)

3.3.2. *S.* Heidelberg D23734

SSH was used to identify 41 sequences that were present in the accessory genome of African bacteraemia isolate D23734, but absent from *S.* Typhimurium LT2. Sequences obtained from *S.* Heidelberg were the most diverse. They were categorized into 8 larger groups. 41 (36%) of the total of subtraction sequences were genuine and different from each other, 54 (47%) out of 114 were non-subtracted, matching the driver strain LT2. Of these genuine sequences the largest group matched hypothetical proteins of unknown function (34%) another large group of sequences were fimbriae-related (15%) followed by 7% of phage/plasmid-related matches. Included in the category "other" are 32% of all sequences with diverse roles and functions including a number of enzymes such as 2-nitropropane dioxygenase. A number of sequences matched adhesion-related (2%) and restriction-enzymes-related (7%) proteins (Table 3.3.1). 61% of D23734 sequences were between 200 and 500bp in length (Table 3.2.).

S. Typhi carries fimbriae designated *tcf*, for Typhi colonising factor (Folkesson *et al.*, 1999). It has been reported previously that the *tcf* operon is present in *S.* Heidelberg and other invasive serovars such as Paratyphi A, Sendai and Cholerasuis (Townsend *et al.*, 2001). We identified two SSH sequences matching genes within this operon, namely *tcfA* (D23734-A5) and *tcfD* (D23734-E10). In addition, we identified an SSH sequence matching *tinR* (D23734-D4), which lies downstream of the *tcf* genes in *S.*

Typhi, and encodes a transcriptional regulator (Folkesson *et al.*, 1999) (Figure 3.11.). The genomes of *S. enterica* possess numerous fimbrial gene clusters implicated in host colonisation and adaptation. Indeed, the genome of *S.* Typhimurium LT2 alone carries eleven fimbrial operons, some of which have been implicated directly in virulence (Humphries *et al.*, 2001;van der Velden *et al.*, 1998). The *stk* gene cluster has been reported as specific for *S.* Paratyphi A (Edwards *et al.*, 2002). However, SSH sequence D23734-A1 matches *stkD* of S. Heidelberg, Hadar and other NTS isolates, which appears to be in conflict with these previous findings (McClelland *et al.*, 2004). Another potential fimbrial operon is represented by the SSH sequence D23734-G3, but can only be identified as a putative fimbrial subunit with matches in a number of NTS serovars including *S.* Heidelberg.

The repertoire of fimbrial operons varies between serovars, with some widely distributed but others restricted to a limited number of serovars (Porwollik & McClelland, 2003;Townsend *et al.*, 2001). Our observations indicate that fimbrial gene clusters make a major contribution to the accessory genome of *S.* Heidelberg.

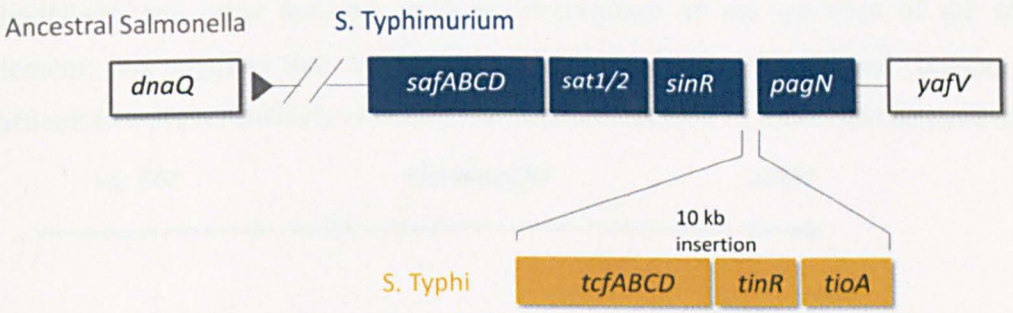


Figure 3.11. The *tinR* regulatory factor of *S.* Typhi is inserted into the *saf* operon

D23734-14 matches a putative autotransporter/pertactin of *S.* Heidelberg and other NTS serovars. Pertactin is a known virulence factor, also designated P69, of *Bordertella pertussis*, the P69 protein is associated with the cell membrane. P69 contains a sequence motif consistent of a arginine-glycine-aspartic acid (RGD) triplet, which functions as a cell binding site in a number of mammalian cells and appears to be involved in host-cell adherence (Charles *et al.*, 1989; Leininger *et al.*, 1992).

Only one phage related sequence was identified in D23734, A9, a putative phage integrase. Further we identified two Rhs family proteins, D23734-C6 and –F10. Rhs elements are large repetitive DNA elements previously described in *E. coli* K12,

where they account for 0.8% of the genome (Feulner *et al.*, 1990; Sadosky *et al.*, 1991). *E. coli* Rhs elements are divided into distinct groups but share common characteristics; each carries a 3.7 kb GC-rich core, which comprises a single ORF spanning the entire core. Adjacent to the core lies an AT-rich region called the core-extension, more than 10 distinctive core extensions have been described(Zhao & Hill, 1995). Each core ORF is overlapped or immediately followed by a downstream ORF (dsORF), which are also AT-rich (Hill *et al.*, 1994). The N-terminus of a number of dsORF appears to carry a signal peptide. Another common feature of Rhs elements is the presence of insertion sequences (ISs), located to the right of the dsORF. Figure 3.12 summarizes the structure of Rhs elements. Rhs elements are not present in all *E. coli* strains and appear to be absent from *S.* Typhimurium LT2 altogether (Hill *et al.*, 1995) (Lin *et al.*, 1984). The function of Rhs elements remain unclear, conditions in which the large ORFs are expressed have not been described, however they appear critically important in *E. coli* population structure (Hill *et al.*, 1994). Rhs family matches D23734-C06 and –F10 are close matches to *S.* Heidelberg and other serovars, with no description of the specifics of the *Rhs* element; this suggests that the annotation has simply been transferred, making it difficult to comprehensively conclude the putative function of these SSH sequences.

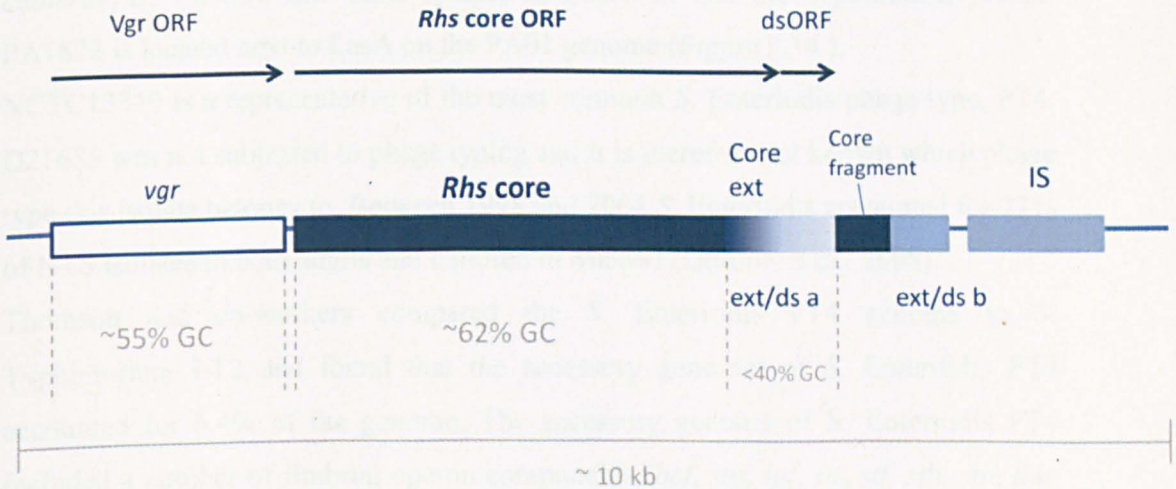


Figure 3.12. Schematic representation of Rhs elements. The most conserved component of Rhs elements is the Rhs core, joined to the core extention (ext) this constitutes the Rhs core ORF. Some cases show an additional extention elsewhere in the element, linked to the 3'core fragment. A segment encoding a dsORF is linked to each core-extension; DNA segments coding both a core-extension and the adjacent dsORF are designated ext/ds. The *vgr* component is associated with *RhsE* and *RhsG* elements, but is absent from the others. *Rhs* elements generally contain one or more insertion sequences, although the ISs may be defective. The G + C contents of the major components are shown in grey. adapted from (Hill, 1999)

3.3.3. *S.* Enteritidis D21685

Subtraction of *S.* Enteritidis D21685 against *S.* Enteritdis PT4 was the least effective. Out of 119 sequences analyzed, 84 (70.5%) matched the driver strain NCTC13349 and with a high number of repeats amongst the genuinely subtracted sequences, only 7 (6%) sequences were both genuinely subtracted and different from each other. The fact that amplicon sequences 1 to 9 in Figure 3.8 were the same size as each other may already have been indicative of the high number of repeats. These results suggest that there is a limited number of additional sequences present in the tester D21685 compared to the driver strain NCTC13349. The predominating match was to a putative LasA (*S.* Virchow, a protease originally from *Vibrio cholerae*), with most of the repeats accounting for this sequence match. There is evidence to believe that this putative LasA sequence has been mis-annotated. A concatenated sequence of "LasA" from D21685-F07 and –SE08 has been used to search the NCBI database, one of the matches was to *P. aeruginosa* PA01 LasA, however when aligning a translated amino acid sequence of the PA01 LasA and the concatenated D21685 "LasA" the match is negligible, as shown in Figure 3.13. The entry has since been changed to hypothetical protein PA1872 of PA01, however the miss-annotation has since self-perpetuated, leading to the annotation of sequence matches in *Vibrio cholerae*, *S.* Virchow and other species as LasA. In fact the hypothetical protein PA1872 is located next to LasA on the PA01 genome (Figure 3.14.).

NCTC13349 is a representative of the most common *S.* Enteritidis phage type, PT4. D21685 was not subjected to phage typing and it is therefore not known which phage type this isolate belongs to. Between 1998 and 2004 *S.* Enteritidis accounted for 21% of NTS isolates in both adults and children in Malawi (Gordon *et al.*, 2008).

Thomson and co-workers compared the *S.* Enteritidis PT4 genome to *S.* Typhimurium LT2 and found that the accessory gene set of *S.* Enteritidis PT4 accounted for 6.4% of the genome. The accessory genome of S. Enteritidis PT4 included a number of fimbrial operon components (*bcf, saf, lpf, sti, stf, stb, sth, fim, sef*) phages and phage remnants (Φ SE10, Φ SE12, Φ SE12A, Φ SE14, Φ SE20) elements of *Salmonella* pathogenicity islands (SPI-1, -2, -3, -6, -9, -10, -13, -14, -16, -17) and other virulence determinants such as RHS elements and C4 carboxylate transporter. (Thomson *et al.*, 2008) . *S.* Enteritidis sequences in this study matched predominantly other NTS serovars (not Enteritidis) and *E. coli*, however due to the limited number of sequences it is difficult to draw any conclusions. Reverse SSH

would have to be carried out in order to detect sequences absent from *S.* Enteritidis D21685 compared to NCTCC13349.

SSH sequence D21685-1 matched D-amino acid dehydrogenase-Alanine racemase fusion protein, originally identified by Thomson et al in *S.* Gallinarum (Thomson *et al.*, 2008). The small subunit of D-amino acid dehydrogenase, a membrane associated enzyme that catalyzes the oxidation of a number of D-amino acid to produce the corresponding imino acid which in turn gets hydrolyzed to alpha-keto acid and ammonia (Olsiewski *et al.*, 1980).

D12685 sequences also matched Rom-like proteins (D12685-3 and -7) and LasA (D21685-F7, -SE3) and an outer membrane protein (D21685-SE4).

The Rom protein controls the copy number of the *E. coli* plasmid ColE1. Chen *et al* identified a Rom-like protein on the plasmid of *S.* Typhimurium isolate by Sanger sequencing (Chen *et al.*, 2007). This protein is of further interest and will therefore be investigated in more detail in chapter 4.

```
  1 AVELLAQDTITETLTQLVGTPIDYLMSKLPKGAEKKIYSLVEKSLNKAAN     50

  0 --------------------------------------------------      0

 51 AALWSLDNEPNR-IASTKTNKFFAALSGAIGG----------AFGFSA--     87
      :.::.:| :||.::...|..|:.|:||          ||.:||
  1 -----MQHKRSRAMASPRSPFLFVLLALAVGGTANAHDDGLPAFRYSAEL     45

 88 ------LAIELPVSTTIMLRSIADIARSEGFDLDRIDTKLECLAVFSYGG    131
       .::.||::..:.|..    ..:|.|||:          |..:...
 46 LGQLQLPSVALPLNDDLFLYG----RDAEAFDLE---------AYLALNA     82

132 PSEDDDAVN----TAYYATRN---LTAEIVQD----------LSKEISNI    164
    |:..|.:..    :.||:...   ||..::|.          |:..:..:
 83 PALRDKSEYLEHWSGYYSINPKVLLTLMVMQSGPLGAPDERALAAPLGRL    132

165 TVKNAASTQTGKWLASLIEKVATR-FGIVITEKMAAQVAPVIG-----AL    208
    :.|.....|    :..:::::::.| :|....:...|.....:|     |.
133 SAKRGFDAQ----VRDVLQQLSRRYYGFEEYQLRQAAARKAVGEDGLNAA    178

209 AGATLNTMFTDYYQDMARGX-----------------------------    228
    :.|.|..:.........:|
179 SAALLGLLREGAKVSAVQGGNPLGAYAQTFQRLFGTPAAELLQPSNRVAR    228

228 --------------------------------------------------    228

229 QLQAKAALAPPSNLMQLPWRQGYSWQPNGAHSNTGSGYPYSSFDASYDWP    278

228 --------------------------------------------------    228

279 RWGSATYSVVAAHAGTVRVLSRCQVRVTHPSGWATNYYHMDQIQVSNGQQ    328

228 --------------------------------------------------    228

329 VSADTKLGVYAGNINTALCEGGSSTGPHLHFSLLYNGAFVSLQGASFGPY    378

228 ----------------------------------------    228

    379 RINVGTSNYDNDCRRYYFYNQSAGTTHCAFRPLYNPGLAL    418
```

Figure 3.13 alignment of D21685 LasA concatenated sequence and LasA of *Pseudomonas aeruginosa* PA01

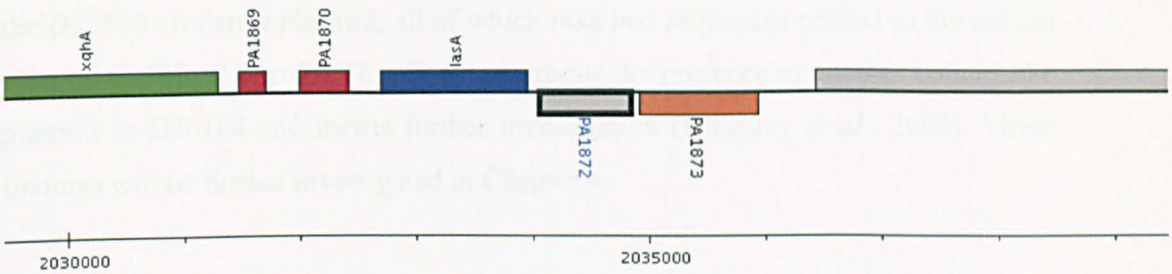


Figure 3.14 Location of PA1872 and *LasA*-encoding gene on PA01 genome

3.3.4. *S.* Typhimurium D26104

S. Typhimurium D26104 sequencing results yielded 108 genuinely subtracted sequences out of 140, but only 32 (23%) of these were also different from each other, with 76 (54%) being sequence repeats. The range of sequences found for *S.* Typhimurium was narrower than that of *S.* Heidelberg, which might be explained by considering the fact that the same serovar was used as a driver. The dominating match found repeatedly were transposases and proteins related to them, accounting for 29% of all the genuinely subtracted sequences found. Colicin (11%) also accounted for a number of repeats, as did extended spectrum beta-lactamases (TEM). 21% of D26104 SSH sequences matched for ORF with no assigned function and other putative proteins, while 11% of D26104 sequences matched hypothetical proteins and 1 sequence (7%) was a match for a DNA transfer protein of *Salmonella* Paratyphi A. In general *S.* Typhimurium sequences appeared to match those of *E. coli*, other *Salmonella* serovars such as Choleraesuis, or a range of bacteriophages.

Given the resistance to amoxicillin and cefuroxime the finding of a number of ESBL-related sequences (D26104-C6, -C11, -D7) is not surprising. D26104-E11 matched a gene encoding aminoglycoside adenylyltransferase (AadA1), which is in keeping with streptomycin resistance. *aadA1* is located on the resistance cassette type A of *S.* Typhimurium DT104 (Ahmed *et al.*, 2005).

In 2009 Kingsley and co-workers genome sequenced the Malawian paediatric isolate D23580, which was also isolated in 2004 and is resistant to amoxicillin, tetracycline, sulphamethoxazole and cefuroxime. Genotyping, using PFGE has been carried out previously on the Malawi collection, and indicated that D26104 and D23580 belong to different genotypes (Prof C.A. Hart, personal communication). A number of bacteriophage-related D26104 SSH sequences (D26104 -ST1, -E02, -C01, -B06, -A10,-H11, -11, -B04 and –F09) mapped to novel prophage regions identified in

D23580. None of the nine plasmid related D26104-SSH sequences was located on the D23580 virulence plasmid, all of which matched sequences related to the colicin plasmid ColE1 of *E. coli*. This finding suggests the presence of another colicin-like plasmid in D26104 and merits further investigation (Kingsley *et al.*, 2009). These findings will be further investigated in Chapter 4.

3.3.5. *S.* Bovismorbificans 3114

The subtraction of *S.* Bovismorbificans against the genome of *S.* Typhimurium LT2 resulted in a total of 52 genuinely subtracted sequences. 29 (56%) of sequences were larger than 500 bp in size. The biggest group of sequences were bacteriophage-related (40%), while 13% putatively encoded for enzymes, 5 (10%) were related to serotyping and 3 (6%) sequences were membrane related. 3114-C12 encoded a putative fimbrial usher subunit, while 3114-B7 was related to a restriction enzyme. Ten (19%) sequences matched hypothetical proteins, with no further attributable function, while two sequences showed no significant match in the database. *S.* Bovismorbificans has not previously been genome sequenced and therefore there is no reference genome available in the database to match SSH sequences to. However the fact that only 2 sequences showed no significant match, may suggest that there is little additional genetic sequence present in *S.* Bovismorbificans compared to other NTS serovars; this theory is further supported by the fact that three sequences were directly related to serotyping. The closest match available to the O-antigen sequences 3114-D4 and 3114-E1 was *S.* Choleraesuis, which is found in group C1, O-antigen 6,7 (Kauffman-White Scheme). 3114 SSH sequences generally matched *S.* Choleraesuis more often than other NTS serovars; among the other close matches were *S.* Enteritidis and *S.* Newport, while 3114-D2, -H5 and E7 matched *S.* Paratyphi, of which H5 and E7 are hypothetical proteins. 22 SSH sequences matched bacteriophages, six sequences matched Gifsy-1, while three others matched phage P27. Gifsy-1 is a fully functional lambdoid prophage of *Salmonella*, which has been described in *S.* Typhimurium. *Salmonella* generally carry five groups of phages, namely P27-like, P2-like, lambdoid, P22-like and T7-like, and three outliers: ε15, KS7 and Felix O1. *Salmonella* only contains one representative of P-27-like (ST64B) and one of T7-like (SP6) phage. *Salmonella* subspecies I genomes, which includes *S.* Typhimurium, contain three complete lambdoid-related prophages of the

Siphoviridae family: Fels-1, Gifsy-1 and Gifsy-2 (Figueroa-Bossi *et al.*, 1997; Figueroa-Bossi *et al.*, 2001).

3.3.5. Previous studies using SSH

Agron *et al* (2001) used SSH for producing a diagnostic test for detecting *S.* Enteritidis in poultry. Strict control of poultry pathogens *Salmonella* Gallinarum and Pullorum has created a niche for *S.* Enteritidis which is fast becoming the most abundant poultry isolate (Baumler *et al.*, 2000). They used the closely related serovar Dublin as a driver to isolate Enteritidis specific-sequences, which could be used as diagnostic markers. They used four different types of restriction enzymes *Rsa*I, *Alu*I, *Sau*3AI and *Hae*III, in separate SSH experiments. They sequenced 48 clones from each subtraction, 192 clones in total. By comparing these sequences with non-Enteritidis sequences in databases they were able to eliminate 92, which showed too much similarity to be of use as a diagnostic marker. They went on to design primers for the other 98 sequences, and using the driver and tester as templates eliminated any unsubtracted sequences. Nine primer pairs (9%) out of 98 amplified sequences only from *S.* Enteritidis and not from *S.* Dublin, and are therefore highly specific. It is important to note that any number of the 92 sequences discarded at the beginning could have been tester specific, but as they match other closely related pathogens they are not useful as a diagnostic tool. Argon *et al* (2001) went on to test these nine primer pairs on a large collection of serovars commonly found in poultry to eliminate any non-specific primer pairs. Interestingly they found that one of the primer pairs detected a different *S.* Dublin strain within this collection, thereby making a point towards recognizing that there exists some diversity between strains of the same serovar.

Kang *et al* (2006) combined microarray analysis and SSH to identify specific sequences in epidemic, multidrug-resistant (MDR) *Salmonella. Salmonella enterica* serovars Typhimurium (DT104 and epidemic pansusceptible DT160) and cephalosporin-resistant MDR Newport strains were chosen for SSH. They used *S.* Typhimurium LT2 as their driver. Three separate subtractions were carried out yielding a total of 885 sequences; it is not clear how many sequences are unsubtracted. A total of 289 sequences were used in constructing the microarray. SSH results for *S.* Newport matched 9 (14%) phage-related sequences, 18 (28%) plasmid-related sequences, and 38 (58%) other sequences from chromosomes or

uncertain sources. All of the 44 SSH sequences specific to *S.* Typhimurium DT160 were found to be phage related. Of the 54 *S.* Typhimurium DT104-specific sequences 22 matched phage-related sequences (40%), 16 (30%) plasmid related and 16 (30%) hypothetical proteins (Kang *et al*, 2006).

3.3.6. Conclusions

The subtraction reaction of *S.* Heidelberg resulted in the most diverse set of sequences, which is not surprising given the fact that it was subtracted against a different serovar; however we identified very few bacteriophage-related sequences. *S.* Bovismorbificans on the other hand, which was also subtracted against *S.* Typhimurium LT2, showed a large collection of phage-related sequences, suggesting that the major differences between the serovars Typhimurium and Bovismorbificans are due to prophages. The subtraction of *S.* Typhimurium D26104 resulted in a number of antimicrobial-resistance sequences, which confirms the MDR status of this isolate. Although *S.* Heidelberg and *S.* Enteritidis were also resistant to multiple antibiotics, we were unable to identify resistance related sequences using SSH. The least successful subtraction reaction in terms of sequence yield was *S.* Enteritidis D21685 against *S.* Enteritidis PT4 NCTC13349, which resulted in a high number of unsubtracted sequences and repeats suggesting that there is little diversity between these strains and possibly within this serovar. The nature of the reaction process of blue/white screening described in section 3.2.1 allows for the selection of single colonies harbouring SSH sequence clones. The rates of repeats for *S.* Heidelberg and *S.* Bovismorbificans were very low (Table 3.2.) and it is therefore reasonable to assume that by randomly selecting and sequencing 100 to 110 clones, we were only able to obtain a reasonable snapshot of sequence diversity. However, SSH will never be 100% efficient at identifying sequences that vary between strains.

SSH is a good way of discerning differences between two closely related strains or serovars, it currently offers a cheaper alternative to genome sequencing methods and is easy to carry out with limited equipment available (Winstanley, 2001), however prices of 2nd and 3rd generation sequencing services are becoming affordable, but the wealth of data arising from sequencing can be challenging.

NCBI BLASTX searches have been used to ascribe SSH sequences a possible function, however the database is not without flaws and some of the sequence

identities appear doubtful. Without further investigations the true functions can neither be confirmed nor disproven.

3.3.7. Further work

- It is important to establish the distribution of SSH sequences involved in virulence among a panel of NTS strains from both sub-Saharan Africa and the UK
- To establish whether D26104 carrys a plasmid similar to ColE1 of *E. coli*

3.4. Summary

- The genomes of *S.* Typhimurium D26104 and *S.* Bovismorbificans 3114 carry a number of phage-related sequences which may indicate the presence of novel prophages
- Fimbrial operons are a major difference between the accessory genomes of *S.* Typhi and NTS such as *S.* Heidelberg, however SSH identified sequences in *S.* Heidelberg D23734 related to the *S.* Typhi fimbrial operon *tcf.*
- S. Enteritidis D21685 showed very little additional genomic sequence compared to *S.* Enteritidis PT4 NCTC13349

CHAPTER 4 INVESTIGATING THE DISTRIBUTION AND PUTATIVE ROLE OF SSH SEQUENCES AMONG A PANEL OF AFRICAN AND UK NTS ISOLATES

4.1. Introduction

Febrile illness is a leading cause for admission to hospitals in Africa; attention to these symptoms focuses on malaria and microbiological identification of other causes is limited.

A recent review of community-acquired bacteraemias in Africa analysed data from 22 studies from 1984 to 2006 and found that from a total cohort of 5578 confirmed bacteraemia patients (2051 children and 3527 adults) from 34 locations, a total of 1643 were due to *S. enterica*. NTS accounted for 960 cases (17%) of bacteraemia in total and was the cause of febrile illness in 14% of adults and 18.7% of children, *Salmonella* serovars Typhimurium and Enteritidis ranked most common and second commonest isolate among *S. enterica* respectively. *S.* Typhi infections are rare in tropical Africa compared to Asia; a total of 650 cases were identified from the literature, 90.5% of which came from two North African countries (Afifi *et al.*, 2005; Hyams *et al.*, 1986 reviewed by Reddy *et al.*, 2010).

NTS are also a common cause of other extraintestinal infections such as meningitis and septic arthritis in children (Lavy *et al.*, 2005). Gordon and co-workers showed that in Malawi, from where the tester strains used in this study originated, 76% of NTS bacteraemias in children and 75% in adults was caused by *S.* Typhimurium; the second commonest cause was *S.* Enteritidis with 21% in both adults and children; *S.* Bovismorbificans ranked third with 2-3%; *S.* Heidelberg was not listed among the identified NTS serovars. The median age of infection in children was 22 months. Over a two year period a case fatality rate of 23% in children was observed (Gordon *et al.*, 2008). In some sub-Saharan African communities the burden of mortality due to paediatric NTS bacteraemia may be greater than that of malaria (Berkley *et al.*, 2005). Case-fatality rates for bacteraemia in children in Africa have been estimated to lie between 4.4 and 27% (Brent *et al.*, 2006; Enwere *et al.*, 2006; Graham *et al.*, 2000; Walsh *et al.*, 2000). The fatality rate for NTS meningitis is likely to be higher than for any other bacterial pathogen; in one study in Malawi 64% of neonates with NTS meningitis died, compared to 26% of group B Streptococcal meningitis (Milledge *et al.*, 2005).

The factors determining this increased pathogenicity are not well understood. Some NTS serovars are categorized as more invasive than others. A recent study from Malaysia found serogroup D isolates, which include serovars Enteritdis, Gallinarum, Pullorum and Javiana, to be more common among blood isolates than other serogroups (Dhanoa & Fatt, 2009). Using genetic profiling by PFGE and antimicrobial susceptibility testing, Kariuki and colleagues showed in 2002 in a study from Kenya that invasive clinical NTS isolates appear to differ from samples taken from livestock and the environment of patients (Kariuki *et al.*, 2002). These factors suggest that the African NTS strains differ from the common food poisoning strains found in Europe in both their route of transmission and possibly their virulence factors (Kariuki *et al.*, 2002). A whole-genome approach by Kingsley and co-workers (2009) has since shown that invasive blood culture isolates of *S.* Typhimurium from adults from Malawi (M. Gordon collection) and Kenya (*S.* Kariuki collection) belong to a distinct new sequence type, ST313, that has rarely been reported outside sub-Saharan Africa, and which harbours a distinct set of novel prophage regions (Kingsley *et al.*, 2009). Furthermore these strains show signs of genome degradation setting them apart from common gastroenteritis-causing strains (Kingsley *et al.*, 2009). As prophage DNA within the genome of a host can pose a metabolic burden, there must be some benefit to the host cell in keeping this DNA. This can range from genes that increase the fitness of the cell to those increasing virulence by influencing bacterial adhesion, colonisation and invasion. Bacteria with six or more prophages integrated in its genome are primarily pathogens (Humbly and Suttle, 2005, Casjens, 1998). *S.* Typhimurium strains normally carry four to five prophages; *S.* Typhi strains Ty2 and CT18 both carry seven prophages (Thomson *et al.*, 2004).

There are various approaches for the identification of prophages within bacterial genomes. The assumption is made that genetic sequences transferred horizontally possess nucleotide signatures (such as percentage GC content) that differ from those of the host genome; other methods rely on predicting genes associated with phage. The previous chapter (Chapter 3) describes a different approach to identifying novel sequences within an isolate, SSH, which is certainly not phage specific but has the advantage of not necessitating actual sequencing of the host genome. Using SSH, sequence libraries were obtained for Malawian paediatric bacteraemia isolates *S.*

Typhimurium D26104, *S.* Enteritidis D21685, *S.* Heidelberg D23734 and *S.* Bovismorbificans 3114.

4.2. Aims of this chapter

- to establish the distribution and thereby representativeness and importance of SSH sequences among a panel of African bacteraemia NTS and UK diarrhoeal isolates
- to further investigate the functionality of some of the strain-specific SSH sequences through expression assays

4.3. Results

4.3.1. Multi-Locus Sequence Typing (MLST) of African *S.* Typhimurium strains

Multi-Locus Sequence Typing (MLST) was carried out on a selection of *S.* Typhimurium isolates from the PCR screening panel to confirm the sequence type. It was not possible to determine the ST of the tester strain D26104. Figure 4.1 shows an example of PCR amplification of loci used in the MLST scheme.

The collections of strains have previously only been characterised by PFGE (Winifred Dove, Caroline Broughton). A selection of strains from Uganda, Malawi, DRC (Zaire) and the UK were therefore further characterised using MLST. A full set of loci were successfully sequenced for seventeen strains. Eleven strains matched sequence type ST313 exactly, four were ST313 clonal complex (cc). One strain was most closely related to both ST313 and ST394, another strain was most closely related to ST313 and ST34 (Table 4.1.). Using the MLST database (http://mlst.ucc.ie/mlst/dbs/Senterica) of published STs an eBurst diagram (http://eburst.mlst.net/) was produced to show the whole *S. enterica* species (Figure 4.2.). Figure 4.3. highlights STs of *Salmonella* serovar Typhimurium, ST19 is the major lineage of *S.* Typhimurium, ST313 circled in red in Figure 4.3. is a common ST of African bacteraemia isolates.

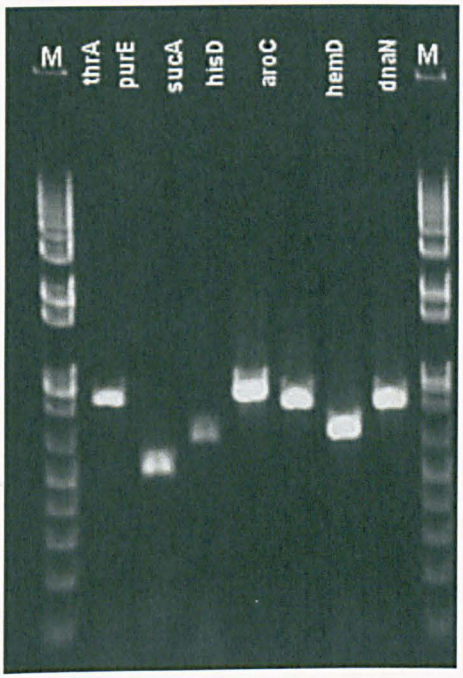


Figure 4.1. PCR amplification of MLST loci from *S.* Typhimurium D26104; 1% agarose gel, M = 1kb plus size marker

Table 4.1. MLST results for 17 *S.* Typhimurium strains

Strain	source	*aroC*	*dnaN*	*hemD*	*hisD*	*purE*	*sucA*	*thrA*	ST match
146	U	10	7	12	133	112	9	53	ST394/313*
868	U	10	7	89	9	112	9	2	ST313cc
4234	U	10	7	12	9	112	9	2	ST313
1060	U	10	7	12	9	112	9	2	ST313
D22209	M	10	7	12	9	112	9	2	ST313
D22337	M	10	7	12	9	112	9	2	ST313
D22404	M	181	7	12	9	112	9	2	ST313cc
D22988	M	10	7	12	9	112	9	2	ST313
D23424	M	10	7	12	9	112	9	2	ST313
D23674	M	123	7	12	9	112	174	2	ST313cc
D22219	M	10	7	12	9	112	9	2	ST313
D23002	M	10	7	12	9	112	9	2	ST313
D24627	M	10	7	12	9	112	9	2	ST313
D25352	M	10	7	12	9	112	149	2	ST313cc
D25834	M	10	7	12	9	112	9	2	ST313
D25840	M	10	7	12	9	112	9	2	ST313
D25907	M	10	19	12	9	112	180	2	ST34/ST313*

U=Uganda, M=Malawi

*best available ST match from the MLST database

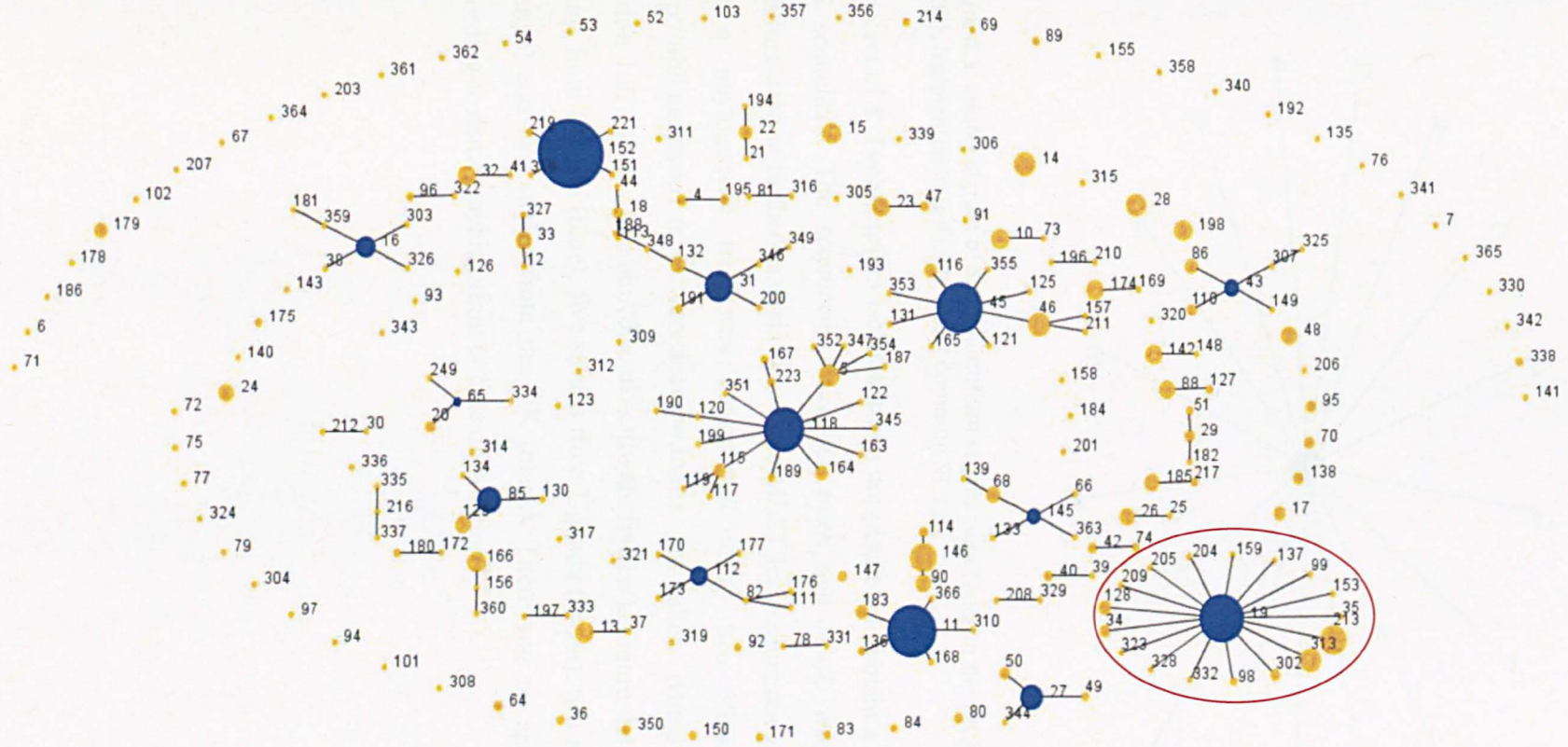


Figure 4.2. eBurst of *S. enterica* strains published on the MLST database (2009), highlighted in red are *S.* Typhimurium STs.

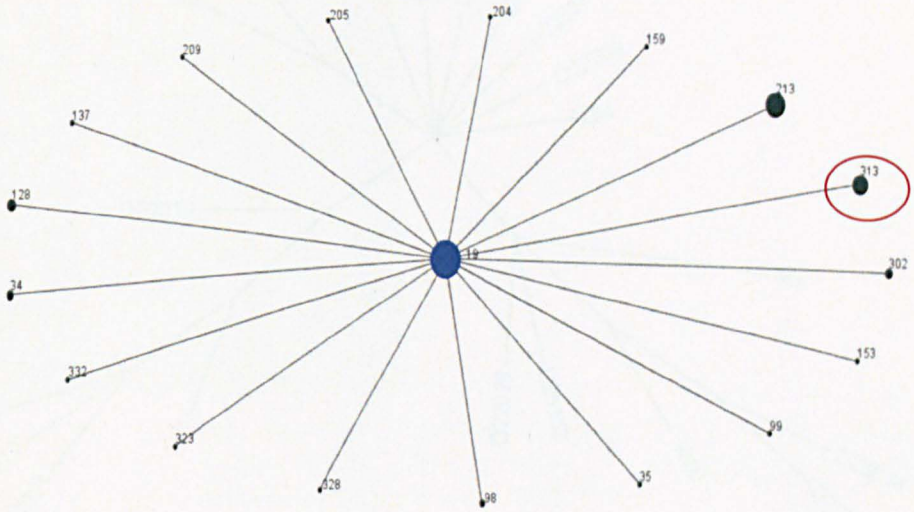


Figure 4.3. eBurst digram of *S*. Typhimurium strains published on the MLST database (2009), highlighted in red is the most common ST 313.

For a set of *S*. Typhimurium isolates it was not possible to obtain a full set of MLST loci sequences; The sequences for *aroC*, *purE* and *hemD* were complete and therefore used in further analysis. The three MLST loci sequences were concatenated and a phylogenetic tree was constructed using Sea View version 4.2.8 (http://pbil.univlyon1.fr/software/seaview.html) and then edited using FigTree version 1.3.1. (http://tree.bio.ed.ac.uk/software/figtree/). Figure 4.4. includes eleven strains from Malawi (blue), five strains from Uganda (green), two strains from DRC (orange) and two strains from the UK (black). There was no apparent clustering according to demographic origin or date of isolation.

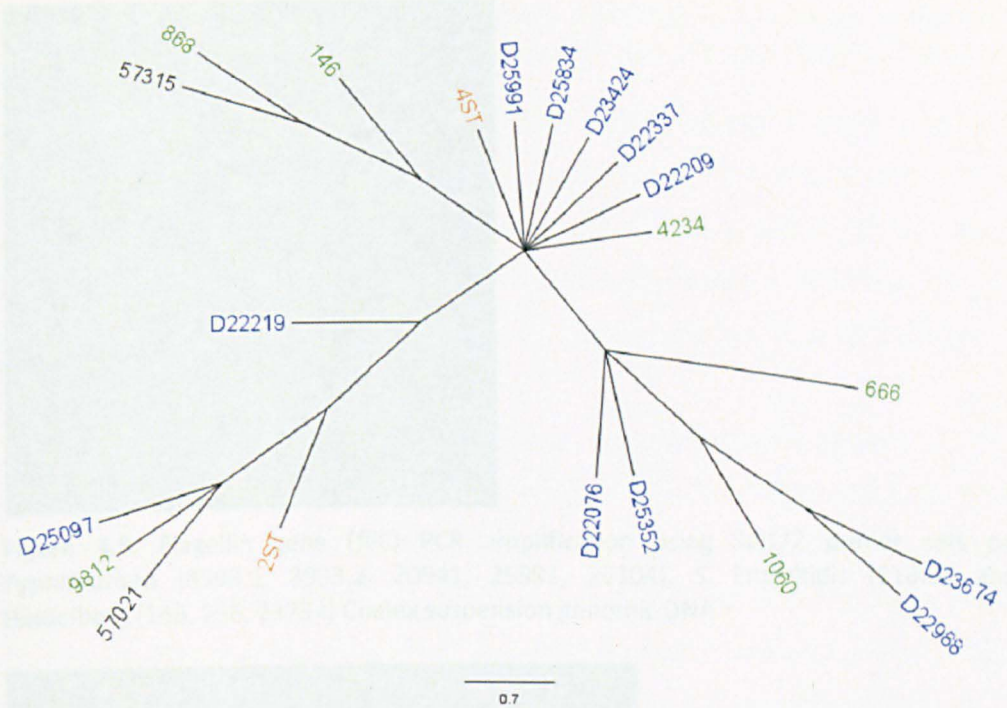


Figure 4.4 parsimony tree constructed of concatenated MLST loci *aroC*, *purE* and *hemD* sequences of 21 strains from the UK (black), Malawi (blue), DRC (orange) and Uganda (green)

4.3.2. Distribution analysis of *S.* Typhimurium D26104, *S.* Enteritidis D21685, *S.* Heidelberg D23734 and *S.* Bovismorbificans 3114 SSH sequences among a panel of African and UK NTS isolates and genome sequenced serovars

A panel of African bacteraemia and diarrhoea and UK diarrhoea isolates were used in PCR assays to determine the distribution of a set of 10 *S.* Typhimurium, 2 *S.* Enteritidis, 14 *S.* Heidelberg and 7 *S.* Bovismorbificans SSH sequences. Genomic DNA was isolated from strains using chelex suspensions and stored at -20°C. The quality of each suspension was assessed by PCR amplification of the *Salmonella* flagellin (*fliC*) gene using the primer set Sal1/2, an example of which is shown in Figure 4.5, demonstrating size variation of *fliC* between *S.* Heidelberg 146 and 256 strains. PCR primers for each of the 33 SSH sequences were designed using the Oligomail (Genosys) program. Gradient temperature PCR reactions were carried out on each primer set with an annealing temperature gradient ranging from 45.0 to 65.6 °C. Figure 4.6 shows an example of gradient temperature PCR reactions for *S.* Typhimurium D26104 SSH sequences. For each positive PCR result, the PCR was repeated for final confirmation.

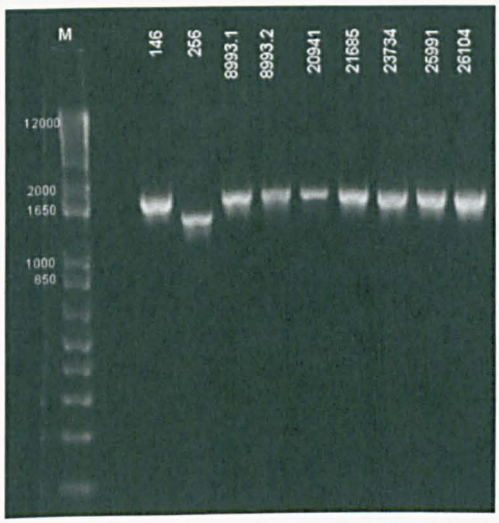


Figure 4.5. Flagellin gene (*fliC*) PCR amplification using Sal1/2 primer sets on *S.* Typhimurium (8993.1, 8993.2, 20941, 25991, 26104), *S.* Enteritidis (21685) and *S.* Heidelberg (146, 256, 23734) Chelex suspension genomic DNA

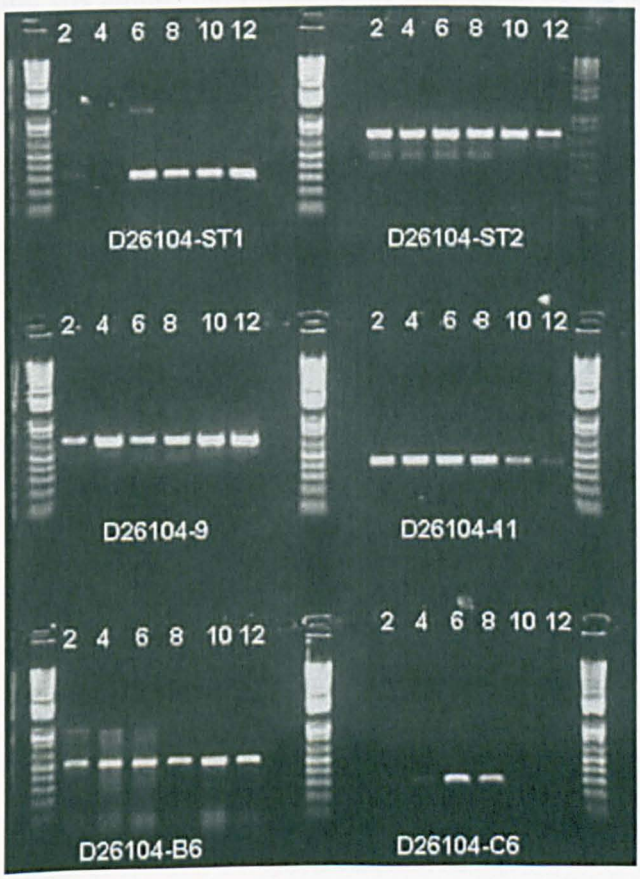


Figure 4.6. An example of temperature gradient PCR for *S.* Typhimurium D26104 SSH sequences, temperature range 2=45.0°C, 4=48.2°C, 6=53.0°C, 8=58.5°C, 10=63.1°C, 12=65.6°C (1% agarose gel, marker 1kb plus)

4.3.2.1. Distribution of 10 S. Typhimurium D26104 SSH sequences among a panel of S. Typhimurium African bacteraemia and UK diarrhoea NTS isolates using PCR assays

A total of ten *S.* Typhiumurium D26104 SSH sequences were selected for distribution analysis by PCR assays; these include five bacteriophage related sequences (D26104-11, -ST1, -B6, -H11, -ST2), one plasmid related sequence (D26104-ST2) and four putative antimicrobial resistance inferring sequences (D26104-C06, -C11, -D7, -E11). Figure 4.7 shows an example of distribution PCR assays for D26104-C11.

Interestingly, it was possible to determine some geographic differences in the distribution. A number of phage sequences, gp24 – D26104-H11 and Tnp-D26104-ST2, and also colicin E1 (D26104-9), were absent from both Malawi and the UK isolates. Further the putative gene for Tum protein of phage 186 was absent from all the UK strains but present in at least some of the strains from each African country. The EaA protein of phage P22 was absent from the UK and Kenyan strains. It is notable that six of the sequences were present in 100% of Ugandan and Malawian strains but in a lower percentage of the Zaire, and Kenyan isolates (Table 4.2.).

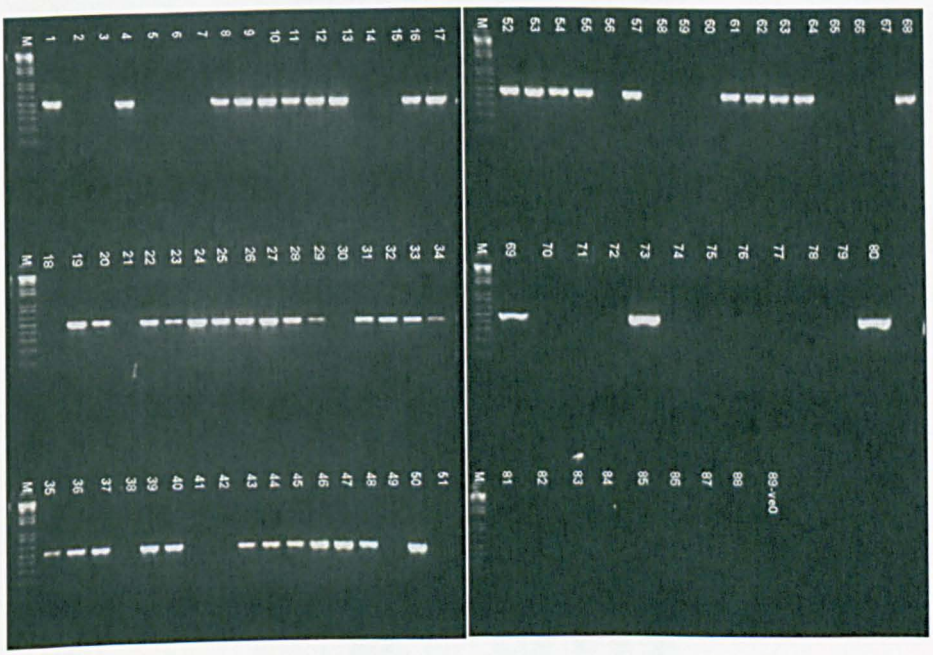


Figure 4.7. Distribution of D26104-C11 among a panel of NTS isolates, 1 = positive control D26104, 2 = negative control G639, 1% agarose gel, M = 1 kb plus size marker

Table 4.2. Distribution of *S.* Typhimurium D26104 SSH sequences according to PCR assays

SSH sequence and best putative function according to database match	*S.* Typhimurium (Uganda, N=9)	*S.*Typhimurium (Malawi, N=17)	*S.* Typhimurium (DRC, N=7)	*S.* Typhimurium (Kenya, N=4)	*S.* Typhimurium (UK, N=6)
D26104-11 (Tum protein, phage 186)	100%	100%	5 (71%)	3 (75%)	0
D26104-ST1 (EaA, phage P22)	100%	100%	6 (86%)	0	0
D26104-B6 (gp10, phage SE1)	100%	100%	5 (71%)	4 (100%)	3 (50%)
D26104-H11 (gp24, phage PsP3)	2 (22%)	0	4 (57%)	3 (75%)	0
D26104-ST2 (Tnp)	4 (44%)	0	4 (57%)	2 (50%)	0
D26104- 9 (Colicin E1 protein)	4 (44%)	0	2 (28%)	2 (50%)	0
D26104-C6 (TEM extended spectrum beta-lactamase)	100%	100%	5 (71%)	2 (50%)	2 (33%)
D26104-C11 (TEM extended spectrum beta-lactamase)	100%	100%	4 (57%)	2 (50%)	2 (33%)
D26104-D7 (TEM extended spectrum beta-lactamase)	100%	100%	4 (57%)	1 (25%)	2 (33%)
D26104- E11 (Aminoglycoside adenylyltransferase)	7 (77%)	15 (88%)	4 (57%)	3 (75%)	2 (33%)

4.3.2.2. Distribution of 10 *S.* Typhimurium D26104 SSH sequences among genome sequenced *S. enterica* serovars using BLASTX

All 74 genomes of *Salmonella enterica* currently available on NCBI BLAST (September 2009) were screened for the presence of 10 *S.* Typhimurium SSH sequences. 30 genomes were negative for all 10 SSH sequences, including 6 genome sequenced *S.* Typhimurium strains (this includes LT2). *S.* Typhi CT18, *S.* Cholerasuis and *S.* Schwarzengrund were found to be positive for all three TEM sequences. *S.* Paratyphi A and *S.* Saintpaul SARA23 were the only strains positive for the SSH sequence matching EaA of P22. *S.* Schwarzengrund str. CVM19633 was positive for the highest number of D26104 SSH sequences, with matches to Tnp (D26104-ST1), all three TEM (D26104-C6,-C11 and -D7) and aminoglycoside adenylyltransferase (D26104-E11) (Table 4.3.). A number of <95% matches to EaA, gp10 and gp24 were found in other serovars not listed in Table 4.3 (Supplementary data, S2).

Table 4.3. Distribution of SSH sequences among genome-sequenced *Salmonella* strains using BLASTX screening

SSH sequence and best putative function according to database match	S. Typhimurium (N=6)	S. Typhi (N=12)	S. Cholerasuis (N=1)	S. Heidelberg (N=2)	S. Newport (N=2)	S. Paratyphi A (N=2)	S. Saintpaul (N=2)	S. Schwarzengrund (N=3)	S. Virchow (N=1)	other serovars (N serovars=69)
D26104-11 (Tum protein, phage 186)	0	0	0	0	0	0	0	0	0	0
D26104-ST1 (EaA, phage P22)	0	0	0	0	0	2	1	0	0	0
D26104-B6 (gp10, phage SE1)	0	0	0	0	0	0	0	0	0	0
D26104-H11 (gp24, phage PsP3)	0	0	0	0	0	0	0	0	0	0
D26104-ST2 (Tnp)	0	0	0	0	0	0	0	1	0	0
D26104- 9 (Colicin E1 protein)	0	0	0	0	0	0	0	0	0	0
D26104-C6 (TEM extended spectrum beta-lactamase)	0	1	1	1	0	0	0	1	0	0
D26104-C11 (TEM extended spectrum beta-lactamase)	0	1	1	0	0	0	0	2	0	0
D26104-D7 (TEM extended spectrum beta-lactamase)	0	1	1	1	0	0	0	2	1	0
D26104- E11(Aminoglycoside adenylyltransferase)	0	0	1	0	1	0	0	1	0	0

Advanced BLASTX searches were used, a percentage match of >95 was considered positive, all 74 genomes available were screened (September 2009) the table above shows the number of strains positive for a given sequence, serovars with 0 positives were excluded from this summary, with the exception of *S.* Typhimurium.

4.3.2.3. Distribution of *S.* Typhimurium prophage regions in a panel of African and UK isolates as determined by PCR assay

In 2009 Kingsley and co-workers genome sequenced the Malawian paediatric isolate D23580, The annotated D23580 genome, which is available from the EBI website: ftp://ftp.era.ebi.ac.uk/vol1/ERA000076, was interrogated for the presence and location of D26104 SSH sequences. Kingsley *et al* identified six distinct prophage regions in D23580. Using Artemis it was possible to map the *S.* Typhimurium D26104 sequences, derived from SSH, D26104-ST1, -E02, -C01, -B06, -A10 to BTP1 and D26104-H11, -11, -B04 and –F09 to BTP5, thereby allocating 9 out of the 12 phage-related sequences identified in D26104 to these two prophages. It was not possible to map any of the remaining five putative transposases identified in D26104 to the prophage regions described in D23580. Sequence D26104-B06 partially matches the packaged DNA stabilization protein STM_MW03841 of BTP1; D26104-E02 and D26104-A10 were identified as regulatory protein and DNA transfer protein, respectively and matched corresponding sequences of BTP1 (STM_MW03481 and STM_MW03891). D26104-H11 partially matches a phage tape measure protein (STM_MW31771), while F09 matches the bacteriophage replication protein (STM_MW32051) of BTP5. None of the nine plasmid related D26104-SSH sequences was located on the D23580 virulence plasmid, all of which matched sequences related to the colicin plasmid ColE1 of *E. coli*. Figure 4.8. shows a circular representation of the D23580 genome; highlighted in green on circle 4 are the novel prophage regions identified by Kingsley and co-workers (2009), while circle 5 shows the location of D26104-SSH sequences in magenta.

Using PCR primer sets from this study it was possible to identify all six prophages in a panel of *S.* Typhimurium from Uganda, Malawi, DRC, Kenya and the UK (Figure 4.9. shows an example of prophage PCR assays). BTP5 (O) was completely absent from the UK isolates. BTP2 was absent both Kenyan and UK strains. BTP6 was present in 100 percent of all the *S.* Typhimurium tested (Table 4.9).

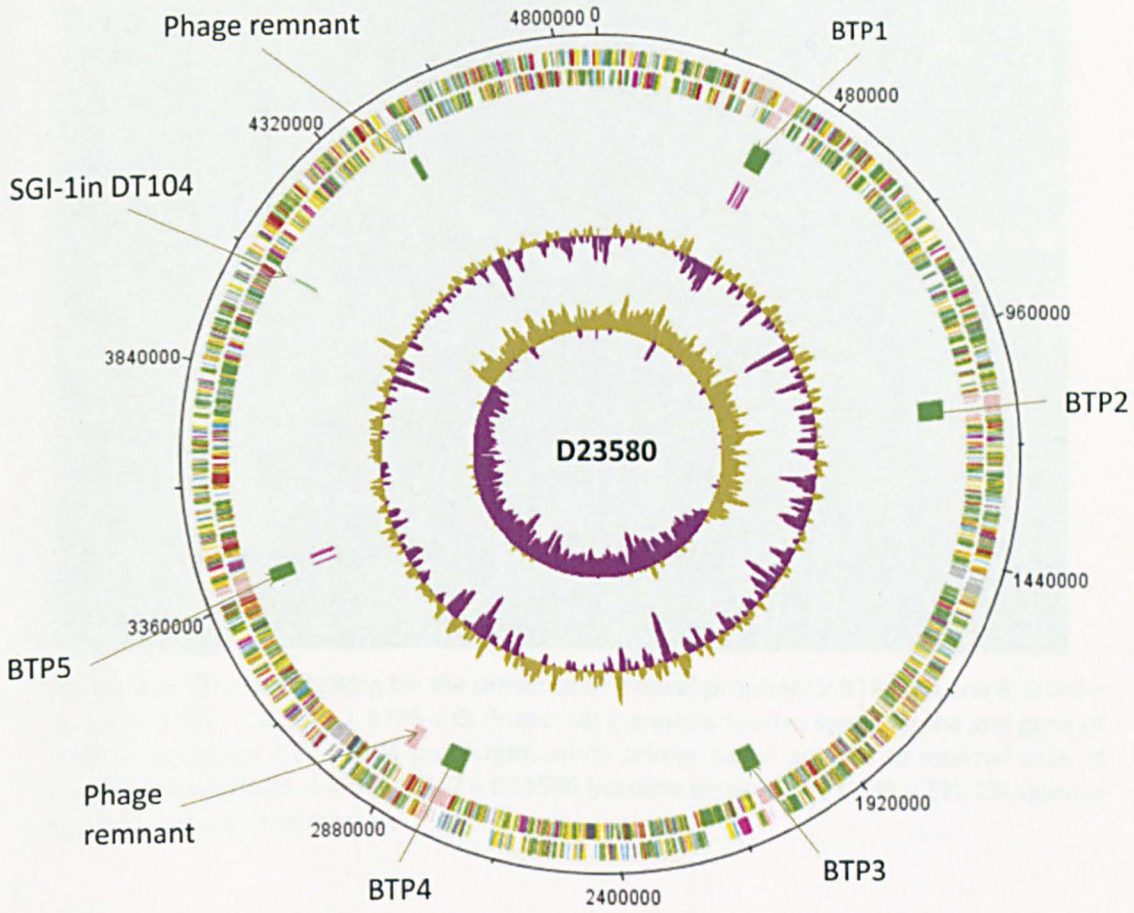


Figure 4.8. Circular diagram showing the D23580 genome: from the outside in, the outer **Circle 1** shows the size in base pairs. **Circles 2 and 3** show the position of CDS transcribed in a clockwise and anti-clockwise direction, respectively. **Circle 4** shows the position of *S.* Typhimurium prophages (BTP) and remnant phages **Circle 5** marks the position of the 9 D26104-SSH sequences (magenta) **Circle 6** shows a plot of G + C content (in a 10-kb window). **Circle 7** shows a plot of GC skew ([G _ C]/[G + C]; in a 10-kb window). Genes in **circles 2 and 3** are colour-coded according to the function of their gene products: dark green, membrane or surface structures; yellow, central or intermediary metabolism; cyan, degradation of macromolecules; red, information transfer/cell division; cerise, degradation of small molecules; pale blue, regulators; salmon pink, pathogenicity or adaptation; black, energy metabolism; orange, conserved hypothetical; pale green, unknown; and brown, pseudogenes.

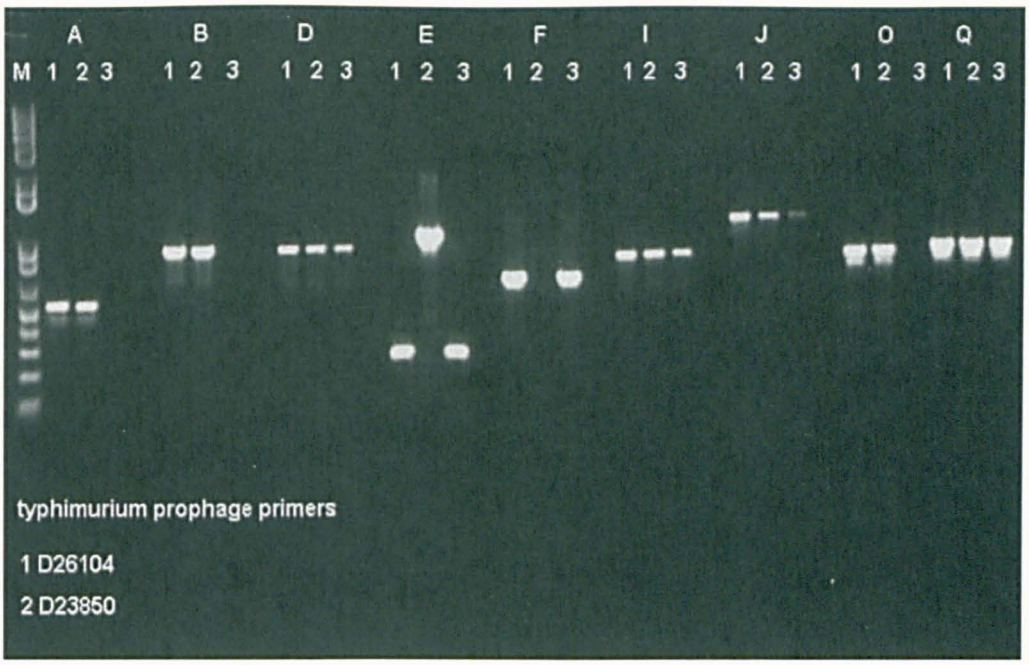


Figure 4.9 PCR assays testing for the presence of 6 novel prophages: BTP1 = A and B, BTP2 = D and E, BTP3 = I, BTP4 = J, BTP5 = O. Primer set E anneals to sites spanning the ssel gene of Gifsy-2, prophage 2 of DT104 prophages, while primer pair F anneals to internal sites of prophage 3 of DT104. 1 = D26104, 2 = D23580 (positive control), 3 = G639 (LT2) 1% agarose gel, M = 1 kb plus size marker.

Table 4.4. Presence and distribution of D23850 novel Prophages (Kingsley *et al.*, 2009) according to PCR assays

S. Typhimurium D23580 prophage regions		Uganda (N=9)	Malawi (N=17)	DRC (N=6)	Kenya (N=4)	UK (N=6)
A BTP1 left hand insertion site		100	82	83	100	50
B BTP1 internal		100	88	83	100	83
D BTP2 internal		100	100	100	100	83
E BTP2 sites spanning *ssel*	IS200 insertion	33	82	17	0	0
	No insertion	44	6	67	100	100
F prophage 3 internal		44	6	50	100	100
I BTP3 internal		89	94	100	75	83
J BTP4 left hand insertion site		89	94	100	100	100
O BTP5 internal		100	100	83	75	0
Q BTP6 internal		100	100	100	100	100

4.3.2.4. ColE1 plasmid-related SSH sequences

The presence or absence of ColE1 related sequence D26104-9 was tested on plasmid DNA using Southern Blotting. Seven *S.* Typhimurium strains (812, 868, 4448, 9664, 9812, 4ST and 5ST) that tested positive for ColE1 (D26104-9) and the tester strain D26104 were selected for plasmid extraction, ColE1 PCR negative strains G639, 146 and 884 were used as negative controls. Using the Benchwork software, suitable restriction enzymes were selected that would cut pSLT and the ColE1 plasmid, without cutting the actual ColE1 gene. *Dra*I, *EcoR*V, *Ava*I, *Pst*I, *EcoR*I and *Rsa*I, were found suitable and tested in a trial digestion (Figure 4.10)

A control PCR assay using D26104-9 primers was carried out on the plasmid DNA using D26104 genomic DNA and plasmid DNA as a positive control and G639 (LT2), 146 and 884 plasmid DNA as a negative control (Figure 4.11.). A Digoxigenin-labelled probe was produced from D26104 (Figure 4.12.) and used to probe a DNA blot of plasmid DNA. D26104, 812, 4234 and 4ST plasmid tested positive for the ColE1 related sequence D26104-9; all four gave multiple signals for all three restriction enzymes tested. This could indicate that the gene has been cut and/or that there are multiple copies of the gene present on the plasmid (Figure 4.13).

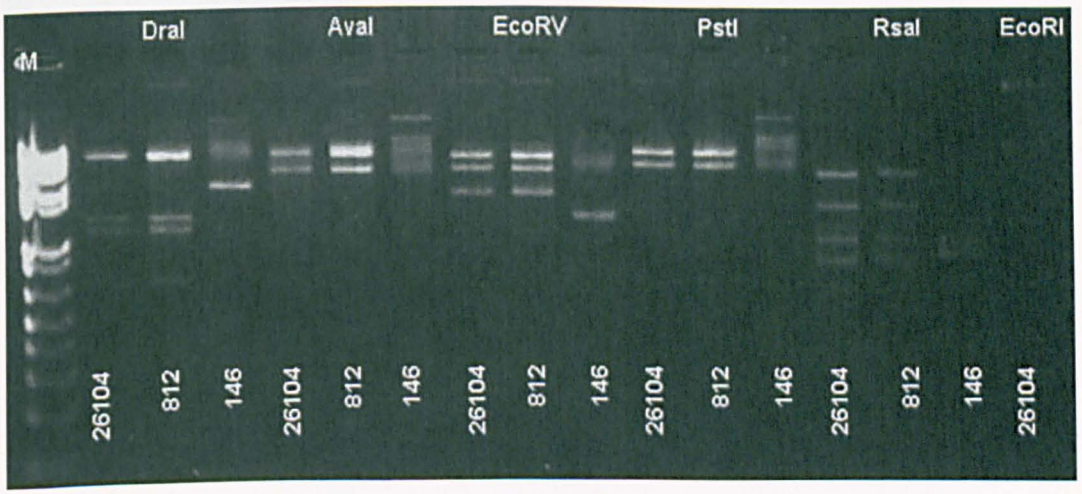


Figure 4.10. RE digest using 6 different restriction enzymes on 3 *S.* Typhimurium plasmid DNA samples 1.2% agarose gel, M = 1 kb plus marker

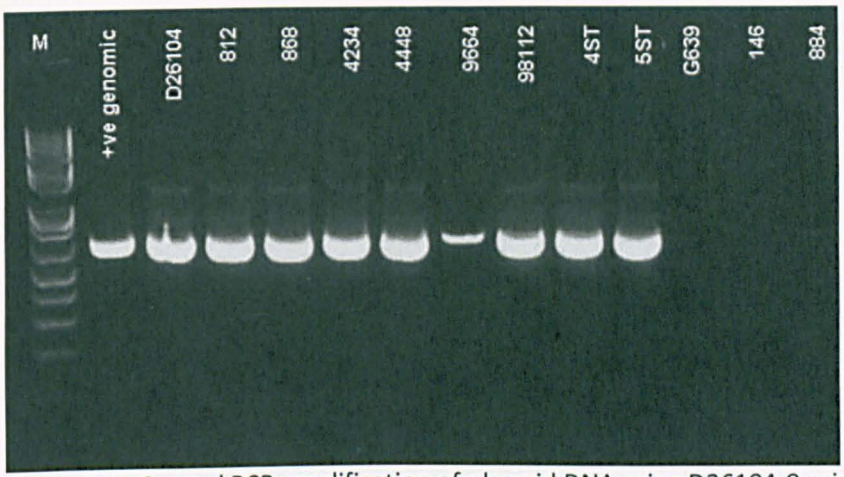


Figure 4.11. Control PCR amplification of plasmid DNA using D26104-9 primers, +ve genomic = genomic DNA of D26104, G639, 146 and 884 plasmid DNA acts as negative controls, 1% agarose gel, M = 1 kb plus marker

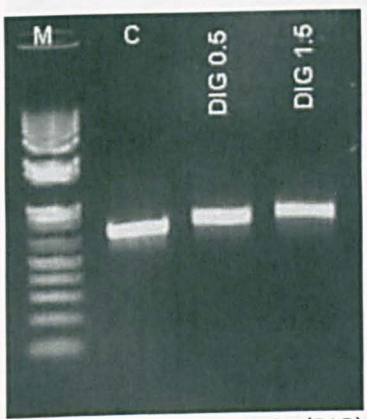


Figure 4.12. Digoxygenin (DIG)-labelling PCR amplification of D26104 genomic DNA using primers D26104-9, C = control D26104 genomic DNA and D26104-9 primers without DIG 1% agarose gel, M = 1 kb plus marker

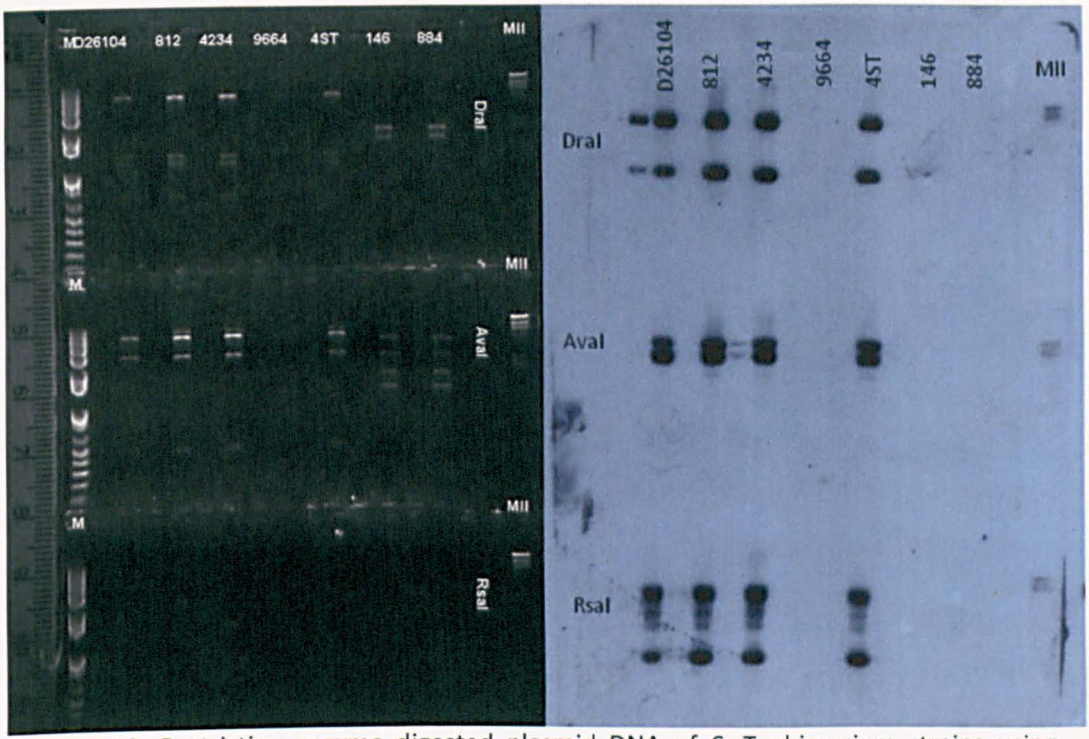


Figure 4.13. Restriction enzyme digested plasmid DNA of *S.* Typhimurium strains using three different restriction enzymes (left), resulting blot after DIG-labelled ColE1 probe has hybridised (right), M 1 kb plus marker, MII = DIG-labelled lambda HindIII marker

4.3.2.2. Distribution analysis of 2 S. Enteritidis D21685 SSH sequences

4.3.2.2.1. SSH sequence distribution of two *S.* Enteritidis D21685 sequences using PCR assays.

PCR primers were designed for two *S.* Enteritidis D21685 SSH sequences, a putative gene for a ROM-like protein (D21685-3) and a putative gene for a protein sequence of unknown function that is derived from two partial sequences (D21685-F7 and –SE8). Distribution of these two sequences across the whole panel of isolates showed these two sequences to be present in three *S.* Enteritidis strains: 1181 from Uganda and D23145 and D23682 from Malawi. The two sequences were absent from all other Enteritidis and all other strains of different serovars from both the panel and the genome database (Figure 4.8. and Supplementary data Tables S1 and S2).

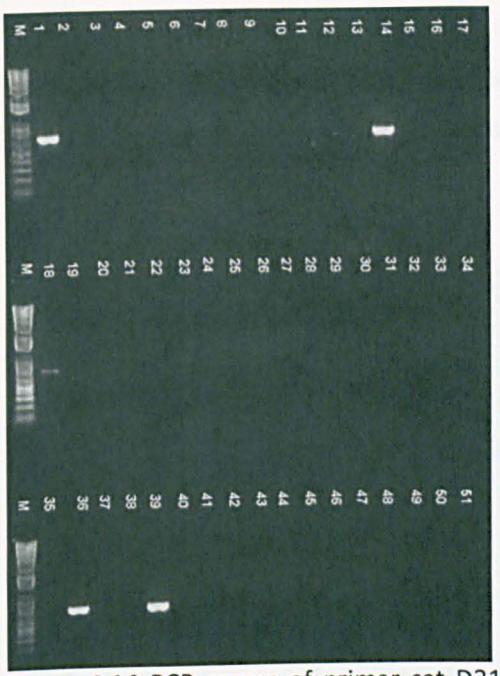


Figure 4.14 PCR assays of primer set D21685-F7/SE8, 1 = positive control D21685, 14 = 1181, 36 = D23145, 39 = D23682 (1% agarose gel, marker (M) 1 kb plus)

4.3.2.3. Distribution of 14 *S.* Heidelberg D23734 SSH sequences

4.3.2.3.1. Distribution of *S.* Heidelberg D23734 SSH sequences using PCR assays

Fourteen S. Heidelberg D23734 SSH sequences with potential roles in virulence according to best BLASTX matches were chosen for distribution analysis.

The collection of invasive NTS isolates from Africa is dominated by *S.* Typhimurium and *S.* Enteritidis. In order to screen for common sequences amongst NTS invasive isolates in this collection regardless of serotype, greater numbers of these more common serovars were included in the panel, alongside *S.* Heidelberg and other less common serotypes. However, none of the invasive *S.* Typhimurium isolates tested positive for any of the subtracted sequences. SSH sequence D23734-G3 represents a putative fimbrial sub-unit present in a restricted number of serovars. All *S.* Enteritidis and *S.* Bovismorbificans, and the majority of *S.* Heidelberg isolates were PCR-positive for this sequence (Table 4.5). According to the PCR assays, two other fimbriae-related sequences (representing *tcf* and *stk* fimbriae respectively) were either present only in *S.* Heidelberg isolates (*stk*) or found in all *S.* Heidelberg isolates, but also in some other serovars (*tcf*). Either *S.* Enteritidis or *S.* Bovismorbificans was PCR-positive for three other SSH sequences, but all *S.* Enteritidis and *S.* Bovismorbificans isolates were PCR-negative for 10 of the SSH sequences (Table 4.5.). For eight of the 14 SSH sequences, all *S.* Heidelberg isolates

were PCR-positive. We found no evidence among the African isolates for distribution according to geographical source (Supplementary data Table S1). Figure 4.15. shows an example of a PCR assay for the distribution of D23734-D2 among NTS isolates from Kenya.

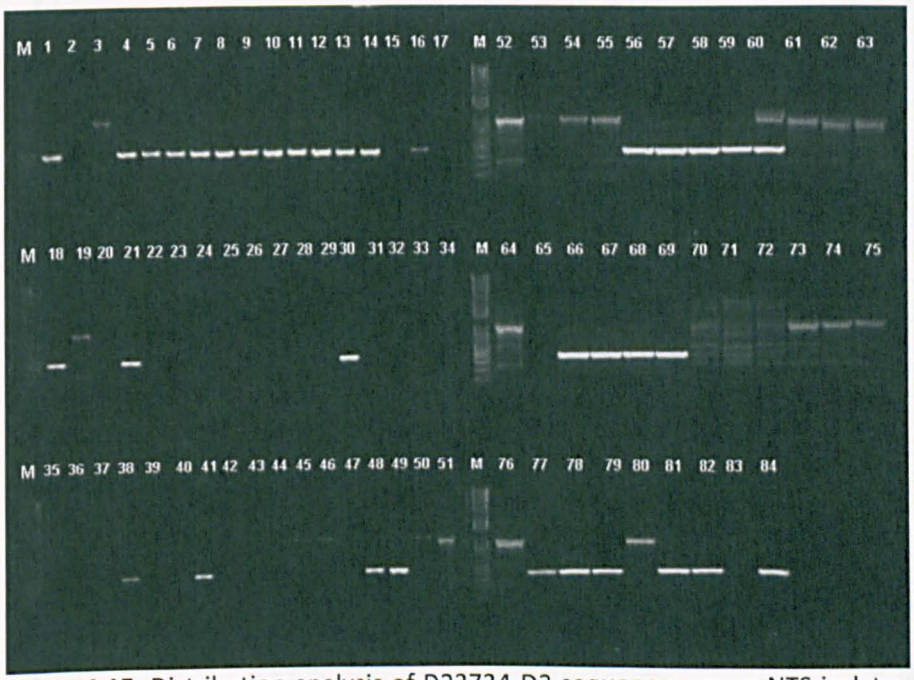


Figure 4.15. Distribution analysis of D23734-D2 sequence among NTS isolates, 1 = positive control D23734, 2 = negative control G639. A 1% agarose gel was used, M= 1kb plus marker.

Table 4.5. Distribution of 14 *S.* Heidelberg D23734 SSH sequences according to PCR assays

SSH sequence putative function based on BLASTX (and sequence identifier)	Sub-Saharan African and Tropical Isolates					UK Isolates			
	S. Typhimurium (n = 38)	*S.* Enteritidis (n = 27)	*S.* Heidelberg (n = 12)	*S.* Bovismorbificans (n = 5)	Other *Salmonella* serovars (n = 3)	*S.* Typhimurium (n = 6)	*S.* Enteritidis (n = 6)	*S.* Heidelberg (n = 3)	*S.* Bovis-morbificans (n = 4)
nitropropane dioxygenase NPD (D23734-4)	-	-	+	-	1 (33)	-	-	+	-
short-chain dehydrogenase/reductase SDR (D23734-10)	-	-	+	-	-	-	-	+	-
putative autotransporter/pertactin (D23734-14)	-	-	11 (92)	+	1 (33)	-	-	+	-
putative fimbrial protein *stkD* (D23734-A1)	-	-	8 (67)	-	-	-	-	2 (66)	-
putative fimbrial protein *tcfA* (D23734-A5)	-	-	+	-	2 (66)	-	-	+	-
fimbrial subunit (D23734-G3)	-	+	10 (83)	+	-	-	+	+	+
probable lipoprotein (D23734-B10)	-	-	7 (58)	+	2 (66)	-	1 (17)	+	+
ferrichrome-iron receptor (D23734-D2)	-	+	10 (83)	-	1 (33)	-	+	+	-
transcriptional regulator *tinR* (D23734-D4)	-	4 (15)	+	-	+	-	1 (17)	+	1 1(25)
conserved hypothetical protein (D23734-B9)	-	-	+	-	-	-	-	+	-
restriction enzyme (D23734-B2)	-	-	+	-	-	-	-	+	-
hypothetical protein (D23734-5)	-	-	+	-	-	-	-	+	-
hypothetical protein (D23734-2)	-	-	+	-	-	-	-	+	-
hypothetical protein (D23734-1)	-	-	11 (92)	-	2 (66)	-	-	+	+

The table shows the number of strains which tested positive for a given sequence by PCR assay; percentages are given in brackets. The group of Sub-Saharan African and Tropical *S.* Heidelberg isolates includes the tester strain D23734. + indicates that all of the strains (100%) were PCR-positive; - indicates that all of the strains tested were PCR-negative.

4.3.2.3.2. A subset of African NTS isolates were screened for the presence of SSH sequences using DNA-DNA Hybridisation techniques (Dot Blot analysis)

The dot blot genomic DNA hybridisation method was employed to screen a panel of 25-53 NTS isolates for the presence of 5 *S.* Heidelberg D23734 SSH sequences. A cross section of strains were used in this technique. Selection was determined by the quality of genomic DNA available. The dots of genomic DNA should be very similar to each other in concentration to avoid false positives. The results obtained varied widely in quality. Figure 4.16 (A) shows a positive control for the membrane which shows that each dot of genomic DNA is capable of hybridisation with a viable probe. Figure 4.16 (B) shows hybridisation of D23734-A5 probes to the same panel of strains, while Figure 4.16 (C) shows hybridisation of D23734-D2 to a more extensive panel of strains. Table 4.6. summarizes the results of dot blot analysis and compares them to PCR assays for the same SSH sequences.

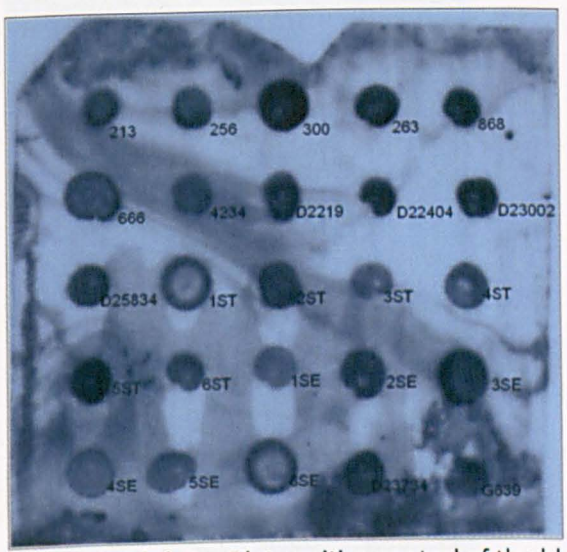


Figure 4.16 (A) Dot Blot positive control of the blot membrane using the 16S ribosomal subunit as a probe (primer pair UNI2/UNI5)

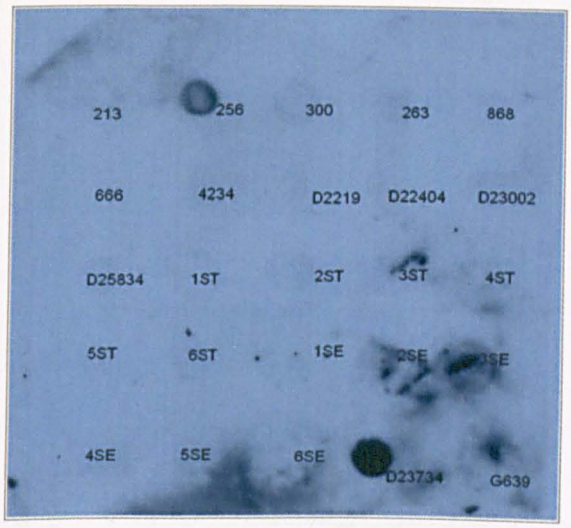


Figure 4.16. (B) Dot Blot using DIG-labelled D23734-A5 DNA as a probe. D23734 acts as a positive control, G639 as a negative control.

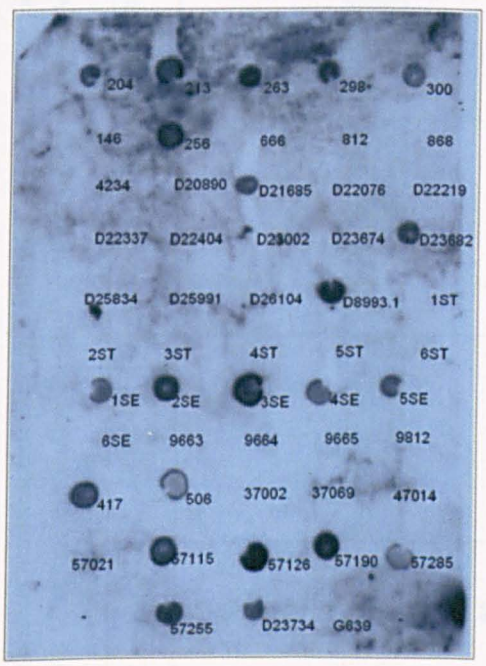


Figure 4.16. (C) Dot Blot using DIG-labelled D23734-D2 DNA as a probe. D23734 acts as a positive control, G639 as a negative control.

Table 4.6 Summary table of distribution analysis comparing Dot Blot (DB) and PCR assay screening results for a panel of NTS isolates, for *S.* Heidelberg SSH sequences D23734-1, -A5, -B10, -D2 and -G3, Origin refers to the country of origin (U=Uganda, M=Malawi, K=Kenya, DRC= Democratic Republic of Congo, UK= United Kingdom)

Strain	serovar	Origin	D23734-1		D23734-A5		D23734-B10		D23734-D2		D23734-G3	
			DB	PCR	DB	PCR	DB	PCR	DB	PCR	DB	PCR
204	*S.* Enteritidis	U	-	-	-	-	-	-	+	+	-	+
213	*S.* Enteritidis	U	-	-	-	-	-	-	+	+	-	+
263	*S.* Enteritidis	U	-	-	-	-	-	-	+	+	-	+
298	*S.* Enteritidis	U	-	-	-	-	-	-	+	+	-	+
300	*S.* Enteritidis	U	-	-	-	-	-	-	+	+	-	+
256	*S.* Stanleyville	U	+	+	+	+	+	+	+	+	-	-
845	*S.* Heidelberg	U	-	+	-	+	-	+	-	+	-	+
868	*S.* Typhimurium,	U	-	-	-	-	-	-	-	-	-	-
4234	*S.* Typhimurium	U	-	-	-	-	-	-	-	-	-	-
D20890	*S.* Typhimurium	M	-	-	X	-	-	-	+	-	-	-
D21685	*S.* Typhimurium	M	-	-	X	-	-	-	-	-	-	-
D22076	*S.* Typhimurium	M	-	-	X	-	-	-	-	-	-	-
D22219	*S.* Typhimurium	M	-	-	-	-	-	-	-	-	-	-
D22337	*S.* Typhimurium	M		-	X	-	-	-	-	-	-	-
D22404	*S.* Typhimurium	M	-	-	-	-	-	-	-	-	-	-
D23002	*S.* Typhimurium	M	-	-	-	-	-	-	-	-	-	-
D23674	*S.* Typhimurium	M	-	-	X	-	-	-	-	-	-	-
D23682	*S.* Typhimurium	M	-	-	X	-	-	-	+	-	-	-
D25834	*S.* Typhimurium	M	-	-	-	-	-	-	-	-	-	-
D25991	*S.* Typhimurium	M	-	-	X	-	-	-	-	-	-	-
D26104	*S.* Typhimurium	M	-	-	X	-	-	-	-	-	-	-
D8993.1	*S.* Enteritidis	M	-	-	X	-	-	-	+	+	-	+
1ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
2ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
3ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
4ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-

5ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
6ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
6.2 ST	*S.* Typhimurium	DRC	-	-	-	-	-	-	-	-	-	-
1SE	*S.* Enteritidis	DRC	-	-	-	-	-	-	+	+	-	+
2SE	*S.* Enteritidis	DRC	-	-	-	-	-	-	+	+	-	+
3SE	*S.* Enteritidis	DRC	-	-	-	-	-	-	+	+	-	+
4SE	*S.* Enteritidis	DRC	-	-	-	-	-	-	+	+	-	+
5SE	*S.* Enteritidis	DRC	-	-	-	-	-	-	+	+	-	+
9663	*S.* Typhimurium	K	-	-	X	-	-	-	-	-	-	-
9664	*S.* Typhimurium	K	-	-	X	-	-	-	-	-	-	-
9665	*S.* Typhimurium	K	-	-	X	-	-	-	-	-	-	-
9812	*S.* Typhimurium	K	-	-	X	-	-	-	-	-	-	-
417	*S.* Enteritidis	K	-	-	X	-	-	-	+	+	-	+
506	*S.* Enteritidis	K	-	-	X	-	-	-	+	+	-	+
37002	*S.* Typhimurium	UK	-	-	X	-	-	-	-	-	-	-
37069	*S.* Typhimurium	UK	-	-	X	-	-	-	-	-	-	-
47014	*S.* Typhimurium	UK	-	-	X	-	-	-	-	-	-	-
57021	*S.* Typhimurium	UK	-	-	X	-	-	-	-	-	-	-
57115	*S.* Entertidis	UK	-	-	X	-	-	-	+	+	-	-
57126	*S.* Enteritidis	UK	-	-	X	-	-	-	+	+	+	+
57190	*S.* Enteritidis	UK	-	-	X	-	-	-	+	+	-	+
57285	*S.* Enteritidis	UK	-	-	X	-	-	-	+	+	-	+
57255	*S.* Enteritidis	UK	-	-	X	-	-	-	+	+	-	+
D23734 +ve	*S.* Heidelberg	M	+	+	+	+	+	+	+	+	+	+
G639 -ve	*S.* Typhimurium LT2	-	-	-	-	-	-	-	-	-	-	-

+ positive, - negative, X not tested

4.3.2.3.3. Distribution of 14 *S.* Heidelberg D23734 SSH sequences among genome sequenced *Salmonella* serovars

We searched *Salmonella* genomes in the database for the presence of the 14 SSH sequences. 13 of the SSH sequences were found to be present in the genomes of both *S.* Heidelberg strains SL476 and SL486. However, D23734-B10, part of a gene encoding a putative lipoprotein, was absent from both genomes. After *S.* Heidelberg, D23734 SSH sequences were most commonly found in *S.* Paratyphi A (Table 4.7).

Table 4.7. BLASTX screening results for D23734 SSH sequences

SSH sequence putative function (and sequence identifier)	*S.* Heidelberg (SL476)	*S.* Heidelberg (SL486)	*S.* Choleraesuis (SC-B67)	*S.* Typhi (Ty2)	*S.* Typhi (CT18)	*S.* Agona (SL483)	*S.* Javiana (GAMM 040433)	*S.* Kentucky (CDC191)	*S.* Kentucky (CVM29188)	*S.* Saintpaul (SARA23)	*S.* Saintpaul (SARA29)	*S.* Schwarzengrund (CVM19633)	*S.* Schwarzengrund (SL480)	*S.* Newport (SL254)	*S.* Newport (SL317)	*S.* Paratyphi A (ATCC9150)	*S.* Dublin (CT_02021853)
2-nitropropane dioxygenase NPD (D23734-4)	+	+	-	-	-	+	-	+	+	-	-	+	+	-	-	-	-
short-chain dehydrogenase/reductase SDR (D23734-10)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
putative autotransporter/pertactin (D23734-14)	+	+	-	-	-	-	-	-	-	-	+	-	-	-	-	+	+
putative fimbrial protein *stkD* (D23734-A1)	+	+	-	-	-	-	-	+	+	-	-	-	-	-	-	+	-
putative fimbrial protein *tcfA* (D23734-A5)	+	+	+	+	+	-	-	-	-	-	-	+	+	-	-	+	-
fimbrial subunit (D23734-G3)	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	+
probable lipoprotein (D23734-B10)	-	-	-	+	+	-	-	-	-	-	+	-	-	+	+	+	-
ferrichrome-iron receptor (D23734-D2)	+	+	-	-	-	+	-	+	+	-	-	-	-	-	-	+	-
transcriptional regulator *tinR* (D23734-D4)	+	+	+	+	+	-	+	-	-	-	-	+	+	-	-	+	-
conserved hypothetical protein (D23734-B9)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
restriction enzyme (D23734-B2)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
hypothetical protein (D23734-5)	+	+	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-
hypothetical protein (D23734-2)	+	+	-	-	-	+	-	+	+	-	-	-	-	-	-	-	-
hypothetical protein (D23734-1)	+	+	-	+	+	+	+	+	+	-	-	+	+	-	-	+	-

4.3.3. *S.* Heidelberg cell invasion assays

The invasiveness of a subset of *S.* Heidelberg strains into nonphagocytic host cells was determined by adding bacteria to Rabbit kidney cells (RKC).

Veterinary strains KMS1977, the sporadic isolate 20070502, as well as the Ugandan bacteraemia isolate 845 were found to be significantly more invasive than *S.* Typhimurium LT2 and the *S.* Heidelberg tester strain D23734, which showed the same level of invasiveness (Figure 4.17).

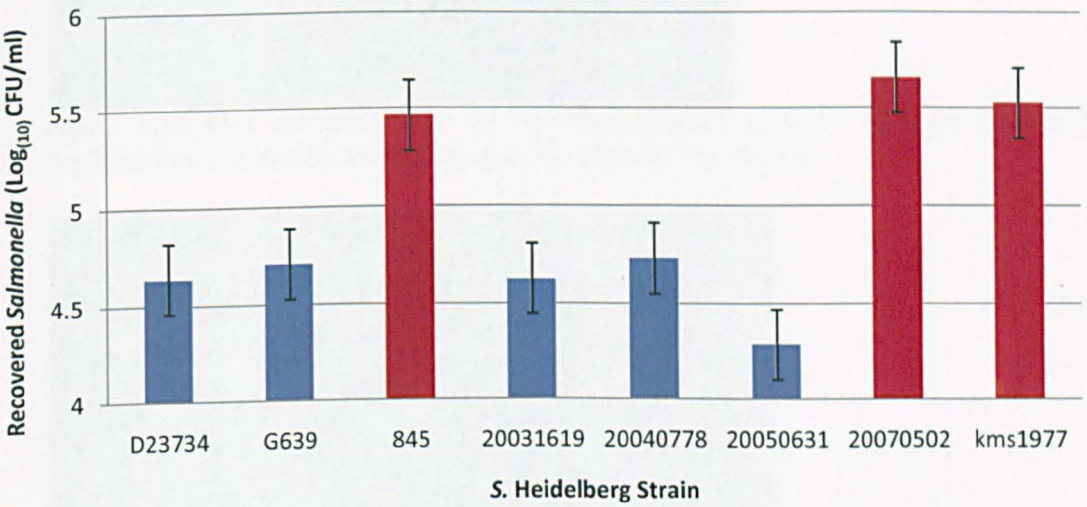


Figure 4.17 showing *Salmonella* cells recovered after host cell lysis (CFU), red bars indicate a statistically significant difference in numbers CFU, compared to the control *S.* Typhimurium LT2 (G639).

4.3.4. Expression of genes represented by *S.* Heidelberg D23734 SSH sequences

A subset of six *S.* Heidelberg strains was chosen for RT-PCR screening to detect the expression profile of seven of the SSH sequences. Figure 4.18. shows an electrophoresis agarose gel of RNA isolated from *S.* Heidelberg strains; RNA was subsequently reverse-transcribed to cDNA (Figure 4.19), which was used as a template in PCR assays using SSH sequence primers. The results are shown in Table 4.8. The majority of SSH sequences lie within genes expressed in all of the African and Tropical *S.* Heidelberg isolates. The exception was SSH sequence D23734-A1 (putative *stk* fimbrial gene), for which expression could not be detected in three of the isolates. For three of the SSH sequences, expression could not be detected in the UK veterinary isolate KMS1977 (Table 4.8).

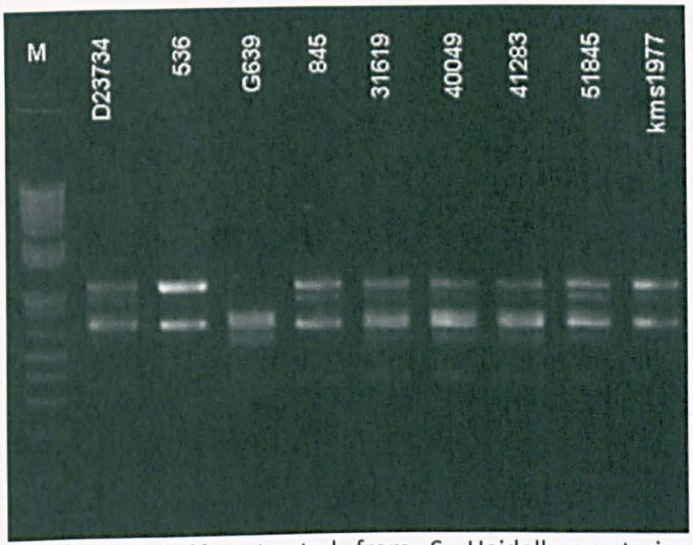


Figure 4.18 RNA extracted from *S.* Heidelberg strains and the negative control *S.* Typhimurium LT2 (G639), 1% agarose gel, M = 1kb plus size marker

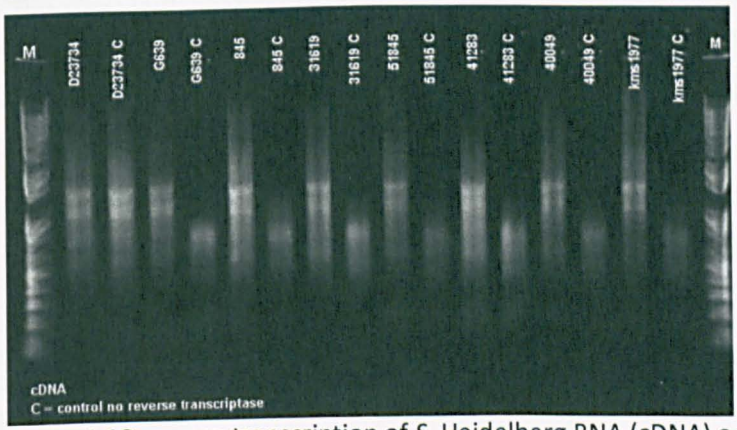


Figure 4.19 reverse transcription of *S.* Heidelberg RNA (cDNA) c = control with no reverse transcriptase added. 1% agarose gel, M = 1kb plus size marker.

Table 4.8. Expression profile of seven SSH sequences among a panel of six *S.* Heidelberg strains

	D23734 (Malawi)	845 (Malawi)	20031619 (Zanzibar)	20040049 (Kenya)	20041283 (Peru)	KMS1977 (UK Vet.)
putative autotransporter/pertactin (D23734-14)	+	+	+	+	+	+
putative fimbrial protein *stkD* (D23734-A1)	+	+	-	-	-	+
putative fimbrial protein *tcfA* (D23734-A5)	+	+	+	+	+	+
fimbrial subunit (D23734-G3)	+	+	+	+	+	-
probable lipoprotein (D23734-B10)	+	+	+	+	+	+
ferrichrome-iron receptor (D23734-D2)	+	+	+	+	+	-
transcriptional regulator *tinR* (D23734-D4)	+	+	+	+	+	-

4.3.2.4. Distribution of seven S. Bovismorbificans 3114 SSH sequences

4.3.2.4.1. Distribution of seven S. Bovismorbificans 3114 SSH sequences among a panel of African and UK NTS isolate using PCR assays

Table 4.9 shows the distribution of seven SSH sequences among Malawian bacteraemia *S.* Bovismorbificans isolates from adults and children and UK veterinary isolates. Five of the SSH sequences match putative prophages within the 3114 genome (3114-18, -A9, B6, -G4, -F10). BLAST searches indicate that 3114-F10 is related to Gifsy-1 and PCR assays showed that phage sequence 3114-F10 was present in a high percentage of both human and veterinary isolates (Figure 4.20 and Table 4.9.). However, it is noteworthy that phage sequences 3114-18, -A9, -B6 and –G4 were PCR-negative in the veterinary strains but positive in up to 89% of children's isolates and 80% of adult bacteramia isolates. Table 4.10 shows the distribution of the same seven SSH sequences among a panel of African bacteraemia and UK diarrhoea isolates of different serovars according to PCR, mainly Typhimurium and Enteritidis. The main observation is the apparent absence of SSH sequences in *S.* Typhimurium isolates from both Africa and the UK. Sequence 3114-11, matching a gene encoding a putative phosphotransferase system component was present in a high percentage of *S.* Enteritidis isolates. Prophage sequences 3114-18,-A9,-B6 and -G4 were PCR-negative in all the isolates tested. Table 4.11 shows the distribution of 3114-SSH sequences among a selection of genome-sequenced *S. enterica* currently available in the NCBI database (October, 2010). The complete table is included in the supplementary data (Supplementary data, S3). Interestingly five out of seven 3114 SSH sequences (3114-18, -A09, -B06, -G04, and -F10) were completely absent from all 36 serovars currently available for interrogation by BLAST searches (including complete and incomplete published *S. enterica* genomes). The putative phosphotransferase system component (3114-11) was widely distributed and was particularly common amongst *S.* Typhi isolates (12 out of 14), as well as in Paratyphi A and B. The putative outer membrane usher protein (3114-C12) was present in serovars Hadar, Newport and Kentucky (Table 4.10).

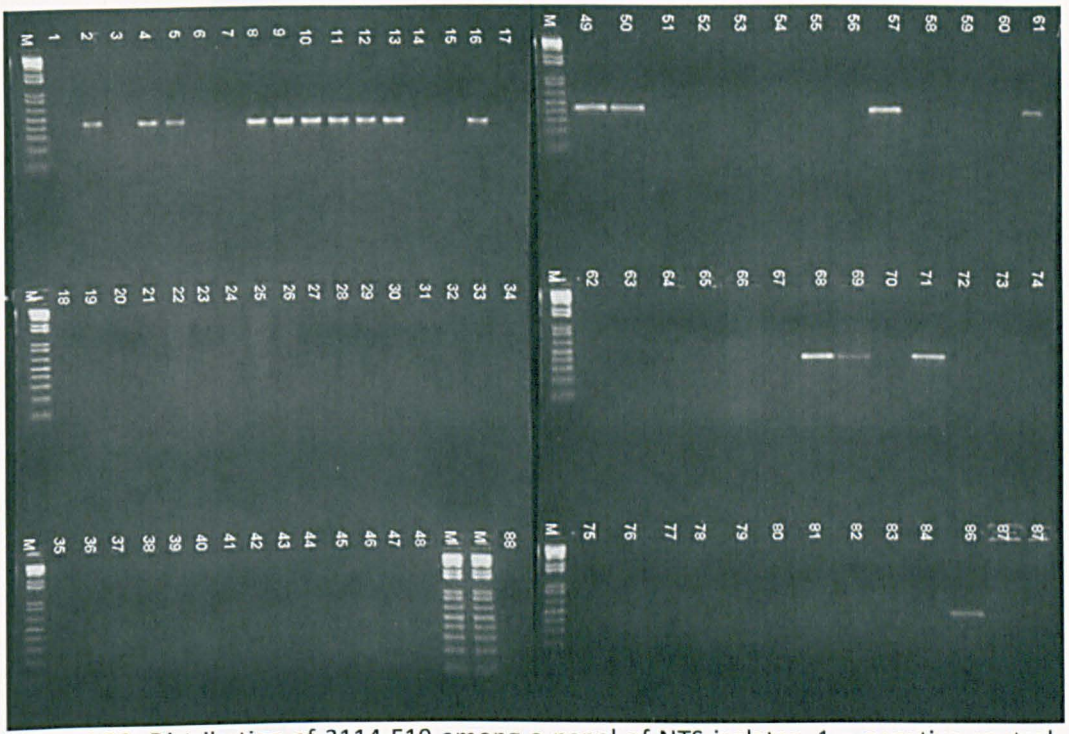


Figure 4.20. Distribution of 3114-F10 among a panel of NTS isolates, 1 = negative control G639, 2 = positive control 3114, 1% agarose gel, M = 1 kb plus marker

Table 4.9. Distribution of 7 *S.* Bovismorbificans SSH sequences among a panel of African bacteraemia and UK veterinary *S.* Bovismorbificans isolates (determined using PCR assays)

	Leahurst		Malawi					
	total		total		adults		kids	
	N=16	%	N=50	%	N=22	%	N=28	%
putative phosphotransferase system component (3114-11)	12	75	44	88	18	82	26	93
peptidase S14, ClpP (3114-18) – ROD34	0	0	39	78	14	64	25	89
phage terminase GpA (3114-A9) -ROD34	0	0	39	78	14	64	25	89
putative tail fiber protein (3114-B6) – ROD13	0	0	38	76	13	59	24	86
hypothetical protein Stx2Ip072 (3114-G4) ROD13	0	0	40	80	15	68	25	89
Gifsy-1 prophage terminase large chaing gp2 (3114-F10)	15	94	41	82	17	77	24	86
putative outer membrane usher protein (3114-C12)	16	100	42	84	17	77	25	89

Table 4.10. Distribution according to PCR assays of 7 *S.* Bovismorbificans SSH sequences among a panel of African bacteraemia and UK diarrhoea isolates of different serovars (STM = *S.* Typhimurium,Ent = *S.* Enteritidis)

	Uganda			Malawi			DRC		Kenya		UK	
	STM n=9	Ent n=13	other n=2	STM n=18	Ent n=5	other n=2	STM n=7	Ent n=5	STM n=4	Ent n=4	STM n=6	Ent n=6
putative phosphotransferase system component (3114-11)	0	13	1	0	4	2	0	5	0	4	0	6
peptidase S14, Clp (3114-18)	0	0	0	0	0	0	0	0	0	0	0	0
phage terminase GpA (3114-A9)	0	0	0	0	0	0	0	0	0	0	0	0
putative tail fiber protein (3114-B6)	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein Stxp072 (3114-C12)	0	0	0	0	0	0	0	0	0	0	0	0
Gifsy-1 prophage terminase (3114-F10)	0	9	0	0	2	0	0	2	0	0	0	0
putative outer membrane usher (3114-G4)	0	0	0	0	0	0	0	0	0	0	0	0

Table 4.11 BLAST search results for 7 3114 SSH sequences

	Typhimurium (n=10)	Choleraesuis (n=2)	Dublin (n=3)	Enteritidis (n=5)	Gallinarum (n=2)	Hadar (n=1)	Heidelberg (n=2)	Infantis (n=1)	Javiana (n=1)	Kentucky (n=2)	Newport (n=4)	Paratyphi A (n=2)	Paratyphi B (n=3)	Paratyphi C (n=2)	Schwarzengrund (n=3)	Saintpaul (n=2)	Typhi (n=14)	Weltevreden (n=1)
putative phosphotransferase system component (3114-11)	0	0	1	1	1	1	0	0	1	0	2	2	1	0	2	1	12	1
peptidase S14, ClpP (3114-18) ROD34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phage terminase GpA (3114-A9) ROD34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
putative tail fiber protein (3114-B6) ROD13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
hypothetical protein Stx2Ip072 (3114-G4) ROD13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gifsy-1 prophage terminase gp2 (3114-F10)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
putative outer membrane usher protein (3114-C12)	0	0	0	0	0	1	0	0	0	2	2	0	0	0	0	0	0	0

4.4. Discussion

4.4.1. NTS serovars

S. Typhimurium ranks amongst the top three human and non-human isolates in the world. *S.* Enteritidis also ranks highly amongst human isolates; in the US *S.* Enteritidis was the second commonest human isolate, while *S.* Heidelberg was the commonest non-human isolate in 2005. *S.* Bovismorbificans ranked third most common amongst non-human isolates in Australia (reviewed by Lan *et al.*, 2009).

4.4.2. *S. enterica* MLST

To date, 1389 STs have been identified in 4202 *S. enterica* isolates (http://pubmlst.org/databases.shtml). There is evidence that at least 11 serovars are not composed of a single clone, including *S.* Typhimurium and *S.* Enteritidis, *S.* Typhimurium is the most divergent serovar. The major lineage of *S.* Typhimurium is ST19, with 17 STs belonging to the same clonal complex with ST19 as its founder (Figure 4.2 and 4.3). ST19 includes previously genome sequenced *S.* Typhimurium isolates LT2, SL1344 and NCTC13348 (Kingsley *et al.*, 2009). Kingsley and co-workers first described the novel ST313 of invasive *S.* Typhimurium isolates from Kenya and Malawi. All of the 31 adult Malawian isolates in their study were ST313, while 13/20 adult Kenyan isolates were ST313, and the remaining matched ST19. Previous to the Kingsley et al (2009) study, only two ST313 isolates have been published in the MLST database, both of which came from outside of sub-Saharan Africa, but interestingly came from patients with invasive disease.

We carried out MLST on 17 invasive African *S.* Typhimurium isolates, including 4 adult isolates from Uganda and 13 paediatric isolates from Malawi. Two of the Ugandan and nine Malawian isolates matched ST313. One Ugandan and three Malawian isolates were ST313cc. One Ugandan isolate (146) was an equal match to ST313 and ST394 (Table 4.1). ST394 is a close relative of ST313, which was also identified in one Malawian adult isolate by Kingsley and co-workers (Kingsley *et al.*, 2009). Another interesting observation is that the Ugandan isolates pre-date the Malawian isolates by 1-7 years, which may indicate an emigration of this invasive clone from the North to the South of sub-Saharan Africa.

4.4.3. Distribution of *S.* Typhimurium D26104 SSH sequences

The distribution analysis of D26104 SSH sequences as determined by PCR assays showed that some sequences appear to follow a geographical distribution pattern, in that gp24 of phage PsP3 (D26104-H11), Tnp (D26104-ST2) and Colicin E1 (D26104-9) were absent from strains from the UK and other Malawi isolates than D26104, Tum of phage 186 (D26104-11) was absent from the UK, while EaA of phage P22 (D26104-ST1) appears to be absent from Kenyan and UK strains but was present in 100% of Ugandan and Malawi strains (Table 4.2).

Bacteriophage-related D26104 sequences were largely absent from genome sequenced isolates of *S. enterica*, There were some exceptions. EaA of phage 22 (D26104-ST1) was found in both *S.* Paratyphi isolates and one *S.* Saintpaul isolate; the transposase (D26104-ST2) was found in one isolate of *S.* Schwarzengrund (Table 4.3).

In 2009 Kingsley and co-workers genome sequenced the Malawian paediatric invasive isolate D23580, which was isolated in 2004 and is resistant to amoxicillin, tetracycline, sulphamethoxazole and cefuroxime. Genotyping, using PFGE has been carried out previously on the Malawi collection, and indicated that D26104 and D23580 belong to different genotypes (Prof C.A. Hart, personal communication). The annotated D23580 genome, which is available from the EBI website: ftp://ftp.era.ebi.ac.uk/vol1/ERA000076, was interrogated for the presence and location of D26104 SSH sequences. Kingsley *et al* identified six distinct prophage regions in D23580. Using Artemis it was possible to map D26104-ST1, -E02, -C01, -B06, -A10 to BTP1, and D26104-H11, -11, -B04 and –F09 to BTP5, thereby allocating 9 out of the 12 phage-related sequences identified in D26104 to these two prophages. The NTS panel tested by PCR included D23580, which was PCR negative for D26104-H11. It was not possible to map any of the remaining five putative transposases identified in D26104 to the prophage regions described in D23580. Sequence D26104-B06 partially matches the packaged DNA stabilization protein STM_MW03841 of BTP1; D26104-E02 and D26104-A10 matched a regulatory protein and DNA transfer protein, respectively and matched corresponding sequences of BTP1 (STM_MW03481 and STM_MW03891). D26104-H11 partially matches a phage tape measure protein (STM_MW31771), while D26104-F09 matches the bacteriophage replication protein (STM_MW32051) of BTP5. Plasmid-related sequences identified in *S.* Typhimurium D26104 matched sequences related to the colicin plasmid ColE1 of *E. coli*. This finding suggests the

presence of another colicin-like plasmid in D26104. A sequence matching the Colicin E1 protein (D26104-9) was isolated from D26104, these findings led to the investigation of D26104-9 positive strains for the presence of a plasmid. ColE1 and 2 are low molecular theta plasmids, these plasmids are typically rare among *Salmonella* isolates (~10%). ColE1 is most commonly found in *S.* Enteritidis. In the presence of a conjugative plasmid ColE1 plasmids can be mobilised and transferred by conjugation due to a specific set of genes. Since *S.* Enteritidis usually does not possess conjugative plasmids, it is quite probable that these plasmids were acquired by *S.* Enteritidis via conjugation. It is therefore thought that *Salmonella* is a non-natural and terminal host of ColE1 and 2 (Gregorova *et al.*, 2004). Plasmid DNA was isolated from 12 *S.* Typhimurium isolates, using probes for D26104-9 hybridisation experiments were performed to establish the presence of ColE1 related sequence among this plasmid DNA, Southern Blotting results confirmed the findings of previous PCR assays for D26104-9 and the plasmid DNA of four *S.* Typhimurium isolates (812, 4234, 9664, 4ST and D26104) tested positive for D26104-9. However, although care was taken when choosing the restriction enzymes to cut the plasmids, D26104-9 probes gave multiple hybridisation signals. This could indicate that the gene has been cut. It is also possible that the plasmids carry multiple copies of the gene.

Figure 4.8. shows a circular representation of the D23580 genome; highlighted in green on circle 4 are the novel prophage regions identified by Kingsley and co-workers, while circle 5 shows the location of D26104-SSH sequences in magenta. Kingsley and co-workers designed PCR assays in order to test for presence or absence of phages BTP1 to BTP6. It was possible to identify all six prophages in our panel of *S.* Typhimurium from Uganda, Malawi, DRC, Kenya and the UK.

BTP1 is absent from LT2 and is similar to DT104 bacteriophage ST104; primer pair A maps to the left hand insertion site of this phage, while primer pair B maps to an internal site of BTP1. BTP1 was therefore present in 100% of Ugandan and Kenyan isolates, and in 83% of DRC isolates. The internal marker (B) of BTP1 was present in 83% of UK isolates but the insertion site (A) was only present in 50%. Similarly the insertion site was present in 82% of isolates from Malawi whereas the internal site was present in 88% of isolates. D26104-ST1 maps to BTP-1 and was found to be absent from all Kenyan and UK isolates tested. This may suggest a sequence

variation within BTP1. BTP2 is similar to Gifsy-2 with the *sseI* gene disrupted by an IS200 element. Primer pair D maps to sites internal of BTP2; primer pair E maps to sites spanning the *sseI* gene of BTP2, Gifsy-2 and prophage 2 of DT104 prophage. Products of these primers therefore vary in size according to whether the IS200 element is present or absent. Primer site D was present in 100% of Ugandan, Malawi, DRC and Kenyan isolates and was present in 83% of UK isolates. The novel prophage with IS200 disruption was present in 33% of Ugandan, 82% of Malawian and 17% of DRC isolates, but it was absent from Kenyan and UK isolates. Primer pair F spans sites internal to prophage 3 of DT104. Prophage 3 was present in 100% of Kenyan and UK isolates but in much lower numbers in the other African isolates. BTP3 is similar to prophage 4, while BTP4 is similar to Gifsy-1. BTP3 and 4 were both present in high percentages of all the isolates. BTP5 is a novel prophage of the P2 family. BTP5 was completely absent from the UK isolates, which confirms the findings of SSH sequence PCR assays for D26104-H11 and D26104-11, however BTP5 was present in 100 percent of Malawian isolates, of which D26104-H11 and -11 were absent, again this may indicate a variation in sequence of BTP5, where D26104-H11 can be present or absent. BTP2 was absent both Kenyan and UK strains. BTP6 was present in 100 percent of all the *S.* Typhimurium tested. BTP6 appears to be a remnant phage which is present in LT2 and DT104, primer pair Q spans sites internal to BTP6. BTP6 was present in 100% of all the isolates (Table 4.4). The tester strain *S.* Typhimurium D26104 was found to be positive for BTP1, BTP3,Prophage 3, BTP4, BTP5 and BTP6 (data not shown).

Of the total number of D26104 SSH sequences, 15% were related to antimicrobial resistance. Gordon and co-workers found *S.* Typhimurium in Malawi to be resistant to ampicillin and trimethoprim-sulfamethoxazole throughout their 7 year study period and cholarmphenicol resistance was rapidly acquired during 2001-2004. All of the NTS studied were susceptible to ciprofloxacin and third generation cephalosporins (Gordon *et al.*, 2008). A number of TEM extended spectrum beta-lactamases were identified through SSH in D26104 and two were chosen for further distribution studies. We found these sequences to be present in 100% of Uganda and Malawi and >50% in Kenya and the DRC and in a third (33%) of UK diarrhoea isolates (Table 4.2.). TEM ESBLs were also present in genome sequenced strains of *S.* Typhi, Choleraesuis and Schwarzengrund. Sequence D26104-E11 aminoglycoside

adenylyltransferase was present in Choleraesuis, Newport and Schwarzengrund (Table 4.3).

4.4.4. Distribution of *S.* Enteritidis D21685 SSH sequences

Distribution analysis by PCR assays of two D21685 SSH sequences showed that together these were only present in three further *S.* Enterica isolates, two from Malawi and one from Uganda. The ROM-like protein (encoded by putative gene matching D21685-3) is an RNA I modulator, associated with ColE1. Interestingly the ROM-like protein and protein of unknown function (D26104-F7/SE8) occur in the same strains and may therefore be linked in some way.

4.4.5. Distribution of *S.* Heidelberg D23734 SSH sequences

Distribution of 14 D23734 SSH sequences was assessed by PCR assays (Table 4.4) and BLAST searches of genome sequenced *S. enterica* isolates (Table 4.6). There was little variation between the genomes of African invasive and faecal isolates of *S.* Heidelberg. The distribution of five SSH sequences (D26104-1, -A5, -B10, -D2 and –G3) was also assessed by Dot blot analysis (Table 4.5). Although results of Dot Blot analysis and PCR largely agreed, Dot Blot analysis was found to be less sensitive and was therefore abandoned in favour of PCR assays.

D23734-B10 was lacking from the two genome sequenced *S.* Heidelberg strains but present in the majority of African isolates. The presence of this sequence in some other invasive serovars, such as Typhi, suggests that the genomic region represented by this sequence may merit further study. However, we found the SSH sequence to be present also in UK veterinary isolates of *S.* Heidelberg, and demonstrated expression of the putative lipoprotein-encoding gene in all six isolates tested, including one of the veterinary isolates. Hence, there was no clear association of either the presence or the gene, or expression of the gene, with invasive isolates. Likewise, the potentially virulence-related putative autotransporter/pertactin-related SSH sequence D23734-14 was found in 92% of the African/Tropical *S.* Heidelberg isolates and all of the UK veterinary isolates, and was expressed in all *S.* Heidelberg isolates tested. Interestingly, based on PCR assays, this sequence was present in the five African invasive isolates of *S.* Bovismorbificans but absent from UK veterinary isolates of this serovar. The sequence was also present in some other invasive serovars. However, overall there was no clear correlation between SSH sequence

distributions and those serovars generally regarded as more virulent/invasive, including serovars Choleraesuis, Dublin, Schwarzengrund and Newport (Chiu *et al.*, 2006;Threlfall *et al.*, 1992).

S. Heidelberg is in the same serogroup as *S.* Typhimurium (serogroup B, antigen profile 1,4,5,12:r:1,2). In a study of the clonal diversity of eight NTS serovars, using enzyme electrophoresis to detect allelic polymorphisms, it was reported that *S.* Heidelberg, when compared to serovars Choleraesuis, Dublin, Derby, Enteritidis, Typhimurium, Infantis, and Newport, showed the least diversity among strains from Europe and the Americas (Beltran *et al.*, 1988). However, we detected variations between African / Tropical isolates of *S.* Heidelberg with respect to six of the SSH sequences used for PCR assays.

Two of the SSH sequences (D23734-B2 and D23734-B9) were present in all *S.* Heidelberg isolates and in none of the other *Salmonella* serovars tested in this study, including all of those in the database (Tables S1 and S2). Thus, these sequences may represent *S.* Heidelberg-specific markers. Genome sequences are available in the database for two American isolates, SL476 (Genbank accession CP001120), a multidrug resistant strain, and SL486 (ABEL01000001-ABEL1000048) a drug-susceptible strain. With respect to the distribution of SSH sequences, we found no difference between these strains. Indeed, with the exception of *S.* Saintpaul and *S.* Newport, all of the same-serovar genomes showed no difference in terms of SSH sequence distributions. *S.* Saintpaul strain SARA23 was the only strain negative for all 14 SSH sequences, and is described as falling within the main clade of the Saintpaul serovar, whereas its partner strain SARA29, which we found to be positive for three SSH sequences (D23734-5, -14, -B10), has been described as an outlier (www.jcvi.org/salmonella). *S.* Newport SL254 and SL317 are strains from the two distinct lineages that exist within the *S.* Newport serovar (www.jcvi.org/salmonella). We found that both strains carry the lipoprotein-associated SSH sequence D23734-B10, but strain SL254 also carries the fimbrial-associated SSH sequence D23734-G3.

The genomes of *S. enterica* possess numerous fimbrial gene clusters implicated in host colonisation and adaptation. Indeed, the genome of *S.* Typhimurium LT2 alone carries eleven fimbrial operons, some of which have been implicated directly in virulence (Humphries *et al.*, 2001;van der Velden *et al.*, 1998). The repertoire of fimbrial operons varies between serovars, with some widely distributed but others

restricted to a limited number of serovars (Porwollik & McClelland, 2003;Townsend *et al.*, 2001). Our observations indicate that fimbrial gene clusters make a major contribution to the accessory genome of *S.* Heidelberg. The *stk* gene cluster has been reported as specific for *S.* Paratyphi A (Edwards *et al.*, 2002). Based on analysis of the distribution of SSH sequence D23734-A1, to this we can add *S.* Heidelberg and *S.* Kentucky, although our PCR assay data suggest that the operon may not be carried by all *S.* Heidelberg. We also demonstrate that expression of the *stk* operon gene could be detected from some, but not all, of the *S.* Heidelberg isolates tested. Given the proven role of fimbriae in pathogenicity, the variable carriage of this fimbrial operon gene (*stkD*), and the variable expression even amongst those strains carrying the gene, leads us to the conclusion that the role of this operon merits further investigation. As described previously in Chapter 3.3.2. *S.* Typhi carries the fimbriae designated *tcf*, for Typhi colonising factor (Folkesson *et al.*, 1999). S. Heidelberg SSH sequences D23734-A5 and D23734-E10 matched *tcfA* and *tcfD* respectively. In addition, SSH sequence D23734-D4 matches *tinR*, which lies downstream of the *tcf* genes in *S.* Typhi, and encodes a transcriptional regulator has been identified (Kisiela *et al.*, 2005).

However, based on the *S.* Heidelberg SL476 genome, it appears that this proximity does not occur in *S.* Heidelberg, and it was observed that the distributions of *tinR* and *tcfA* differ amongst this strain collection and the serovars represented in the genome database. PCR assays indicated that the SSH sequence D23734-A5 (representing *tcfA*) was present in 100% of African, Tropical and UK isolates of *S.* Heidelberg, as well as in the two genome sequenced isolates SL476 and SL486. Expression of *tcfA* was detected in all six isolates of *S.* Heidelberg tested.

A third fimbrial operon, represented by the SSH sequence D23734-G3, was present in serovars Enteritidis, Bovismorbificans, Saintpaul, Newport and Dublin as well as all but two *S.* Heidelberg isolates, and was expressed in all *S.* Heidelberg isolates tested except for the UK veterinary isolate KMS1977. Expression of genes associated with a further two of the seven SSH sequences tested, encoding a putative ferrichrome-iron receptor and transcriptional regulator respectively, were also expressed in all *S.* Heidelberg isolates except KMS1977. These observations suggest that although we found little evidence for genome content variations between invasive and gastrointestinal isolates of *S.* Heidelberg, we did find variations in gene expression. It is possible that differences in gene expression play a role in the

different pathogenic abilities exhibited by isolates of this serovar. Using cell invasion assays the ability of a subset of seven *S.* Heidelberg isolates (845, D23734, 20031619, 20040778, 20050631, 20070502) to invade non-phagocytic host cells was tested; *S.* Typhimurium LT2 (G639) was used as a control. Invasiveness was measured in colony forming units (CFU) per ml and results for each strain were compared to each other. Barrow and Lovell first carried out invasion assays using *S. enterica* serovars Typhimurium, Infantis Gallinarum and Pullorum, in Vero cells in 1989. *S.* Typhimurium showed the highest level of invasiveness into Vero cells with a median count of 5.3 (4.5-6.1) Log10 CFU/ml in 29 *S.* Typhimurium isolates, which corresponds to the level of 4.7 Log10 CFU/ml found for *S.* Typhimurium LT2 (G639) recovered from Rabbit Kidney Cells (RKC). *S.* Heidelberg tester strain D23734 showed a similar level of invasiveness with 4.6 Log10 CFU/ml, Barrow and Lovell showed an invasion level of 4.5-5.4 Log10 CFU/ml for *S.* Heidelberg. *S.* Heidelberg strains significantly more invasive than LT2 (G639) and D23734 in this study were the veterinary UK isolate KMS1977 (5.5 Log10 CFU/ml), a Ugandan adult bacteraemia isolate 845 (5.5 Log10 CFU/ml) and a sporadic human isolate 20070502 (5.6 Log10 CFU/ml), which are not significantly higher than Barrow and Lovell's food poisoning isolates (*S.* Heidelberg is included in a group of food-poisoning isolates and the data does not specify the exact results for *S.* Heidelberg) (Barrow & Lovell, 1989). Further expression analysis under different growth conditions would be of interest to establish expression of virulence genes for example in the host cell. Barrow and Lovell showed that both cell culture temperature and bacterial culture temperature had an effect on invasiveness; invasiveness fell from >5.4 to <3.6 Log10 CFU/ml across different *S enterica* serovars with decreasing cell culture temperature. Bacterial culture temperature affected invasiveness both positively and negatively depending on serovar (Barrow & Lovell, 1989). Knock-out mutants for certain genes could also be tested in these invasion assays. Van Asten and co-workers demonstrated that knocking out *fliC* in *S.* Enteritidis reduces the ability of *S.* Enteritidis to invade differentiated Caco-2 cells 50-fold, without affecting bacterial adherence. Complementation of *fliC* mutants fully restores invasiveness (Van Asten *et al.*, 2000).

4.4.5. Distribution of *S.* Bovismorbificans 3114 SSH sequences

Of the seven *S.* Bovismorbificans 3114 SSH sequences targeted using PCR assays, four (3114-18, -A9, -B6 and –G4) were completely absent from UK veterinary Bovismorbificans isolates but present in >76% of Malawian bacteraemia isolates, and they were present in higher percentages of paediatric isolates (86-89%) than in adult isolates (59-68%) according to PCR (Table 4.7). These phage-related sequences have subsequently been mapped to two novel prophages ROD34 and ROD13, which are discussed in detail in chapter 5. These four SSH sequences and 3114-F10 (BLASTX match to a Gifsy-1 terminase large chain gp2) were completely absent from all genome sequenced *Salmonella* serovars. Interestingly the putative phosphotransferase system component (3114-11) was common in invasive serovars such as *S.* Typhi, Paratyphi A and B and Dublin (Table 4.9). Other African NTS isolates have been screened for the presence of *S.* Bovismorbificans SSH sequences and interestingly 3114 sequences were completely absent from all *S.* Typhimurium isolates and present in high numbers of *S.* Enteritidis isolates (Table 4.8.). This observation is in keeping with findings described in Chapter 5 which show high similarity between the genomes of *S.* Typhimurium and *S.* Bovismorbificans (Chapter 5). According to PCR assays SSH sequences 3114-18, -A9, -B6, -C12 and –G4 were completely absent from all NTS isolates tested.

4.5. Conclusions and Summary

- Using MLST, the presence of ST313 isolates in Uganda that pre-date the 2004 Malawi isolates has been confirmed; further it was demonstrated that Malawian paediatric isolates were also ST313 or ST313cc. This confirms the presence of a distinct ST in bacteraemia isolates from sub-Saharan Africa, as previously described by Kingsley and co-workers (2009).
- PCR assays have shown some evidence of geographical distribution among SSH sequences of *S.* Typhimurium and *S.* Bovismorbificans. *S.* Enteritidis D21685 sequences were absent from other serovars and rare amongst other *S.* Enteritidis isolates. *S.* Bovismorbificans phage sequence distribution implied a difference in bacteriophage carriage between African bacteraemia and UK veterinary isolates.
- The *S.* Typhi fimbrial operon *tcf* and an associated regulator *tinR* was found to be present in *S.* Heidelberg isolates and it was confirmed that *tcfA* is

expressed in all isolates tested carrying the gene. A number of *S.* Heidelberg isolates from different sources have been identified that are significantly more invasive in non-phagocytic cells than *S.* Typhimurium.

- The identification of colicin E1 on a number of *S.* Typhimurium plasmids may suggest the presence of ColE1-like plasmids in African *S.* Typhimurium isolates.

CHAPTER 5: GENETIC CHARACTERISATION OF *SALMONELLA ENTERICA* SEROVAR BOVISMORBIFICANS

5.1. Introduction

5.1.1. *S.* Bovismorbificans background

In 1894 Basenau described the properties of *Bacillus bovis-morbificans* isolated from a cow with metritis (Basenau, F. Archiv für Hygiene, 20:242, Feb 1894 from (Kross & Schiff, 1940). The bacterium was then identified across Europe, including the UK (Haines & Wilson, 1957), India (Hayes & Freeman, 1945) and South Africa (Henning & Greenfield, 1937). It was also identified as the cause of human infections in the United States (Angrist & Mollov, 1946; Seligmann *et al.*, 1943; Seligmann *et al.*, 1946). *S.* Bovismorbificans was found to be widely distributed in Australia and was first described there in 1940 by Stewart, who found it in sheep, pigs and horses (Stewart, 1940a; Stewart, 1940b). Atkinson described a number of human gastroenteritis outbreaks caused by *S.* Bovismorbificans (Atkinson *et al.*, 1944; Atkinson *et al.*, 1947). Outbreaks in Malaysia between 1973 and 1996 showed that *Salmonella* serovar Bovismorbificans accounted for 2-11% of Salmonellosis cases (Jegathesan, 1984; Lee *et al.*, 1998). However, earlier in the current decade, peaking in 2005, there was an increase in gastroenteritis in humans due to *S.* Bovismorbificans PT24 in pork products in Germany (Gilsdorf *et al.*, 2005), and there have been previous outbreaks associated with this serotype in Finland (Puohiniemi *et al.*, 1997), Denmark (Schiellerup *et al.*, 2001), Australia (PT32) (Stafford *et al.*, 2002) and Italy (Nastasi *et al.*, 1994). Although contaminated meat products are a potential source of such infections (Finley *et al.*, 2008; Magistrali *et al.*, 2008), alfalfa sprouts have also been implicated (Puohiniemi *et al.*, 1997). However, *Salmonella* surveys generally report Bovismorbificans as rare (<2%) amongst human isolates (Monno *et al.*, 2007). In Germany there was a marked increase in *S.* Bovismorbificans infection between November 2004 and March 2005. Enter-Net EuroSurveillance showed no increase in *S.* Bovismorbificans in other European countries during the same time period. Bovismorbificans was among the ten most commonly notified serovars between 2001 and 2003 in Germany

accounting for between 0.3 and 0.5% of all reported *Salmonella* infections (Gilsdorf *et al.*, 2005).

Although invasive NTS isolates from Africa are dominated by the serovars Typhimurium and Enteritidis (Gilks *et al.*, 1990; Gordon *et al.*, 2008; Kariuki *et al.*, 2006a; Kassa-Kelembho *et al.*, 2003; Sigauque *et al.*, 2009), a 7 year study of NTS bacteramias from Malawi showed that *S.* Bovismorbificans was the third commonest isolate, with 46 cases, which accounts for 1% of the total number of NTS isolates (Gordon *et al.*, 2008). To our knowledge this is the only report of bacteraemia caused by *S.* Bovismorbificans in the recent literature, although as mentioned above, sporadic outbreaks of *S.* Bovismorbificans gastroenteritis have been reported.

Multi-drug resistance is widespread in Malawi and other parts of Africa, where a large proportion of NTS are resistant to several commonly available antibiotics, including ampicillin, streptomycin, cotrimoxazole, chloramphenicol and tetracycline (Bachou *et al.*, 2006; Gordon *et al.*, 2008; Kariuki *et al.*, 2005). Chloramphenicol is the drug of choice in Malawian children; whereas *S.* Typhimurium remains susceptible the sudden appearance of chloramphenicol resistant *S.* Enteritidis is worrying. Interestingly, *S.* Enteritidis has become a more common cause of bacteraemia since it has developed resistance. Chloramphenicol resistance in Malawi went from 0% in January 1999 to over 80% in December of the same year (Graham, 2002). During this period the prevalence of *S.* Enteritidis as a blood culture isolate in children rose from 13 to 37% (Graham, 2002).

Salmonella serovars vary in their host range and pathogenicity, *S. enterica* subspecies I encompasses human adapted serovars including the human-specific typhoidal *Salmonella* serovars Typhi and Paratyphi, and generalists such as Typhimurium and Enteritidis, which are the most common NTS isolates around the world. Genetic variation between serovars generally accounts for ~10% of the genome. Variable regions include antigenic structures, regions implicated in host specificity and virulence determinants (Edwards *et al.*, 2002). Genomic variation is mediated through mobile genetic elements such as transposons, prophages, plasmids and insertion elements (Thomson *et al.*, 2004).

Since the publication of the first *S. enterica* genome (McClelland *et al.*, 2001), a large number of representative isolates from a number of *Salmonella* serovars have been genome sequenced. The NCBI Blast database currently logs 38 genome projects "in progress", 23 genomes are "in assembly" and 18 *Salmonella* enterica I

genomes have been completed. Furthermore the site also includes 19 *Salmonella* plasmids including the *S.* Bovismorbificans plasmid (http://www.ncbi.nlm.nih.gov/sites/entrez, 02/06/10).

Members of *Salmonella* subspecies I carry a low copy number (1-2 copies per chromosome) virulence plasmid of variable size (50-90kb), which is required for systemic infection. The *Salmonella* virulence plasmid is important for bacterial multiplication within the reticulo-endothelial system of warm blooded vertebrates. A 7.8 kb region *spv* (Salmonella plasmid virulence) is important to confer the virulence phenotype (Baird *et al.*, 1985; Williamson *et al.*, 1988). The *spv* region consists of five genes *spvR*, *spvA*, *spvB*, *spvC*, and *spvD*. The *spvR* gene encodes a regulatory protein which together with *rpoS* regulates the *spvABCD* genes (Fang *et al.*, 1992; Gulig *et al.*, 1993; Taira *et al.*, 1991). The *spv* genes are not required for colonization of the intestine or invasion of the epithelium. The *spv* region affects proliferation and cytotoxicity to macrophages. It was found that other ancestral lineages of *Salmonella* subspecies II, IIIa and VII as well as some isolates of subspecies IV carry *spv* genes on their chromosome (Boyd & Hartl, 1998).

Not every isolate of a plasmid-bearing serovar carries the virulence plasmid (Boyd & Hartl, 1998). It is found in a number of serovars of subspecies I, including Typhimurium, Enteritidis, Cholerasuis and Dublin. *S.* Typhi, Paratyphi A and Paratyphi B lack this plasmid; it is present in Paratyphi C but its contribution to virulence is unclear. Other loci of the plasmid such as *pef*, the fimbrial operon and *traT*, the conjugal transfer gene, as well as *rck* and *rsk* are involved in the infectious process. The role of the virulence plasmid in invasive disease is not proven conclusively, however *spv* is essential for Dublin induced salmonellosis in calves (Libby *et al.*, 1997). It was also observed that the plasmid is more often found in strains of *S.* Typhimurium and Enteridtidis isolated from blood and other extraintestinal sites then from strains isolated from faeces (Fierer *et al.*, 1992). The *Salmonella* Typhimurium virulence plasmid is self-transmissable whereas others, such as the virulence plasmid of *S.* Cholerasuis and *S.* Enteritidis, lack a complete *tra* operon. *Salmonella* virulence plasmids are very stable despite their low copy number (Tinge & Curtiss, 1990).

Lysogenic bacteriophages play a major role in the generation of genetic diversity within *S.enterica* I (Thomson *et al.*, 2004). Genes can be exchanged between bacteria through conjugation, transformation and transduction (Jain *et al.*, 2002).

Bacteriophages that mediate transduction can be isolated from the environment, including the intestine. Some bacteriophages can enter a lysogenic state and thereby contribute to the bacterial chromosome. Furthermore they can carry genes that are not essential for phage proliferation but may alter host pathogenicity (Campbell *et al.*, 1992). *S.* Typhimurium strains usually carry four to five prophages, whilst *S.* Typhi strains CT18 and Ty2 carry seven prophages (Thomson *et al.*, 2004).

Initially, suppression subtractive hybridization (SSH) was used in order to identify novel regions of the genome of an invasive paediatric bacteraemia *S.* Bovismorbificans isolate (3114), details of which are described in Chapter 3.

5.1.2 Aims of this chapter

In this chapter the genome sequence of *S.* Bovismorbificans ST142 3114, is characterised, general features of the *Salmonella* serovar Bovismorbificans genome are described, and the Bovismorbificans virulence plasmid is characterised. Using genomic sequencing techniques a number of invasive isolates of *S.* Bovismorbificans from sub-Saharan Africa were used, along with veterinary UK isolates, in a comparative genomic analysis in order to identify genomic characteristics associated with the invasive bacteraemia strains.

5.2. Results

5.2.1. Initial characterisation of *S.* Bovismorbificans strains

Antimicrobial susceptibility testing was carried out previously on these strains (Melita Gordon). A total of 11 antimicrobials were tested, and strains were resistant to either Sulphamethoxazol (RL), Cefuroxime (CXM) and Rifampicine (RD) or Cefuroxime (CXM) and Rifampicin (RD) only (Table 5.1.).

Table 5.1. shows an overview of clinical data and background information of Malawian paediatric bacteraemia and veterinary strains studied in this chapter.

Table 5.1. Overview of *S.* Bovismorbificans strains studied in this chapter. pSLT-like refers to the presence or absence of a virulence plasmid; age is given in years (Y) and/or months (M); strains were resistant to Sulphamethoxazole (RL); Cefuroxime (CXM) and Rifampicin (RD); ST refers to the MLST type as determined by Illumina/Solexa sequencing and sequencing of PCR amplicons. (ND=no data, Unk= unknown)

Strain	Isolation date	Isolation year	host	age	outcome	pSLT-like	Resistance profile
human isolates							
3114	20.01.1997	1997	child	ND	ND	pVIR+	RL_CXM_RD
3180	26.01.1997	1997	child	ND	ND	pVIR+	RL_CXM_RD
3476	18.02.1997	1997	child	ND	ND	pVIR+	RL_CXM_RD
3160	26.01.1997	1997	child	ND	ND	No	CXM,RD
3064	04.01.1997	1997	child	ND	ND	No	RL,CXM,RD
D993	03.04.1999	1999	child	ND	ND	pVIR+	CXM_RD
D1253	27.04.1999	1999	child	4M	ND	pVIR+	CXM_RD
D4451	09.01.2000	2000	child	1Y11M	ND	pVIR+	RL_CXM_RD
D4891	31.01.2000	2000	child	11M	ND	pVIR+	RL_CXM_RD
D19094	15.03.2003	2003	child	8Y	ND	No	RL,CXM,RD
A1104	11.06.1998	1998	adult	ND	3	pVIR+	CXM_RD
A1608	ND	Unk	adult	ND	ND	pVIR+	RL_CXM_RD
A1668	17.08.1998	1998	adult	ND	1	pVIR+	RL_CXM_RD
A16982	08.06.2002	2002	adult	ND	2	pVIR+	RL_CXM_RD
A31126	19.09.2004	2004	adult	23Y	1	pVIR+	RL_CXM_RD
A5893	31.08.1999	1999	adult	35Y	ND	pVIR+	RL_CXM_RD
A8737	03.05.2000	2000	adult	30Y	ND	pVIR+	RL_CXM_RD
A24091	12.07.2003	2003	adult	50Y	3	No	RL,CXM,RD
A22921	01.05.2003	2003	adult	36Y	2	No	CXM,RD
A22920	01.05.2003	2003	adult	45Y	3	No	CXM,RD
veterinary isolates							
499208	ND	Unk	alpaca	ND	ND	pVIR-	Unk
653308	ND	Unk	pig	ND	ND	pVIR-	Unk
276608	ND	Unk	pig	ND	ND	pVIR-	Unk
51892776	ND	Unk	pig	ND	ND	pVIR+	Unk

5.2.2. Multi Locus Sequence Typing (MLST)

MLST sequence types (ST) were determined for each strain from Illumina reads by locating the sequence in Artemis and entering it into the MLST search function of the *S. enterica* I MLST database (http://mlst.ucc.ie/mlst/dbs/Senterica). Ambiguous results were confirmed by carrying out MLST PCR assays using the PCR primers published on the MLST website (http://mlst.ucc.ie/mlst/dbs/Senterica/documents /primersEnterica_html) and Sanger sequencing the PCR products (Lark/Cogenics/). The Sequence Type (ST) was determined for 20 paediatric (D) and adult (A) MDR *S.* Bovismorbificans bacteraemia isolates from Malawi, isolated between 1997 and 2004 and 4 veterinary isolates from the UK isolated from pigs and alpaca. Further

details on each strain are summarized in Table 5.1. 14 human bacteremia and all 4 veterinary strains were either ST 142 or of the 142cc (ST142 clonal complex), which is the major *S.* Bovismorbificans lineage in the MLST database (www.mlst.net), whilst six human bacteraemia strains gave conflicting results, not matching any particular ST or matching STs not previously described for *S.* Bovismorbificans (Table 5.1). Strains depicted in red in table 5.2. were characterised through Illumina sequences only and were not followed up further due to their unusual or conflicting STs.

Figure 5.1. shows a phylogenetic representation of all the major lineages of *Salmonella enterica* subspecies 1, published on the MLST database (www.mlst.net) . Concatenated sequences were loaded into seaview (http://pbil.univ-lyon1.fr/software/seaview.html) in fasta format to construct the tree and figtree (http://tree.bio.ed.ac.uk/software/figtree/) was used to re-draw and label the final figure. The defining ST of the most important lineages of *S. enterica* subspecies I have been labelled at the tips.

Table 5.2. MLST data derived from Illumina sequences

strain	ST	aroC	dnaN	hemD	hisD	purE	sucA	thrA
human strains								
3064	~79, ~275	11	11	17	32	33	23	137
3160	???	72	167	-	~42 (1 SNP)	157	101	3
A22920	603	14	178	3	12	182	171	143
A22921	603 cc	14	178	3	-	182	171	143
A24091	446	15	126	101	88	8	19	18
D19094	~645, ~63	285	26	30	213	21	87	28
3114	142	2	59	23	64	38	61	12
3180	142	2	59	23	64	38	61	12*
3476	142	2	59	23	64*	38	61	12
D1253	142	2	59	23	64	38	61	12
D4451	142	2	59	23	64	38	61	12
A1104	142	2	59*	23	64	38	61	12*
A1608	142	2	59	23*	64	38	61	12
A1668	142	2	59	23	64	38	61	12
A16982	142	2	59	23	64	38	61	12
A31126	142	2	59	23	64	38	61	12
A5893	142	2	59	23	64*	38	61	12
A8737	142	2*	59	23	64	38	61	12
D4891	142 cc	2	59*	23	64	38	61	12
D993	142 cc	2	~59 (1 SNP)	23	64	38	61	12
veterinary strains								
51892776	142	2	59	23	64*	38	61	12
276608	142	2	59	23	64*	38*	61*	12*
499208	142	2	59	23	64	38	61	12
653308	142	2	59	23	64*	38	61	12

*loci confirmed by sequencing of PCR amplicons

cc = clonal complex, ~ = closest match with one SNP difference, - = sequence could not be confirmed

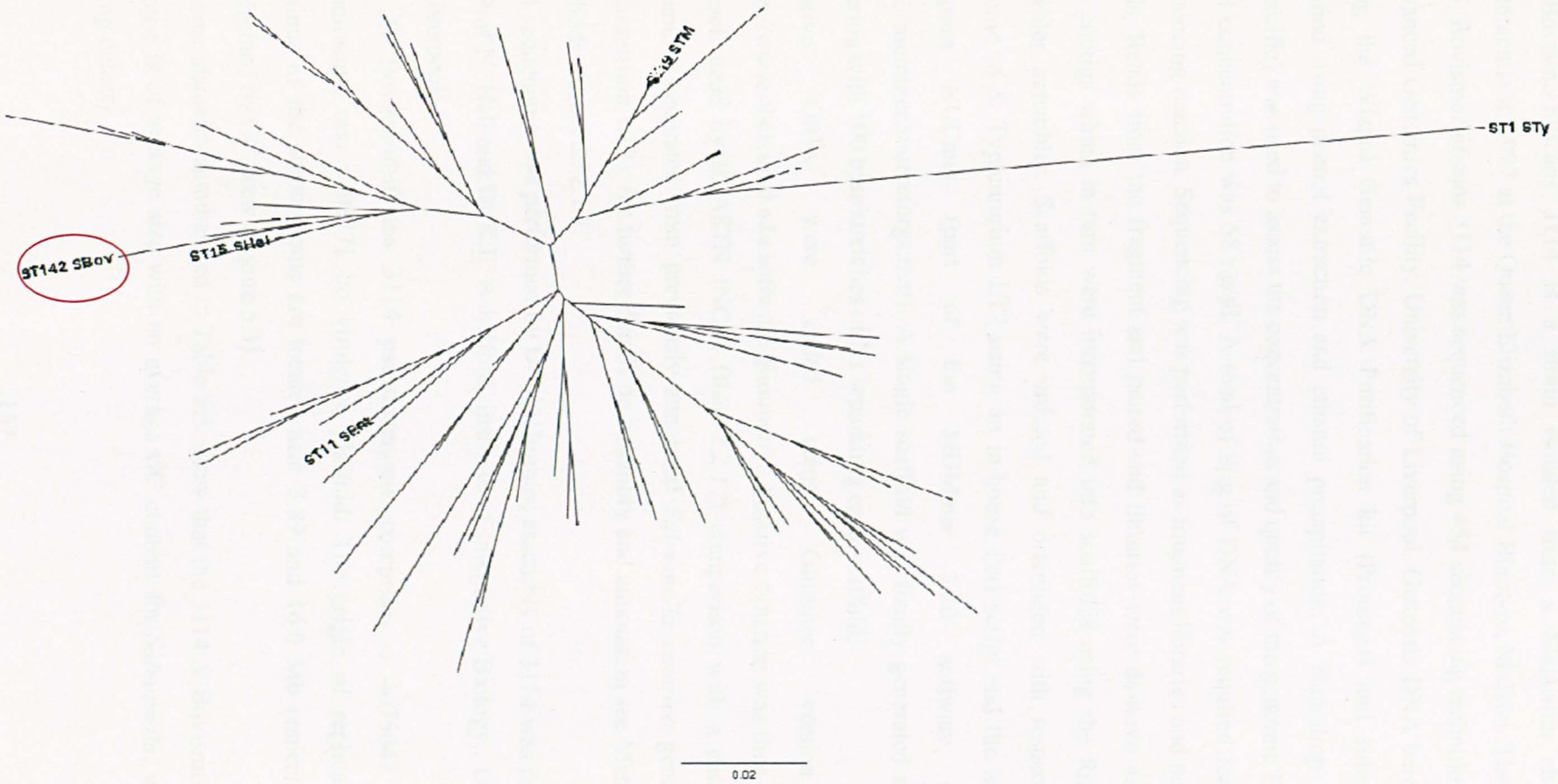


Figure 5.1. Concatenated MLST sequences of *S. enterica* subsp. I published on the MLST database (www.mlst.net). Tip labels show major lineages of the *Salmonella enterica* subsp. I serovar identified (ST11-*S.* Enteritidis, ST15-*S.* Heidelberg, ST19-*S.* Typhimurium and ST142-*S.* Bovismorbificans, circled in red) *S.* Typhi ST1 has been included as an outlier

5.2.3. The *S.* Bovismorbificans 3114 genome sequence (454 sequencing)

S. Bovismorbificans 3114 is a strain isolated from a Malawian child with bacteraemia in 1997 at the Queen Elizabeth Hospital Blantyre, Malawi. The genome of *S.* Bovismorbificans 3114 was sequenced using 454 sequencing technology at the Advanced Genomics Facility, University of Liverpool. Genomic DNA was isolated using the Wizard Genomic DNA Purification kit (Promega) and subsequently purified using phenol extraction and ethanol precipitation. A Nanodrop (Thermo Scientific) was used to assess the concentration and quality of the genomic DNA; the final concentration was 58 ng/µl. A total of 8µg of DNA was required for the full sequencing reaction. Sequencing was performed as fragment libraries and paired-end reads. Reads from the fragment and paired-end libraries were de-novo assembled into contigs which in turn were incorporated into scaffolds using the Roche 454 Newbler assembler. Scaffolds were ordered and orientated with respect to the genome of *S.* Typhimurium LT2 using an in-house Perl script and the alignment program NUCmer (part of the MUMmer 3.20 software package; http://mummer.sourceforge.net/). A single scaffold was finally generated from this ordering with 100 base stretches of Ns separating each scaffold.

Putative ORFs were called using Glimmer version 3.02 (http://www.cbcb.umd.edu/software/glimmer/). Putative function was then assigned to each gene by BLASTN (NCBI Blast 2.2.17) comparison with a database of sequences generated from previously annotated *Salmonella enterica* genomes (*S.* Typhimurium LT2), for further details on assembly and annotation see Material and Methods section 2.12.1.2.

(454 sequencing was performed by Dr M. Hughes; assembly of 3114 was performed by Prof N. Hall and Dr K.E. Ashelford, Institute of Integrative Biology,, University of Liverpool).

The *S.* Bovismorbificans 3114 pseudogenome comprises a 4675047 bp size chromosome and a 93771 bp virulence plasmid. The origin of replication and terminus of the chromosome are located near 3.89 and 16.0 Mb respectively, as determined by GC skew (Figure 5.3).

Genome statistics summarized in Table 5.3. show that the 3114 *S.* Bovismorbificans genome is of average size with an average GC content for *Salmonella*, with high coding density.

Table 5.3. Properties of the *S.* Bovismorbificans 3114 genome compared to other *S.* enterica genomes

	Bovismorbificans	Typhimurium	Typhi	Enteritidis P125109	Gallinarum
strain	3114	LT2	CT18	(PT4)	287/91
Size (bp)	4768822	4857432	4809037	4685846	4658697
percent G&C	52.02 (CDS)	52.22	52.09	52.17	52.2
No of CDS	4931	4451	4599	4318	4274
coding density	88%	86.80%	87.60%	85.50%	79.90%
average gene size	968	947	958	953	939
pseudogenes	113	25	204	113	309

*taken from (Chiu *et al.*, 2005; Thomson *et al.*, 2008)

Whole-genome comparison between *S.* Bovismorbificans and *S.* Typhimurium LT2 showed a high level of synteny and colinearity (Figure 5.3.), with no inversions. The Figure shows the location of important parts of the *Salmonella enterica* genome, such as *Salmonella* pathogenicity islands (SPIs) (blue arrows) and fimbrial operons (yellow arrows), as well as a number of *Salmonella* prophages such as Gifsy and Fels (black arrows).

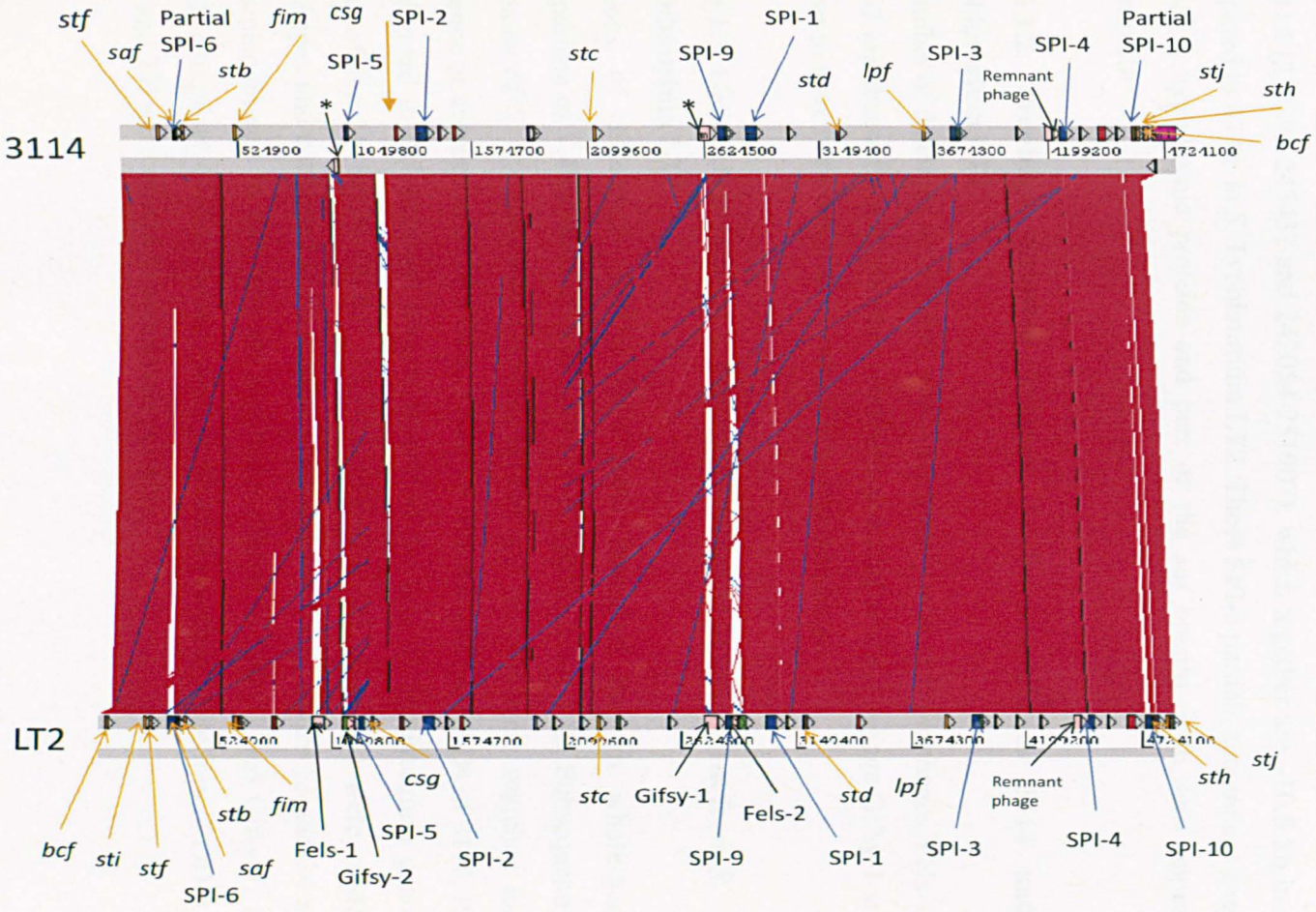


Figure 5.2 ACT comparison (http://www.sanger.ac.uk/Software/ACT) between *S.* Bovismorbificans 3114 and *S.* Typhimurium LT2 of amino acid matches between the complete six-frame translations (computed using TBLASTX) of the whole-genome sequences of *S.* Bovismorbificans and *S.* Typhimurium (LT2). Forward and reverse strands of DNA are shown for each genome (light grey horizontal bars). The red bars between the DNA lines represent individual TBLASTX matches, with inverted matches colored blue. The position of all the fimbrial operons marked in orange, the positions of *Salmonella* pathogenicity islands (SPI) are marked in blue, the position of prophages inserted into the genome are marked in pink. Analogous features are coloured the same. Prophages marked with a * in 3114 share some overlapping features with Gifsy-2, Gifsy-1 and Fels-2 respectively but have been found to be sufficiently different.

5.2.3.1. The *S.* Bovismorbificans 3114 chromosome

5.2.3.1.1. Gene sets common to both *S.* Bovismorbificans 3114 and *S.* Typhimurium

S. Bovismorbificans shares common features with Typhimurium and other *Salmonella* serovars such as a repertoire of 12 fimbrial operons and *Salmonella* Pathogenicity Islands (SPIs) 1-5 and 9, as well as a partial of SPI-6. SPI-3 in 3114 shows a small deletion of 4.4 kb, compared to *S.* Typhimurium LT2 which contains *sugR* and *rhuM*, as well as 2 additional CDS. There are two partials of SPI-6 present in 3114 (239898-243437 and 242054-251907), which together are ~10.6 kb in size, compared to 47 kb in *S.* Typhimurium LT2. These SPI-6 partials encompass a region encoding cytoplasmic proteins and part of the *saf* operon, with *safA* appearing incomplete.

5.2.3.1.2. Prophage variation between *S.* Bovimorbificans 3114 and *S.* Typhimurium LT2

A number of deletions are notable, which are mainly related to prophages. Fels-1 and Fels-2 are absent from the Bovismorbificans genome, while prophage Gifsy-1 and -2 appear to have been replaced by other Gifsy-like prophages.

5.2.3.1.3. Gene sets present in *S.* Bovismorbificans 3114 but absent from *S.* Typhimurium LT2

Regions of difference (RODs) have been identified through whole-genome comparison of *S.* Bovismorbificans 3114 and *S.* Typhimurium LT2. Subsequently the genomes of *S.* Typhimurium D23580, DT104 and SL1344 were searched for the presence or absence of 3114 RODs. Table 5.4 summarizes the RODs of 3114. RODs highlighted in red were not present in any of the four *S.* Typhimurium genomes. Eight of the 31 RODs (ROD7,-12, -13, -14, -17, -21, -30, -31) were related to prophages and contained at least some phage-related CDS, RODs 13 and 34 are of particular interest in that they are prophages related to Gifsy-2 and Gifsy-1. Eight RODs comprised CDS matching only hypothetical proteins, while four (ROD-9, -25, -28, and 33) were related to membrane proteins.

Table 5.4. Regions of difference (RODs) determined by ACT comparison of *S.* Bovismorbificans 3114 and *S.* Typhimurium LT2, further 3114 RODs were compared to *S.* Typhimurium D23580, DT104 and SL1344, RODs absent in all 3 are shaded in red; location refers to the location on *S.* Bovismorbificans 3114

ROD	label	CDS	location	D23580 match and CDS	DT104 match and CDS	SL1344 match and CDS	putative function
3	hypothetical	SBOV00591	63477 - 63975	Different	absent	absent	hypothetical
4	oxidoreductase	SBOV00901-SBOV00911	96912 - 98667	Absent	absent	absent	aldo/keto reductase, transcriptional regulator LysR family
5	hypothetical	SBOV01161-SBOV01181	124965 - 126985	Different STM MW01651	different SDT0165	partial match SL0160-SL0161	hypothetical, phosphotyrosine protein phosphatase, yacL pseudo
6	hypothetical	SBOV02291-SBOV02401	235478 - 243389	different STM_MW02641 – STM_MW03401	partial match SPI-6 remnant	partial match SPI6	hypothetical proteins, putative cytoplasmic protein, hypothetical, start of SPI-6
7	prophage-related	SBOV0269651-SBOV02701,	269973 - 270753	different BTP1	different prophage 1 (ST104)	partial match SL0319-SL0321	hypothetical, IS3 transposase orfA, pseudogene
8	Ab resistance-related	SBOV03191-SBOV03241	323995 - 331861	partial difference STM_MW04431	partial match SDT0412-SDT0416	partial match SL0367-SL0371	delta-aminolevulinic acid dehydratase, hypothetical, putative flagellin structural protein, hypotehtical, penicillin binding protein, protein SbmA
9	membrane-related	SBOV05111-SBOV05121	512750 - 515990	different STM_MW06211-STM_MW06241	different SDT0580-SDT0584	different SL0546-SL0548	putative membrane protein, glycosyl transferase
10	TolA	SBOV06831	693728 - 694165	partial difference STM_MW08041	partial match	partial match SL0729	protein TolA
11	hypothetical	SBOV06871	697721 - 698972	partial difference tRNAs	partial match tRNAs	partial match tRNAs	hypothetical
12	*integrase*	SBOV085411	859455 - 859832	partial difference STM_MW09591-	different SDT0917-SDT0919	different SL0885-SL0887	phage integrase

				STM_MW09611			
13	Gifsy-2 related prophage	SBOV09131-SBOV09521	932203 - 962362	Different BTP2	different Gifsy-2	different Gifsy2	excisionase, exodeoxyribonuclease 8, putative regulator, diguanylate cyclase, alkyl/aryl-sulfatase BDS1, 27 hypothetical, phage holin-analog protein, bacteriophage lysis protein, phage terminase large subunit, Gifsy-1 prophage head-tail preconnector gp4 , gp5, phage major capsid protein E, Gifsy-1 prophage VmtH
14	remnant phage	SBOV11471 - SBOV12121	1142532 - 1188945	Absent	absent	absent	SifA, 55 hypothetical, 2 exodeoxyribonuclease 8, replication P family, Eaa protein, HNH endonuclease, phage prohead protease, HK97 family, tail protein, spermidine/putrescine transporter, spermidine/putricine ABC transporter, putative peptidase T (aminotripeptidase)
15	hypothetical	SBOV15581-SBOV15611	1502134 - 1504843	Present STM_MW15481-STM_MW15491	match SDT1515-SDT1520	match SL1483-SL1484	hypothetical, pseudogenes
16	GTP-binding	SBOV18071-SBOV18101	1754017 - 1758742	Absent	partially absent SDT1749-SDT1751	partially absent SL1712-SL1713	GTP-binding protein EngD, 2 hypothetical, putative cytoplasmic protein
17	prophage remnant	SBOV18881-SBOV19071	1827565 - 1839850	partial match STM_MW18471-STM_MW18531	partial match SDT1818-SDT1843 end of prophage 3	partial match SL1785-SL1792	putative cytoplasmic protein, 13 hypotehtical, pseudo, putative O-acetyl transferase related, putative phage integrase, transposase Mutator family, putative

							cytoplasmic, putative fimbriae usher
18	periplasmic proteins	SBOV19051-SBOV19551	1872155 - 1875962	partial match STM_MW18801	partial match SDT1907-SDT1909		3 hypothetical, putative periplasmic protein, putative periplasmic phosphate binding protein,
19	O-antigen related	SBOV21431-SBOV21561	2044252 - 2057050	Different STM_MW21141-STM_MW			undecarpenyl-phospahte galactose ,phosphomannomutase, mannose-1-phospahte guanyltransferase/mannose-6-phosphate isomerase, glycosyl transferase group1, second mannosyl transferase, O-antigen polymerase, ramnosyl transferase, hypothetical, glycosyl transferase group 2 family protein, putative membrane protein, CDP-abequose synthase, lipolpolysaccharide biosynthesis protein RfbH
20	transcription	SBOV25581-SBOV25601	2474782 - 2478697	Absent	absent	absent	glycerate kinase, gluconate transporter, transcriptional regulator CdaR
34	prophage-related	SBOV26591-SBOV27131	2600887-2646661	Different BTP4	different prophage 5	partial match Gifsy-1 SLP272	novel phage
21	prophage-related	SBOV27691-SBOV27921	2708845 - 2725693	different phage insertion	partial match, mostly different SDT2737-SDT2770	partial match SL2712-SL2757 (region similar to CT18)	16 hypothetical, phage DNA binding protein, phage immunity repressor protein, repressor of phase 1 flagellin gene, flagellin pseudo, H inversion protein, putative integrase core domain protein, transposase subfamily,

22		SBOV29301-SBOV29351	2857799 - 2861624	partial match STM_MW28701	absent	partial match SL2886-SL2887	putative cytoplasmic protein - pseudo, serine/threonine-protein phosphatase 2, 4 hypothetical
23	CRISPR sequences	SBOV29651-SBOV29791	2888057 - 2898695	partial match STM_MW29001-STM_MW29071	mostly different SDT2931-SDT2938	partial match CRISPR region-SL2923	9 hypothetical, CRISP-associated protein Cas1, CRISPR-associated protein Cse3 family, CRISPR-associated protein Cas5, CRISPR-associated protein Cse4 family
24	fimbriae-related	SBOV29941-SBOV30001	2912924 - 2919863	absent/difference STM_MW29180	absent	absent	putative fimbrial subunit, outer membrane usher protein, chaperone protein PapD, fimbrial subunit, putative minor fimbrial subunit, fimbrial subunit, hypothetical
25	membrane - related	SBOV35431-SBOV35501	3430870 - 3444424	Match STM_MW34641-STM_MW34691	match SDT3454-SDT3459	match SL3440-SL3445	putative membrane protein, hypothetical, nitrite reductase NAD(P)H small subunit, nitrite transporter NirC, siroheme synthase, porin autotransporter,
26	hypothetical	SBOV37321-SBOV37341	3641260 - 3643330	Different	different	absent	hypothetical proteins
27	membrane-related	SBOV37701-SBOV37731	3677881 - 3680464	Absent	absent	absent	glutathione S-transferase C-domain, secretion protein HlyD family protein, inner membrane protein YiaW, hypothetical
28	membrane protein	SBOV37821	3686577 - 3691122	Match STM_MW36781-STM_MW36791	match SDT3670	match SL3655-SL3656	putative inner membrane protein
29	hypothetical	SBOV38131	3720876 - 3721893	Absent	absent	absent	hypothetical protein
30	prophage-	SBOV42691-	4190455 -	Different	different	partial match	DeoR family regulatory protein,

	related	SBOV42741	4196989	STM_MW41441-STM_MW41481 start of BTP6	SDT4181-SDT4186 (end of SPI)	phage remnant SLP443	hypothetical, putative membrane protein, tail assembly chaperone gp38, gp19
31	prophage-related	SBOV44051-SBOV44151	4334188 - 4341361	partial match STM_MW42621-STM_MW42701	partial match SDT4299-SDT4307	partial match SL4248-SL4258	putative phage protein, cytoplasmic protein, putative luxR family bacterial regulatory protein, AraC family regulatory protein, cytoplasmic protein, 2 hypothetical, putative periplasmic protein, TnpA pseudo, putative acid phosphatase
32	hypothetical	SBOV45841-SBOV5851	4505424 - 4507314	partial match STM_MW44441	partial match SDT4480-SDT4481 SPI-10 insertion in DT104	partial match SL4417-SL4429	hypothetical proteins
33	membrane proteins	SBOV46071-SBOV46211	4525053 - 4541118	partial match STM_MW44641-STM_MW44761	partial match SDT4501-SDT4514	partial match SL4448-SL4461	4 hypothetical protein, 2 putative membrane protein, unknown function, endoribonuclease SymE, putative type I restriction-modification system S subunit, M subunit, type I restriction enzyme EcoEI R protein, putative membrane protein, putative cytoplasmic protein

5.2.3.2. 3114 orthologous genes in *Salmonella enterica*, *E.coli* and *Yersinia*

A circular figure of the 3114 genome was constructed using DNA plotter (http://www.sanger.ac.uk/ resources/software/dnaplotter/) which is based on Artemis (Carver *et al.*, 2009). SSH sequences identified for *S.* Bovismorbificans 3114 were mapped onto the genome and are depicted in green on the outermost track.

Common *Salmonella* features such as SPIs and fimbrial operons (track 6) were identified in *S.* Bovismorbificans by overlaying tab files containing the location of such onto the 3114 genome in Artemis. Pseudogenes represented in red on track 5 were identified by scanning ORFs manually in Artemis and by comparing 3114 to *S.* Typhimurium D23580 in ACT. Reciprocal fasta searches were carried out and orthologs of the following serovars/species are shown on separate tracks, from the outside in: *S.* Typhimurium (LT2), *S.* Typhimurium (SL1344), *S.* Typhimurium (D23580), *S.* Enteritidis (SEN), *S.* Cholaeraesuis (Schol) , *S.* Paratyphi A (SpA), *S.* Paratyphi C (ParaC), *S.* Typhi (CT18), *S.* Gallinarum (SGAL) and *S.* Arizonae shown as red tracks, *E. coli* (M1655) and *E. coli* (Sakai) shown in blue, and *Yersinia* (YE), *Yersinia* (YPSTB) and *Yersinia* (YP91001) shown in green (Figure 5.3.).

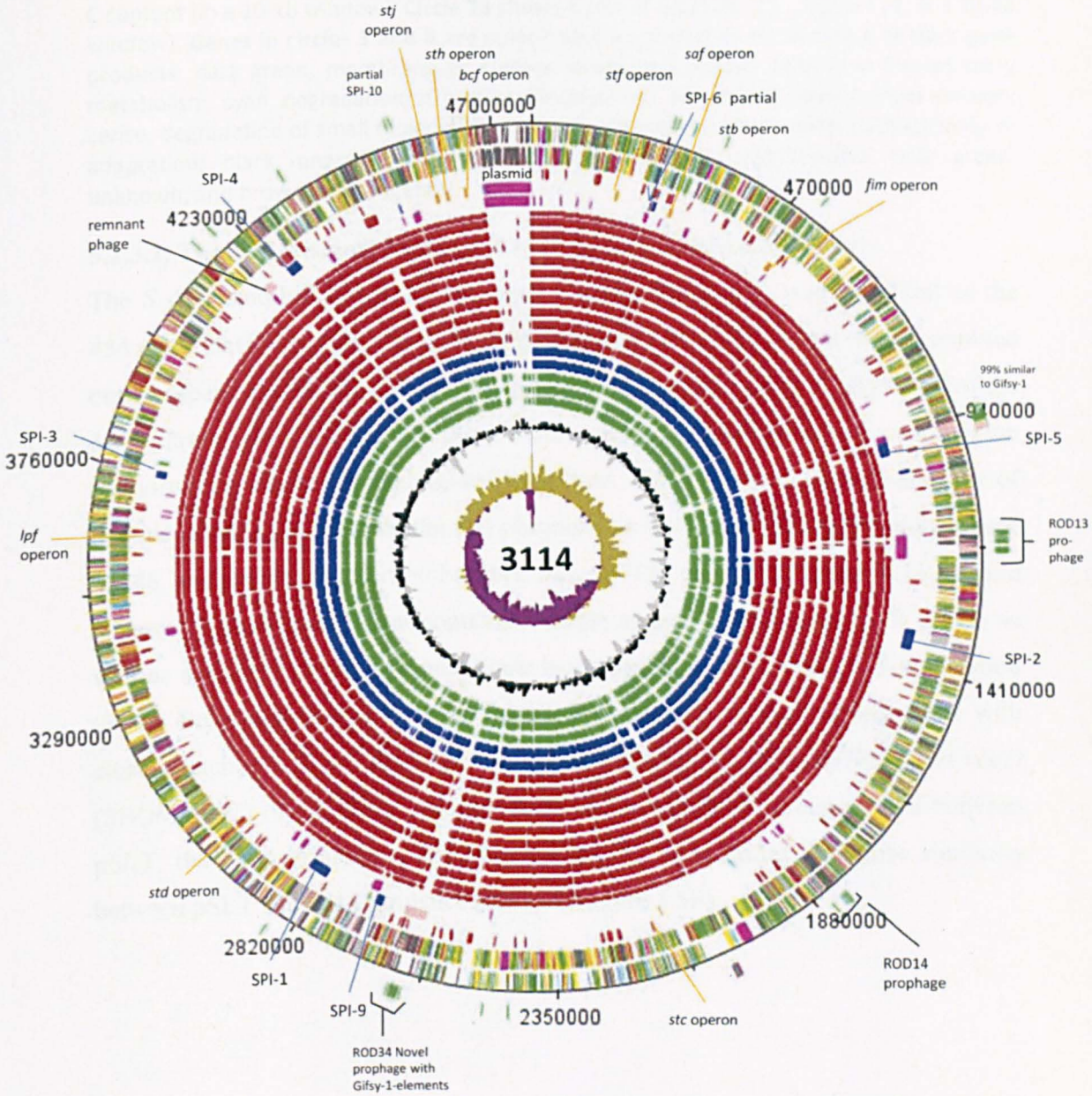


Figure 5.3. Representation of the *S.* Bovismorbificans chromosome. From the outside in, the outer **circle 1** marks the position of the 52 SSH sequences (green). **Circle 2** shows the size in base pairs. **Circles 3 and 4** show the position of CDS transcribed in a clockwise and anti-clockwise direction, respectively. **Circle 5** shows the position of pseudogenes (red). **Circle 6** shows Regions of Difference (RODs) common to several NTS, including pathogenicity islands and fimbrial operons, while **Circle 7** shows (RODs) in *S.* Bovismorbificans compared to *S.* Typhimurium (magenta). **Circles 8. to 21** show orthologous genes of *S.* Bovismorbificans (as determined by reciprocal FASTA analysis) in: *S.* Typhimurium (LT2), *S.* Typhimurium (SL1344), *S.* Typhimurium (D23580), *S.* Enteritidis (SEN), *S.* Cholaeraesuis (Schol) , *S.* Paratyphi A (SpA), *S.* Paratyphi C (ParaC), *S.* Typhi (CT18), *S.* Gallinarum (SGAL) and *S.* Arizonae in red, *E. coli* (M1655) and *E. coli* (Sakai) in blue and *Yersinia* (YE), *Yersinia* (YPSTB) and *Yersinia* (YP91001) in green. **Circle 22.** shows a plot of G +

C content (in a 10-kb window). **Circle 23** shows a plot of GC skew ([G _ C]/[G + C]; in a 10-kb window). Genes in **circles 3 and 4** are color-coded according to the function of their gene products: dark green, membrane or surface structures; yellow, central or intermediary metabolism; cyan, degradation of macromolecules; red, information transfer/cell division; cerise, degradation of small molecules; pale blue, regulators; salmon pink, pathogenicity or adaptation; black, energy metabolism; orange, conserved hypothetical; pale green, unknown; and brown, pseudogenes.

5.2.3.2. The *S.* Bovismorbificans 3114 virulence plasmid (pVIRBov)

The *S.* Bovismorbificans 3114 virulence plasmid (pVIRBov) was identified as the 454 reads that did not align to the *S.* Typhimurium LT2 chromosome were assembled onto a separate scaffold (scaffold 44) Figure 5.4 (A) shows an ACT alignment of the 3114 plasmid (top) and the LT2 pSLT virulence plasmid (bottom), the location of the *spv* virulence cassette is highlighted in yellow. The alignment shows a number of regions of variation between the two plasmids, as well as a 7463 bp deletion at base 38646, and a 6705 bp insertion between bases 67125 and 73830 in the 3114 plasmid (arrows). The deleted region contains a single strand binding proteins B (ssbB) as well as a number of putative membrane associated proteins on pSLT. The insertion on the Bovismorbificans virulence plasmid contains a number of sequences with closer matches to *E.coli* than *Salmonella*, including *vagC* (SBOV47881) and *vagD* (SBOV47891) (Table 5.5.). Another ACT alignment shows a comparsion between pSLT, the 3114 virulence plasmid and F-plasmid confirming the close similarity between pSLT and 3114 virulence plasmid (Figure 5.5B).

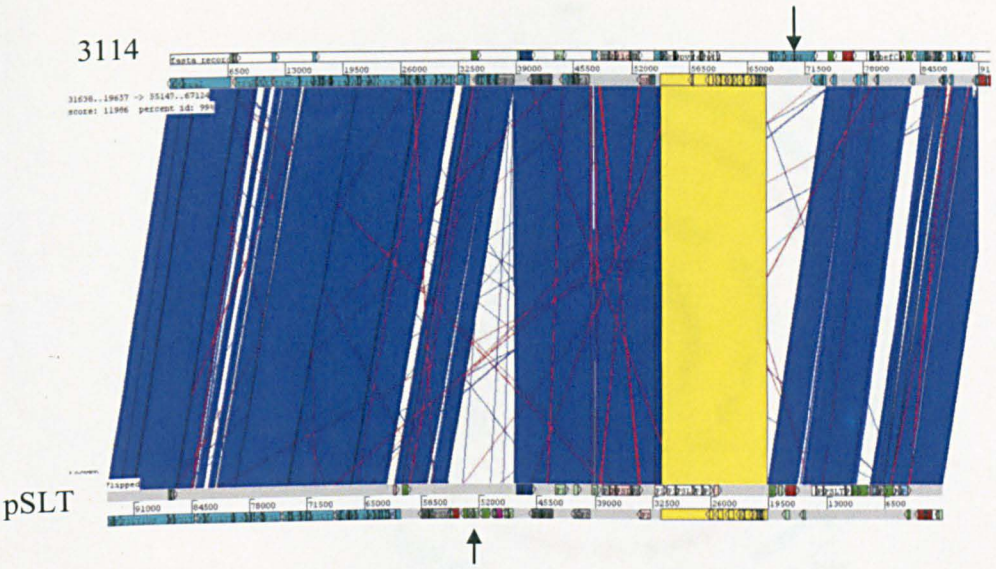


Figure 5.4 (A) ACT comparison of the 3114 plasmid to *S.* Typhimurium pSLT, highlighted in yellow is the *spv* operon. Forward and reverse strands of DNA are shown for each genome (light gray horizontal bars). The blue bars between the DNA lines represent individual TBLASTX matches, with inverted matches coloured red.

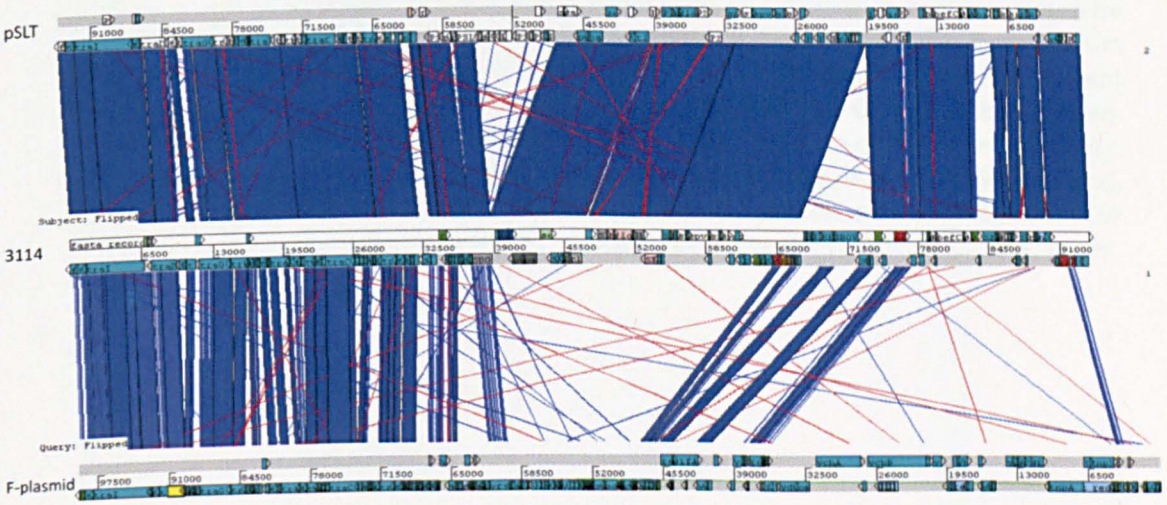


Figure 5.4. (B) ACT comparison of pSLT (top), the 3114 plasmid (middle) and F-plasmid (bottom) Forward and reverse strands of DNA are shown for each genome (light gray horizontal bars). The blue bars between the DNA lines represent individual TBLASTX matches, with inverted matches coloured red.

A circular figure of the 3114 virulence plasmid is shown in Figure 5.6. Orthologous genes of 3114 in LT2 pSLT are shown on track 4 in red. Important features such as the *pef* operon and *spv* have been labelled. Table 5.5. lists in detail the contents of the 3114 virulence plasmid

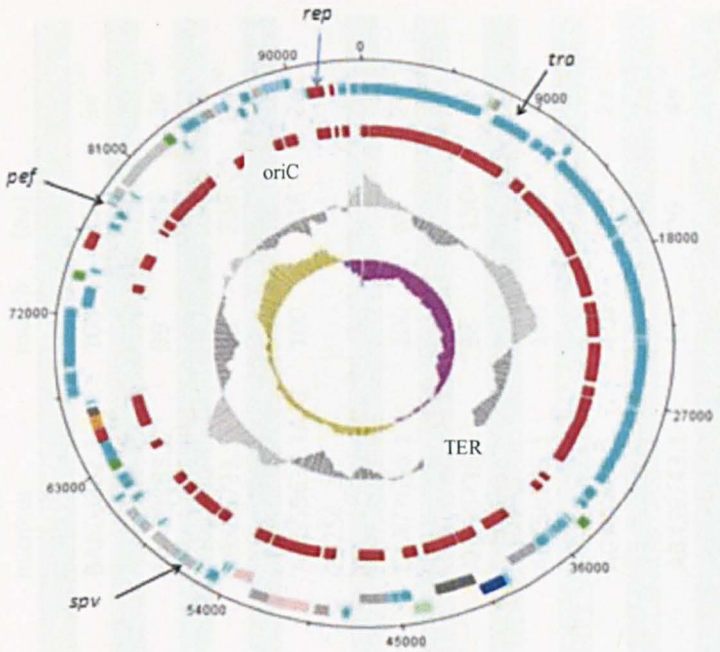


Figure 5.5. The virulence plasmid of *S.* Bovismorbificans 3114. From the outside: Circle 1 shows the size in basepairs, Circle 2 and 3 show CDSs in a clockwise and anti-clockwise direction, respectively. Circle 4 shows othologous genes of 3114 in pSLT of *S.* Typhimurium LT2 (red) as determined by reciprocal fasta analysis. Circle 4 shows a plot of G + C content (in a 10-kb window). Circle 5 shows a plot of GC skew ([G _ C]/[G + C]; in a 10-kb window). Genes in circles 2 and 3 are colour-coded according to the function of their gene products: dark green, membrane or surface structures; cyan, degradation of macromolecules; red, information transfer/cell division; pale blue, regulators; salmon pink, pathogenicity or adaptation; black, energy metabolism; orange, conserved hypothetical; pale green, unknown.

Table 5.5 Table summarizing the *S.* Bovismorbificans 3114 virulence plasmid, CDS numbers are continuous from the chromosome as the plasmid was assembled onto the last scaffold of the 3114 genome, however location (bp) refers to the separate 3114plasmid embl file, length refers to the length of the protein match in the NCBI database according to blastp matches, carried out through artemis. CDS highlighted in red are part of a 6705 bp insertion in 3114 compared to LT2, CDS highlighted in green are different compared to pSLT.

CDS	location	putative function	accession number	% match	length (aa)	E-value
SBOV47461	423-587	possible lysozyme	ACA51216.1	100	54	$5e^{-24}$
SBOV47471	761-1528	SpvA	BAA03382.1	100	255	$9e^{-149}$
SBOV47481	1491-1658	hypothetical protein SeD_B0021	ACH73578.1	100	55	$5e^{-23}$
SBOV47491	1710-3485	SpvB	BAA03383.1	99	589	$3e^{-124}$
SBOV47501	3607-3753	hypothetical protein SeD_B0023	ACH73503.1	97	47	$4e^{-18}$
SBOV47511	3766-4491	SpvC	ABX56731.1	98	238	$7e^{-97}$
SBOV47521	4545-4664	hypothetical protein SeD_B0025	ACH73517.1	100	39	$9e^{-14}$
SBOV47531	4753-5403	SpvD	AAG59632.1AF329254_1	100	216	$2e^{-122}$
SBOV47541	5530-5889	transposase	ACH73588.1	100	119	$5e^{-65}$
SBOV47551	6096-6362	SpvE	CAA77654.1	100	88	$2e^{-37}$
SBOV47561	6429-6845	YeeJ	ABX56735.1	100	107	$3e^{-56}$
SBOV47571	6869-7327	sch_V10	ABX56736.1	98	150	$3e^{-81}$
SBOV47591	8188-8676	Sch_V11 hypothetical inner membrane protein	ABX56738.1	100	161	$1e^{-83}$
SBOV47601	8701-9414	Sch_V12 putative inner membrane protein	ABX56739.1	100	162	$6e^{-91}$
SBOV47611	9423-10205	putative cyclic diguanylate phosphodiesterase (EAL) domain protein	ACH73531.1	100	237	$2e^{-133}$
SBOV47621	10240-10761	resolvase	ACY86464.1	100	260	$2e^{-149}$
SBOV47631	10758-11048	Sch_V15 putative cytoplasmic protein	ABX56742.1	100	173	$2e^{-96}$
SBOV47641	11050-11355	Sch_V16 putative cytoplasmic protein	ABX56743.1	100	96	$4e^{-48}$
SBOV47651	11357-11575	cytotoxic protein CcdB	ACH73561.1	100	101	$4e^{-52}$
SBOV47661	11797-11913	toxin addiction system: antidote	ACY86460.1	100	72	$2e^{-34}$
SBOV47671	12321-12551	hypothetical protein SeD_B0040	ACH73541.1	81	31	$6e^{-11}$
SBOV47681	1254812964	VagC [Escherichia coli]	ADL14115.1	94	76	$6e^{-35}$
SBOV47691	12996-13739	VagD [Escherichia coli]	ADL14117.1	94	136	$6e^{-71}$

SBOV47701	13977-15065	hypothetical protein Daro_2501	AAZ47235.1	38	235	$7e^{-38}$
SBOV47711	15067-17310	putative transcriptional repressor PifC [Escherichia coli]	EFK66797.1	89	362	0.0
SBOV47721	17434-18333	KAP family P-loop domain protein [Escherichia coli O157:H7	EDU89616.1	42	743	$8e^{-161}$
SBOV47731	18323-18613	hypothetical protein HMPREF9347_04314 [Escherichia coli	EFK66795.1	90	299	$5e^{-159}$
SBOV47741	18915-19403	yebA [Plasmid F]	BAA97908.1	96	96	$9e^{-47}$
SBOV47751	19537-19767	Sch_V19	ABX56746.1	96	162	$9e^{-87}$
SBOV47761	20677-21666	RepFIB replication protein A	AAS58893.1	100	329	0.0
SBOV47771	22207-22593	protein YgiW precursor	AAS58895.1	98	128	$4e^{-67}$
SBOV47781	22960-23271	hypothetical protein SPC_p024	ACN48797.1	90	103	$2e^{-46}$
SBOV47791	23300-23602	PefB	BBAB20526.1	100	100	$7e^{-52}$
SBOV47801	23728-23880	no significant match				
SBOV47811	23877-24395	PefA	BAB20527.1	89	172	$2e^{-81}$
SBOV47821	24423-24395	hypothetical protein SeI_A1194	EDZ14221.1	92	38	$5e^{-11}$
SBOV47831	24622-27030	PefC	BAB20528.1	99	802	0.0
SBOV47841	27023-27715	PefD	BAB20529.1	99	227	$3e^{-128}$
SBOV47851	27775-28290	putative outer membrane protein	ACN48802.1	94	171	$1e^{-89}$
SBOV47861	28439-28564	no significant match				
SBOV47871	28902-29810	putative outer membrane protein	EDZ14218.1	30	133	$1e^{-60}$
SBOV47881	29954-30067	hypothetical protein SeI_A1185	EDZ14191.1	77	36	$6e^{-09}$
SBOV47891	30116-30766	Dlp(SrgA)	BAB20573.1	97	215	$1e^{-120}$
SBOV47901	30836-31048	PefI	BAB20571.1	96	57	$2e^{-24}$
SBOV47911	31033-31368	putative bacterial regulatory protein	ACN48805.1	100	110	$1e^{-57}$
CDS	31365-31526	hypothetical protein SentesT_2802	ZP_03385880.1	96	30	$1e^{-06}$
SBOV47921	31847-32008	no significant match				
SBOV47931	31929-32165	no significant match				
SBOV47941	32282-32749	putative outer membrane protein –sequence gap	AAO71980.1	100	15	1.6
SBOV47951	32742-32861	no significant match				
SBOV47961	32909-33466	resistance to complement killing (rck)	ACY86444.1	98	185	$5e^{-101}$
SBOV47971	33520-33639	hypothetical protein SeI_A1180	EDZ14214.1	94	39	$4e^{-13}$

SBOV47981	33603-34472	SrgC	BAB20576.1	100	246	$2e^{-144}$
SBOV47991	34555-34992	hypothetical protein SeI_A1177	EDZ14202.1	97	43	$4e^{-17}$
SBOV48001	35526-35741	hypothetical protein SeD_B0066	ACH73530.1	88	71	$6e^{-23}$
SBOV48011	35753-36622	DNA replication	AAL23448.1	98	289	$2e^{-166}$
CDS	36940-37185	DNA replication	AAL23445.1	100	81	$7e^{-39}$
SBOV48021	37385-37825	putative phospholipase D	AAL23444.1	97	147	$1e^{-77}$
SBOV48031	37991-38443	dsba oxidoreductase	ACH73544.1	98	150	$1e^{-81}$
SBOV48041	38560-39123	conjugal transfer fertility inhibition protein FinO	YP_209284.1	100	187	$2e^{-103}$
SBOV48051	39178-39918	type-F conjugative transfer system pilin acetylase TraX	EDZ14222.1	99	246	$7e^{-138}$
SBOV48061	39938-45193	conjugative transfer: oriT nicking-unwinding	AAL23509.1	97	1755	0.0
SBOV48071	45272-45502	putative cytoplasmic protein	AAL23544.1	98	76	$4e^{-36}$
SBOV48081	45502-45900	homologue of mvpA, Shigella flexneri	AAL23543.1	99	132	$3e^{-72}$
SBOV48091	45909-48116	TraD	ABX56760.1	99	735	0.0
SBOV48101	48396-49007	hypothetical protein [Escherichia coli]	BAF33992.1	82	203	$2e^{-98}$
SBOV48111	49127-49858	TraT	ABX56761.1	98	243	$5e^{-136}$
SBOV48121	49957-50268	conserved hypothetical protein	CAQ87996.1	31	107	$5e^{-07}$
SBOV48131	50234-53089	conjugative transfer: assembly and aggregate stability (TraG)	AAL23504.1	93	937	0.0
SBOV48141	53086-54717	traH pilus assembly protein (traH)	CBA11441.1	99	377	0.0
SBOV48161	54623-54799	no significant match				
SBOV48171	54707-55255	type-F conjugative transfer system pilin assembly thiol-disulfide isomerase TrbB	EDZ14131.1	98	182	$1e^{-101}$
SBOV48181	55242-55535	type-F conjugative transfer system pilin chaperone TraQ	EDZ14160.1	98	97	$2e^{-47}$
SBOV48191	55649-55780	type-F conjugative transfer system pilin assembly protein TraF	EDZ14147.1	98	250	$1e^{-143}$
SBOV48201	55749-56501	hypothetical protein SeI_A1802	EDZ14161.1	93	43	$8e^{-15}$
SBOV48211	56510-56725	conjugative transfer protein	ACY86523.1	100	71	$2e^{-32}$
SBOV48221	56736-58562	type-F conjugative transfer system mating-pair stabilization protein TraN	EDZ14156.1	97	608	0.0
SBOV48231	58559-58978	HNH endonuclease	EDZ14151.1	97	139	$2e^{-75}$
SBOV48241	59015-59653	conjugative transfer assembly protein	AAY88111.1	94	212	$9e^{-113}$
SBOV48251	59650-59946	conjugative transfer protein	AAY88112.1	97	98	$2e^{-43}$

SBOV48261	59965-60957	conjugative transfer assembly protein	AAY88113.1	99	330	0.0
SBOV48271	60954-61484	conjugative transfer: assembly	AAL23494.1	98	176	$4e^{-97}$
SBOV48281	61583-61888	type IV secretion-like conjugative transfer system protein TrbI	EDZ14159.1	100	101	$4e^{-52}$
SBOV48291	61944-64592	type IV secretion-like conjugative transfer system protein TraC	EDZ14149.1	99	882	0.0
SBOV48301	64608-64991	putative multidrug efflux pump, Major facilitator superfamily	EDZ14132.1	96	87	$8e^{-40}$
SBOV48311	64988-65455	conjugative transfer protein	EDZ14125.1	99	155	$5e^{-82}$
SBOV48321	65448-65669	type IV secretion-like conjugative transfer system protein TraR	EDZ14145.1	100	73	$5e^{-36}$
SBOV48331	65804-66319	type IV conjugative transfer system protein TraV	EDZ14158.1	91	171	$7e^{-84}$
SBOV48341	66316-66552	conjugative transfer protein	EDZ14155.1	94	78	$2e^{-35}$
SBOV48351	66539-67075	conjugative transfer protein	EDZ14139.1	80	198	$5e^{-85}$
SBOV48361	67089-68486	protein TraB	EDZ14134.1	97	465	0.0
SBOV48371	68486-69226	type-F conjugative transfer system secretin TraK	EDZ14130.1	98	246	$1e^{-136}$
SBOV48381	69213-69779	type IV secretion-like conjugative transfer system protein TraE	EDZ14150.1	98	188	$6e^{-105}$
SBOV48391	69801-70112	type IV conjugative transfer system protein TraL	EDZ14133.1	100	103	$2e^{-54}$
SBOV48401	70127-70489	type IV secretion-like conjugative transfer system pilin TraA	EDZ14144.1	98	120	$2e^{-58}$
SBOV48411	70895-71584	protein TraJ	ACD54539.1	54	223	$3e^{-64}$
SBOV48421	71766-72146	mating signal protein (TraM)	EDZ14142.1	100	126	$6e^{-67}$
SBOV48431	72507-73031	transglycolylase	EDZ14137.1	98	156	2e-86
SBOV48441	73074-73190	hypothetical protein SeI_A1775	EDZ14141.1	97	38	$2e^{-13}$
SBOV48451	73343-73513	conserved hypothetical protein	ACF57125.1	70	41	$5e^{-09}$
SBOV48461	73555-73788	yubP protein [Escherichia coli B088]	EFE64087.1	75	73	$3e^{-27}$
SBOV48471	73892-74161	conserved hypothetical protein [Escherichia coli MS 146-1]	EFK88260.1	88	89	$3e^{-39}$
SBOV48481	74264-74383	hypothetical protein SeI_A1771	EDZ14136.1	100	39	$2e^{-15}$
SBOV48491	74508-74636	hypothetical protein SeI_A1770	EDZ14163.1	97	42	$9e^{-15}$
SBOV48501	74671-75237	PsiA (plasmid sos inhibition)	EDZ14128.1	100	188	$8e^{-107}$
SBOV48511	75264-75428	plasmid SOS inhibitor protein	AAY88122.1	87	62	$4e^{-22}$
SBOV48521	75471-77222	parB-like partition protein	EDZ14154.1	99	577	0.0
SBOV48531	77493-77915	mutagenesis by UV and mutagens; related to umuDC operon	AAL23541.1	98	140	$2e^{-75}$
SBOV48541	77915-79193	mutagenesis by UV and mutagens; related to umuDC operon	AAL23540.1	98	425	0.0
SBOV48551	79275-80249	plasmid partition protein B	AAL23465.1	99	324	0.0

SBOV48561	80249-81454	plasmid partition protein A	AAL23464.1	100	401	0.0
SBOV48571	81869-82810	conserved hypothetical protein [Escherichia coli MS 146-1]	EFK88260.1	91	89	$3e^{-39}$
SBOV48581	82807-83412	hypothetical protein SeI_A1771	EDZ14136.1	100	39	$2e^{-15}$
SBOV48591	83469-83804	hypothetical protein SeI_A1770	EDZ14163.1	97	42	$9e^{-15}$
SBOV48601	83857-83976	YacC	EDZ09452.1	100	38	2e-14
SBOV48611	83988-84497	YacC	ABX56770.1	98	169	8e-93
SBOV48621	84490-85605	TlpA	EDZ13892.1	99	371	0.0
SBOV48631	86044-86454	putative cytoplasmic protein	AAL23537.1	100	52	8e-23
SBOV48641	87109-87849	putative carbonic anhydrase	AAL23536.1	100	246	1e-141
SBOV48651	88056-88616	RlgA	ACH73522.1	100	186	1e-101
SBOV48671	88600-90264	integrase, catalytic region	ACH73575.1	99	554	0.0
SBOV48681	90254-91222	putative phosphoribulokinase / uridine kinase family	AAL23533.1	100	322	0.0
SBOV48691	91411-92448	putative integrase protein	AAL23461.1	99	345	0.0
SBOV48701	92656-92787	hypothetical protein SeI_A2420	EDZ13936.1	100	43	7e-17
SBOV48711	92953-93093	no significant match				
SBOV48721	93056-93646	SpvR	BAA03385.1	100	194	9e-109

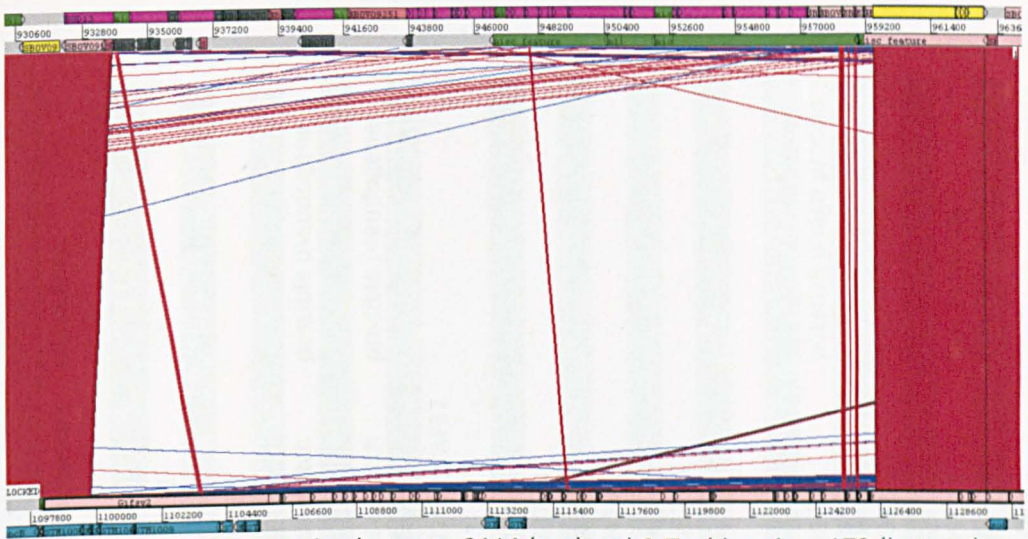
5.2.4. The *S.* Bovismorbificans novel prophages

The ROD labelled ROD13, matches the location of the Gifsy-2 prophage in *S.* Typhimurium LT2. SSH sequences 3114-D12, -D03,-F06, -F10, H07 and -F04 were found to cluster to this region (Figure 5.6 (B)). Gifsy-2 is 45486bp in size in LT2. In *S.* Bovismorbificans, a 15327bp conserved region matches the C-terminal end of Gifsy-2, whereas 30159 bp of ROD13 differs between LT2 and 3114. The contents of ROD13 are summarized in Table 5.6. ROD13 appears to be a variable region within serovar Typhimurium as it also differed in *S.* Typhimurium D23850, where it has been lablled BTP2, as well as in DT104 and SL1344 (Table 5.4.).

A large insertion, 46413 bp in size, downstream of SPI-5 has been labelled ROD14 (Figure 5.8 (A)). Its contents are summarized in Table 5.8. The region appears to be prophage-related, carrying putative phage head and tail protein as well as a number of genes matching hypothetical proteins and a *sifA* gene (SBOV11471). SSH sequences, 3114-D12, -6, -C02, -G04, -B06, -G04, -G02, -2, -H10, -D07, -A01, -B06 cluster to ROD14 (Figure 5.7 (B)). ROD14 was also absent from *S.* Typhimurium D23580, DT104 and SL1344.

A third prophage-related region of difference is ROD34, which is located at the insertion site of Gifsy-1 in LT2. Gifsy-1 in LT2 is 47840 bp in size whilst ROD34 is 45774 bp in size. ROD34 carries Gifsy-1 like elements at the beginning and end, as well as one Fels-1 like element (for details see Table 5.8). A total of nine SSH sequences cluster to ROD34 (3114-15, -C04, -19, -18, -05, -A09, -C10, -G11, -D10, -H01 and H06) (Figure 5.8(B)). Comparsions with *S.* Typhimurium D23580, DT104 and SL1344 show different phages present in this location in each strain, the ROD34 phage partially matches Gifsy-2 SLP272 of SL1344 (Table 5.4).

3114



LT2

Figure 5.6 (A) ACT comparsion between 3114 (top) and *S.* Typhimurium LT2 (bottom) showing the ROD13 prophage

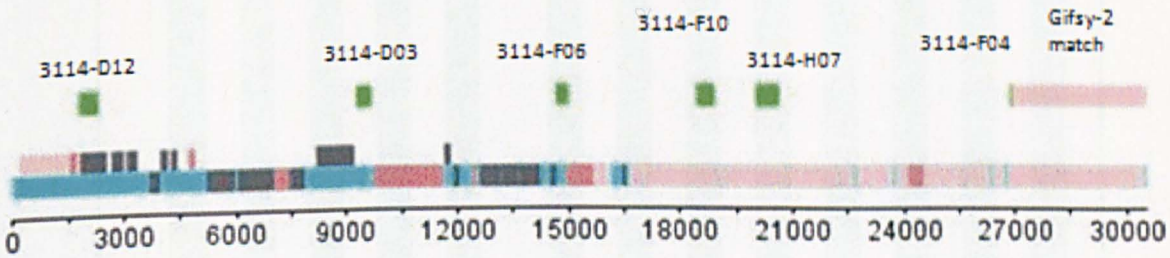


Figure 5.6(B) shows the position of 3114-SSH sequences matching the novel ROD13 prophage

Table 5.6. ROD13 prophage, underlined are CDS matching SSH sequences

CDS	location	putative function	Accession number match	comments
SBOV09131	932230-933522	putative phage integrase	ACH75199.1	
SBOV09141	933567-933815	prophage excisionase (Gifsy-2)	AAL19940.1	
SBOV09151	933852-934538	exonuclease IIIV	EDX45199.1	SSH 3114-D12
SBOV09161	934676-935011	LygB	EDX44991.1	
SBOV09171	935086-935370	conserved hypothetical protein, low match to Gifsy-1 protein	EDZ19035.1	
SBOV09181	935970-936200	hypothetical protein	no close match	possible pseudogene
CDS	935678-936004	type 2 secretion F domain	low match (33%)	
SBOV09191	936306-936461	hypothetical protein	no close match	possible pseudogene
SBOV09201	936758-936955	predicted phage regulatory protein	CAR37070.1	
CDS	937255-937509	Cro repressor-like protein	AAK57707.1|AF378725_1	
SBOV09211	937496-937990	predicted phage protein	CAR37072.1	
SBOV09221	938037-939044	predicted phage protein	CAR37073.1	
SBOV09231	939076-939498	replication P family protein	EDX46975.1	
CDS	939511-939906	LygF	EDZ29410.1	
SBOV09241	940193-941254	diguanylate cyclase	EDZ06275.1	
SBOV09251	941732-943699	alkyl/aryl-sulfatase BDS1	ACH75579.1	
SBOV09261	943717-943875	conserved hypothetical protein	EDZ06161.1	
SBOV09271	943940-944092	no significant match		
SBOV09281	944373-944591	putative phage protein	CAR32719.1	
SBOV09291	944655-945254	IrsA	EDX48904.1	putative phage protein
SBOV09301	945251-945445	conserved hypothetical protein	EDX48949.1	
SBOV09311	945427-945723	conserved hypothetical protein	EDZ29331.1	
SBOV09321	945720-946274	putative bacteriophage protein	ACY87930.1	
SBOV09331	946551-946739	hypothetical protein SeW_A1423	EDZ29532.1	SSH 3114-F06
SBOV09341	946943-947245	phage holing analog protein	EDZ15099.1	SSH 3114-F06
SBOV09351	947274-947711	phage lysozyme	EDZ15095.1	

SBOV09361	947696-948172	bacteriophage lysis protein	ACF64677.1	
SBOV09371	948508-948627	no significant match		possible pseudogene
SBOV09381	948641-949186	to Gifsy-1 prophage DNA packaging protein	EDX48365.1	low match (78%)
SBOV09391	949158-951092	phage-terminase large subunit	ACY87942.1	SSH 3114-F10
SBOV09401	951076-951279	Gifsy-1 head to tail joining protein	AAX65122.1	
SBOV09411	951276-952862	Gifsy-1 prophage head-tail preconnector gp4 [Phage Gifsy-1]	AAX65123.1	SSH 3114-H07
SBOV09421	95852-954366	gifsy-1 prophage head-tail preconnector gp5	EDZ33645.1	

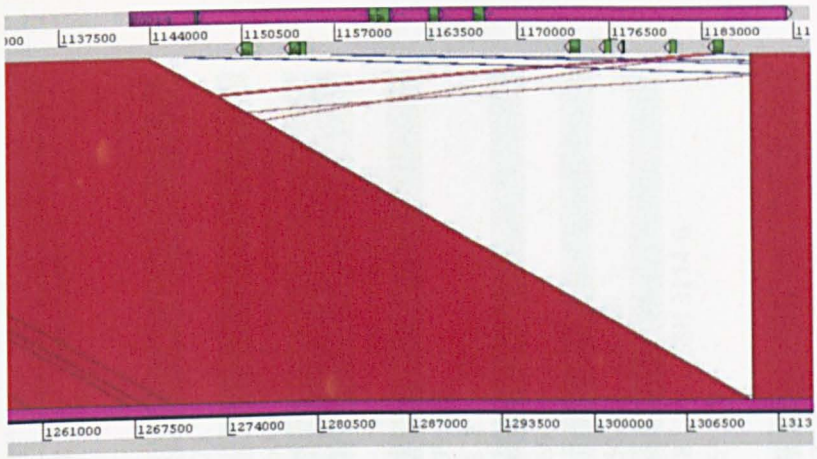


Figure 5.7 (A) ACT screenshot of the 3114 genome (top) and *S.* Typhimurium LT2 (bottom) showing the location of the ROD14 prophage

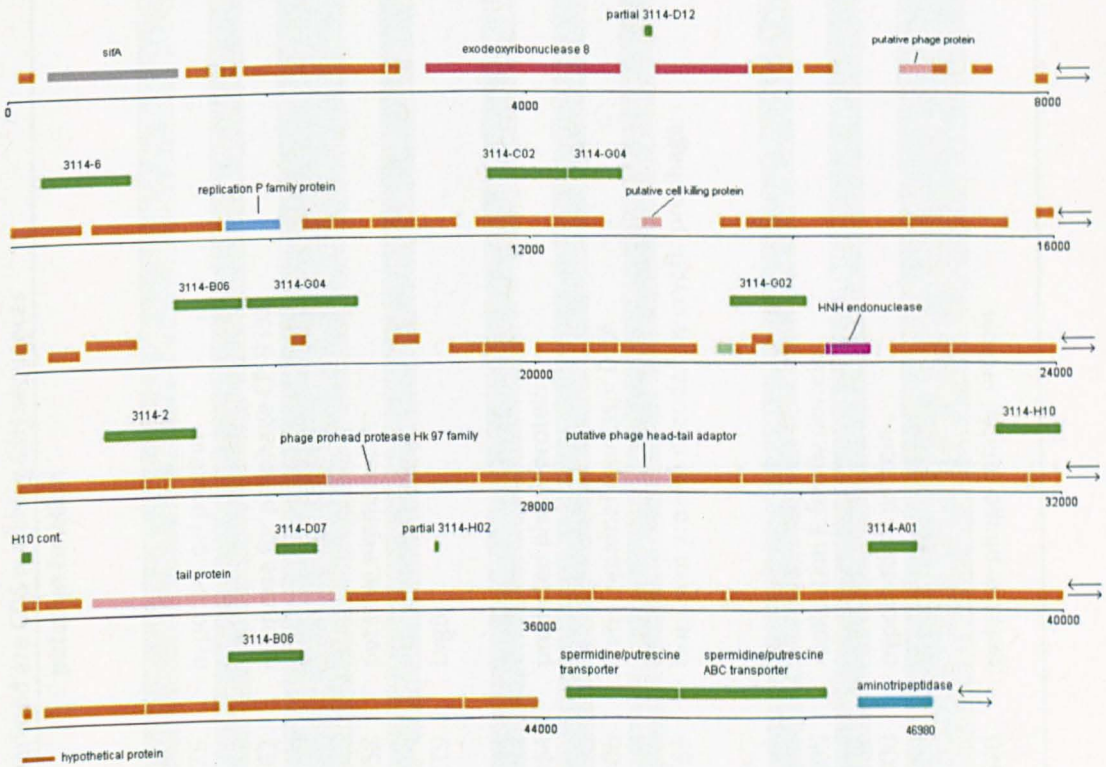


Figure 5.7 (B) Figure 5.7 (B) shows the position of 3114-SSH sequences matching the novel ROD14 prophage

Table 5.7. ROD14 prophage, underlined are CDS matching SSH sequences

CDS	location	putative function	Accession number match	comments
SBOV11471	1142530-1143540	SifA	EDZ02055.1	
SBOV11481	1143599-1143775	hypothetical protein	ACN46625.1	
SBOV11491	1143869-1143991	hypothetical protein		no close match
SBOV11501	1144044-1145162	integrase for prophage CP-933N	EDX28460.1	*E. coli* closest match
SBOV11501	1145137-1145259	putative excisionase for prophage CP-933N	AAG55870.1| AE005322_4	
SBOV11521	1145462-1146958	putative exonuclease		low match (52%)
SBOV11531	1147224-1147934	exodeoxynuclease 8	ACH74194.1	low match (70%)
SBOV11541	1147962-1148279	LygB	AAL27299.1|AF370716_2	
SBOV11551	1148364-1148585	putative phage encoded cell division inhibitor protein	CAR32960.1	
CDS	1149088-1149354	predicted phage protein	CAR37069.1	
CDS	149351-1149458	hypothetical protein		*E.coli* low match
SBOV11561	1149654-1149806	transcriptional repressor DicA	ACT71556.1	*E.coli*
SBOV11571	1150133-1150360	DNA-binding transcriptional regulator Dic	EFE63956.1	
SBOV11581	1150344-1150769	Unknown protein encoded by cryptic prophage	AAN79880.1| AE016759_154	SSH 3114-6
SBOV11591	1150841-1151851	conserved hypothetical protein	EEJ45413.1	low match, SSH 3114-6
SBOV11601	1151883-1152305	replication P family protein	ACH76172.1	
SBOV11611	1152483-1152713	LygF	low match	
SBOV11621	1152710-1153000	hypothetical protein	ABV18768.1	
SBOV11631	1152997-1153287	hypothetical protein	low match E.coli	
SBOV11641	1153284-1153670	putative phage-related protein	ABB65631.1	low match *Shigella*

SBOV11651	1153816-1154388	Eaa protein		low match (55%), SSH 3114-C02
SBOV11661	1154388-1154783	hypothetical protein ECDG_01107	EFF07576.1	SSH 3114-G04
CDS	1155071-1155226	putative cell killing protein encoded within cryptic prophage CP-933M (Hok/gef family protein)	ABE06954.1	
SBOV11671	1155671-1155826	no significant match		
CDS	1155875-1156075	conserved hypothetical protein	low match Salmonella	
SBOV11681	1156077-1157120	hypothetical protein	low match	
SBOV11691	1157132-1157506	putative endonuclease	low match Shigella (70%)	
SBOV11701	1157496-1157882	phage antitermination Q type 1 family	EDU64777.1	
SBOV11711	1158100-1158321	LVIVD repeat protein	very low match (48%)	
SBOV11721	1158472-1158720	putative bacteriophage protein	EDU65577.1	low match (78%)
SBOV11731	1158760-1159152	hypothetical protein	low match (45%)	
SBOV11741	1160347-1160466	no significant match		SSH 3114-G03
SBOV11751	1161141-1161344	similar to inter-alpha (globulin) inhibitor H4 (plasma Kallikrein-sensitive glycoprotein) [Taeniopygia guttata]	GENE ID: 100230771 LOC100230771	very low match
SBOV11762	1161574-1161999	conserved hypothetical protein	EFK43350.1	*E.coli*
SBOV11771	1161996-1162148	conserved hypothetical protein	EFK43068.1	*E.coli*
SBOV11781	1162238-1162630	phage holing protein	EFF07725.1	low match (71%)
SBOV11791	1162620-1162901	putative prophage membrane protein	CAR32973.1	
SBOV11801	1162898-1163476	conserved phage protein	CAR32974.1	possible lysozyme
CDS	1163632-1163745	hypothetical protein	EDV60666.1	various *E.coli*
SBOV11811	1163774-1163929	hypothetical phage protein	BAG77594.1	various *E.coli*, SSH

				3114-G02
SBOV11821	1163895-1164053	no significant match		SSH 3114-G02
SBOV11831	1164144-1164491	no significant match		match to drosophila gene (24%) SSH 3114-G02
SBOV11841	1164460-1164810	Hnhc	EDV60660.1	various *E.coli*
SBOV11851	1164958-1165440	phage terminase, small subunit, P27 family	EFK15390.1	
SBOV11861	1165440-1167197	putative phage terminase, large subunit	EDX38951.1	*E.coli*, SSH 3114-02
SBOV11871	1167209-1167391	conserved hypothetical protein	EDX39038.1	*E.coli*, SSH 3114-02
SBOV11881	1167391-1168632	phage portal protein, HK97 family	EDX38976.1	SSH 3114-02
SBOV11891	1168610-1169260	putative prohead protease	CAC83557.1	Bacteriophage P27
SBOV11901	1169275-1169820	putative major capsid protein	CAC83558.1	Bacteriophage P27
SBOV11911	1169759-1170490	putative major capsid protein	CAC83558.1	Bacteriophage P27
SBOV11921	1170544-1170867	hypothetical phage protein	CAC83559.1	Bacteriophage P27
CDS	1170831-1171280	putative phage head-tail adaptor	EDX38959.1	*E.coli* 101-1
SBOV11931	1171246-1171770	hypothetical protein	CAC83560.1	Bacteriophage P27
SBOV11941	1171767-1172336	hypothetical protein	CAC83561.1	Bacteriophage P27
SBOV11951	1172340-1172501	conserved domain protein	EDX38932.1	*E.coli* 101-1
SBOV11961	1172498-1173994	putative sheath protein	CAC83562.1	Bacteriophage P27, SSH 3114-H10
SBOV11971	1173994-1174350	hypothetical protein P27p45	CAC83563.1	Enterobacteria phage phiP27, SSH 3114-H10
SBOV11981	1174350-1174679	hypothetical protein	CAC83564.1	Bacteriophage P27
SBOV11991	1174764-1176638	putative tail protein	AAN43571.1	*S. flexneri* 2a str. 301, SSH 3114-D07
SBOV12001	1176724-1177188	hypothetical protein SFxv_2260	ADA74406.1	*S.* flexneri 2002017
SBOV12011	1177234-1178625	hypothetical protein P27p48	CAC83566.1	Enterobacteria phage phiP27, SSH 3114-H12
SBOV12021	1178622-1179677	putative tail protein	CAC83567.1	Bacteriophage P27
SBOV12031	1179677-1180210	bacteriophage Mu Gp45 protein	EDX39029.1	*E. coli* 101-1

SBOV12041	1180216-1180629	hypothetical protein	CAC83569.1	Bacteriophage P27
SBOV12051	1180622-1181704	hypothetical protein	CAC83570.1	Bacteriophage P27, SSH 3114-A01
SBOV12061	1181704-1182294	tail protein	ACD09239.1	*S. boydii* CDC 3083-94
SBOV12071	1182395-1183156	side tail fiber protein	ACF68999.1	low match 76%
SBOV12081	1183156-1183737	phage tail assembly protein	ACF62949.1	low match 66%
SBOV12091	1183801-1185615	tail fiber protein	EEP88985.1	low match 54% Yersinia, SSH 3114-B06
SBOV12101	1185624-1186167	hypothetical phage protein	BAG77164.1	low match 66% *E.coli* SE11
SBOV12111	1186406-1187269	spermidine/putrescine transporter	AAL20154.1	
SBOV12121	1187253-1188389	spermidine/putrescine ABC transporter, ATP-binding protein	AAL20155.1	
SBOV12131	1188640-1189869	putative peptidase T(aminotripeptidase)	AAX65084.1	

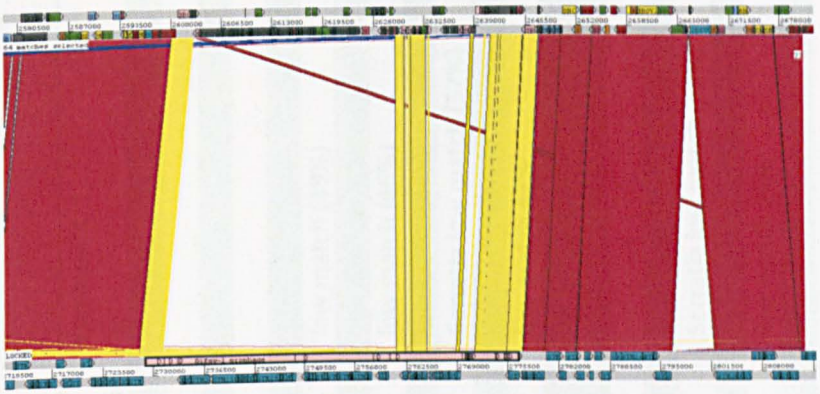


Figure 5.8 (A) ACT screenshot of the 3114 genome (top) and *S.* Typhimurium LT2 (bottom) showing the location of ROD34.

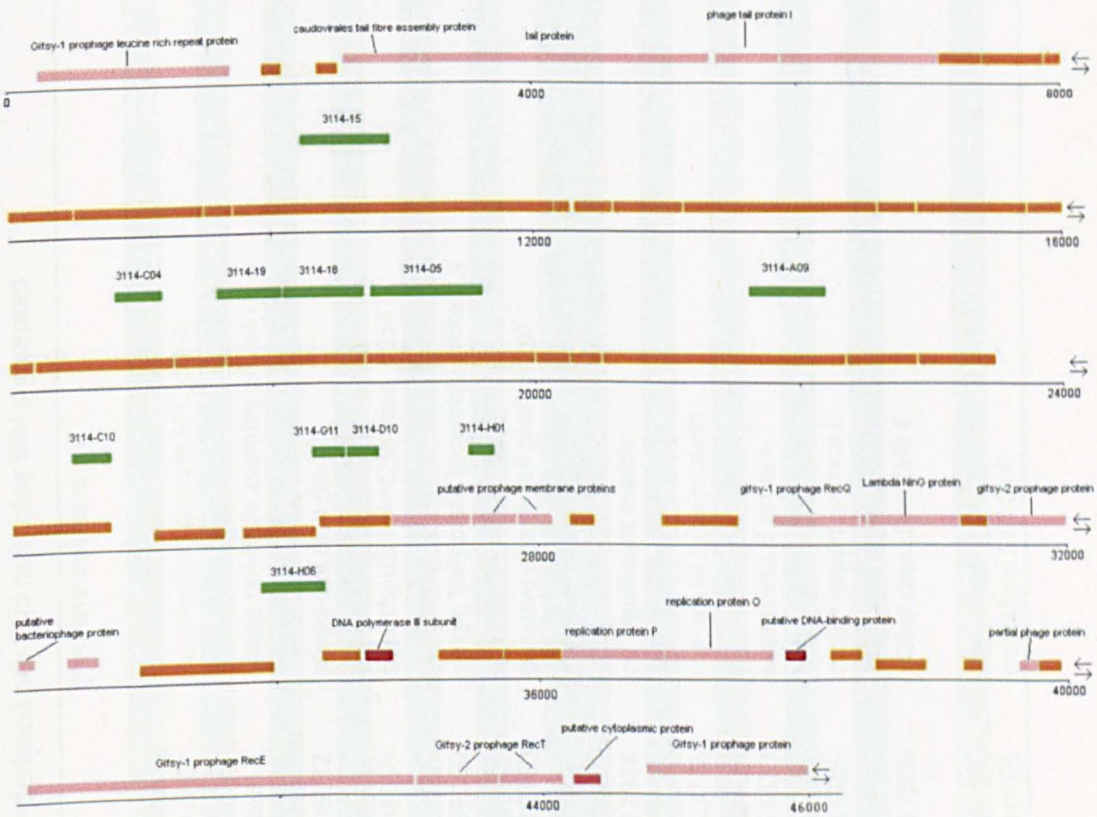


Figure 5.8 (B) shows the position of 3114-SSH sequences matching the novel ROD34 prophage

Table 5.8 ROD34 prophage, underlined are CDS matching SSH sequences

CDS	location	putative function	Accession number match	comments
SBOV26591	2600886-2602352	Gifsy-1 prophage leucine-rich repeat protein	AAX66485.1	Phage Gifsy-1
SBOV26601	2602597-2602749	hypothetical protein SPC_1064	ACN45230.1	
SBOV26611	2603023-2603181	hypothetical protein SeSA_A0712	ACF92271.1	
SBOV26621	2603226-2603813	phage tail-fibre asembly protein	CBG26188.1	
SBOV26631	2603813-2605996	tail protein	ACF92043.1	
CDS	2606051-2606632	phage tail protein I	ABV44048.1	low match (49%)
SBOV26641	2606625-2607749	Baseplate J family protein	ABV44049.1	Serratia low match (54%)
CDS	2607746-2608069	hypothetical protein Spro_4958	ABV44050.1	low match (46%)
SBOV26651	2608066-2608533	conserved hypothetical protein	ABV44051.1	low match (51%)
SBOV26661	2608530-2609150	phage baseplate assembly protein V	ABV44052.1	Serratia low match (57%)
SBOV26671	2609151-2610152	late control D family protein	ABV44002.1	serratia low match (62%)
SBOV26681	2610142-2610360	putative phage tail protein gpX	ADE57234.1	low match 42%
SBOV26691	2610353-2612809	phage tail tape measure protein, TP901 family	ABV44003.1	Serratia low match (66%), SSH 3114-15
SBOV26701	2612809-2612928	no significant match		
SBOV26711	2612964-2613245	hypothetical protein		low match 48%
SBOV26721	2613255-2613776	major tail tube protein	ABV44005.1	Serratia low match (58%)
SBOV26731	2613789-2615258	conserved hypothetical protein	ABV44006.1	Serratia low match 64%
SBOV26741	2615258-2615557	hypothetical protein		low match 25%
SBOV26751	2615557-2616042	hypothetical protein		low match 44%
SBOV26761	2616039-2616383	major capsid protein E	ABV44010.1	low match 74%
SBOV26771	2616390-2616845	transcriptional regulator (GntR family)	BAB50251.1	low match 34%
SBOV26781	2616846-2617889	major capsid protein E	ABV44010.1	low match 74%, SSH 3114-C04
SBOV26791	2617905-2618291	hypothetical protein Spro_4919	ABV44011.1	low match 52%
SBOV26801	2618302-2619366	peptidase S14 ClpP	ABV44012.1	low match 58%, SSH 3114-19

				and -18
SBOV26811	2619359-2620921	phage portal protein, lambda family	ABV44013.1	low match 72%, SSH 3114-05
CDS	2620918-2621157	conserved hypothetical protein	ACD54189.1	low match 48%
SBOV26821	2621169-2623004	terminase GpA	ABV44014.1	SSH 3114-A09
SBOV26831	2623010-2623555	conserved hypothetical protein	ABV44015.1	low match 56%
SBOV26841	2623555-2624145	ParB domain protein nuclease	ABV44016.1	low match 64%
SBOV26851	2624654-2625400	conserved hypothetical protein	ABV44017.1	low match 35%, SSH 3114-C10
CDS	2625729-2626265	conserved hypothetical protein	EDV61462.1	low match 34%
CDS	2624606-2626966	hypothetical protein SPAB_01002	ABX66424.1	low match 77%
SBOV26861	2626992-2627540	hypothetical protein SPAB_01000	ABX66422.1	SSH 3114-G11 and –D10
SBOV26871	2627537-2628151	lytic enzyme	EDY29407.1	
SBOV26881	2628151-2628432	hypothetical protein	BAE20156.1	SSH 3114-H01
SBOV26891	2628419-2628763	putative prophage membrane protein	EDZ28031.1	
SBOV26901	2628895-2629083	hypothetical protein STM14_3199	ACY89629.1	
SBOV26911	2629590-2630159	antirepressor-like protein	ACY89630.1	
SBOV26921	2630428-2631105	antiterminator-like protein	ACY89632.1	
CDS	2631102-2631242	hypothetical protein STM14_3204	ACY89633.1	
SBOV26931	2631239-2631850	bacteriophage Lambda NinG protein	ACF67531.1	
SBOV26941	2631853-2632059	hypothetical protein SPAB_02177	ABX67560.1	
SBOV26951	2632059-2632661	predicted bacteriophage protein	CBW18683.1	possible pseudogene
SBOV26961	2632696-2632812	putative bacteriophage protein	ACY87926.1	
SBOV26971	2633061-2633294	conserved domain protein	EDZ33636.1	
SBOV26981	2633609-2634637	protein of unknown function P63C	EFI46826.1	low match 50%, SSH 3114-H06
SBOV26991	2635006-2635296	retrotransposon protein, putative, Ty1-copia subclass	ABG22406.1	low match 33%, SSH 3114-H06
CDS	2635334-2635537	DNA polymerase III theta subunit	ADF60873.1	low match 59%
SBOV27001	2635888-263679	EA22-like protein	EFE62984.1	low match 46%
SBOV27011	2636376-2636825	conserved hypothetical protein	EDZ06389.1	

SBOV27021	2636822-2637523	replication protein P	EDZ23962.1	
SBOV27031	2637520-2638425	replication protein O	EDZ24001.1	
SBOV27041	2638517-2638666	Gifsy-1 prophage cI protein	AAX66539.1	
CDS	2638857-2639093	hypothetical protein SPAB_02191	ABX67574.1	
SBOV27051	2639197-2639580	hypothetical protein SPAB_02192	ABX67575.1	
SBOV27061	2639866-240009	hypothetical protein SPAB_02193	ABX67576.1	
SBOV27071	2640278-2640433	Gifsy-1 prophage protein	AAL21523.1	Phage Gifsy-1, possible pseudogene
SBOV27081	2640426-2640596	hypothetical protein STM14_1425	ACY87911.1	
SBOV27101	2640737-2643684	enterobacterial exodeoxyribonuclease VIII family protein	EDZ28239.1	possible pseudogene
CDS	2643695-2644318	gifsy-2 prophage RecT	EDZ01913.1	low match 78%
SBOV27111	2644327-2644803	Gifsy-2 prophage RecT	AAX64871.1	
SBOV27121	2644879-2645085	putative bacteriophage protein	ACF67123.1	
SBOV27131	2645431-2646660	bacteriophage integrase protein	YP_001700643.1	Phage Gifsy-1

5.2.4.1. Induction of bacteriophages from 3114

Induction of bacteriophages from 3114 was performed using norfloxacin, mitomycinC and UV light. None of these methods were successful in inducing bacteriophages from *S.* Bovismorbificans 3114, by forming plaques in the host strain *S.* Typhimurium LT2, however phages were successfully induced from *S.* Typhimurium D26104, which acted as a positive control, using the inducing agents mentioned above.

5.2.5. Pseudogene formation in *S.* Bovismorbificans 3114

Genes that are likely to have lost their function due to any one of a number of mutations have been identified and catalogued. Figure 5.2.9 sumarizes the type of mutations that have led to pseudogene formation in 3114. 113 pseudogenes were identified in 3114 using Artemis and ACT alignments with *S.* Typhimurium D23580, DT104 and SL1344. Table 5.9 summarizes pseudogenes identified in 3114. Highlighted in red are possible pseudogenes which may be due to sequencing errors and CDS assignments rather than true pseudogenes. The authenticity of these remains to be evaluated.

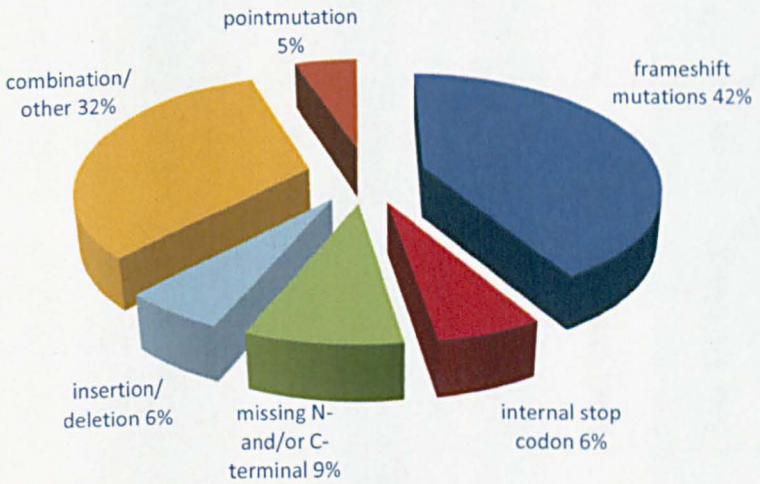


Figure 5.9. Summary of disabling mutations leading to pseudogene formation in 3114

Table 5.9. Position of pseudogenes on the genomes of *S.* Bovismorbificans 3114 , *S.* Typhimurium D23580, DT104 and SL1344. Pseudogenes that need further confirmation are highlighted in red and may be due to annotation errors, a ? indicates that no putative function could be determined.

SBOV 3114	Strand	position	STM D23580		SL1344		DT104		putative function	diabling mutation
CDS	-	29752 - 29844	STM_MW00861	Pseudo	SL0083A	pseudo	SDT0086	pseudo	putative secreted peptidase PipD (pseudogene)	SBOV CDS contains stop codon
CDS	-	32820 – 32885	STM_MW00891	Pseudo	SL0085	pseudo	SDT0089	pseudo	putative carnitinyl-CoA dehydratase (pseudogene)	Created CDS, was not present
SBOV00231	+	23046 – 23231	-	Absent	-	Absent	-	Absent	Putative uncharacterized protein	N-terminal missing in SBOV
SBOV00641	-	69044 – 69169	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV01181	+	126483 – 126843	STM_MW01661	Intact	SL0161	Intact	SDT0166	Intact	UPF0231 protein yacL	gap in sequence
SBOV01261	-	131844 – 131990	-	Absent	?	?	-	Absent	?	No close blast match
SBOV02101	+	218198 – 218633	STM_MW02461	Intact	SL0241	Intact	SDT0246	Intact	Peptidyl-tRNA hydrolase domain	frameshift in SBOV, pointmutation
SBOV02401	-	240461 – 240774	STM_MW03061	pseudo	SL0292	pseudo	SDT0297	pseudo	conserved hypothetical cytoplasmic protein	Frameshift, same as D23580
SBOV02411	+	241104 – 241569	STM_MW03081	pseudo	SL0294	pseudo	SDT0299	pseudo	putative IS1400 transposase B (pseudogene)	Frameshift
SBOV02481	-	250034 –	-	Absent	-	Absent	-	Absent	?	No close blast match

		250153								
SBOV02701	-	270053 – 270352	-	Absent	SL2753	Intact	-	Absent	IS3, transposase orfA	Internal stop codon
SBOV02721	+	270572 – 270860	STM_MW03961	pseudo	SL0320	pseudo	SDT0366	pseudo	glycerol dehydratase reactivation factor large subunit	Large deletion in SBOV, N-terminal missing, frameshift, was edited to match STM
SBOV02791	+	276895 – 277859	STM_MW04031	Intact	SL0328	Intact	SDT0373	pseudo	putative lysR family transcriptional regulator	Frameshift in SBOV
SBOV03101	-	312689 – 312850	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV03621	-	365379 – 365495	-	Absent	-	Absent	-	Absent	Putative uncharacterized protein	Missing start codon
SBOV03911	+	390913 – 392207	STM_MW05091	Intact	SL0432	Intact	SDT0480	Intact	Tetratricopeptide repeat protein	Frameshift
SBOV04341	-	436868 – 437916	STM_MW05491	Intact	SL0472	Intact	SDT0519	Intact	Putative transposase	Frameshift
SBOV 04521	+	452753 – 454405	STM_MW05641	Intact	SL0487	Intact	SDT0533	pseudo	UDP-sugar hydrolase	SBOV insertion of RV=ACG GGT at AA386
SBOV05141	-	517278 – 517523	-	Absent	-	Absent	-	Absent	CopE1	Internal stop codon
SBOV05171	+	518101 – 518337	STM_MW06271	pseudo	SL0550A	pseudo	SDT0587	pseudo	cation efflux protein (fragment)	Lacks start and stop
SBOV05271	-	528195 –	-	Absent	-	Absent	-	Absent	?	Lacks function

		528314								
SBOV06871	+	698367 – 698834	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV07771	-	784432 – 784965	STM_MW08911	Intact	SL0815	Intact	SDT0850	Intact	Sugar phosphatase SupH	Two internal stop codons
SBOV07961A	+	802608 – 803122	STM_MW09061	Intact	SL0830	Intact	SDT0865	Intact	Electron transfer flavoprotein	Frameshift
SBOV07961B	+	803041 – 803894	STM_MW09071	Intact	SL0831	Intact	SDT0866	Intact	electron transfer flavoprotein beta subunit	Frameshift
SBOV07981	+	804885 – 805148	-	Absent	-	Absent	-	Absent		Middle part missing
SBOV08031	+	810375 – 811263	STM_MW09121	Intact	SL0837	Intact	SDT0871	Intact	Putative oxidoreductase	Two frameshifts, changes, deletions, poor seq
SBOV09181		935970 – 936200	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV09191		936306 – 936461	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV09371	+	948508 – 948627	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV26951		2632059 – 2632661	STM_MW10311	Intact	SL0958	Intact	SDT0991	pseudo	Gifsy-2 hypothetical phage protein	Inverted in SBOV, substitution A--> T at AA 69 (GCC--->ACC)
SBOV09651	+	974241 – 974659	STM_MW10651	pseudo	SL0992	pseudo	SDT1029	pseudo	transposase (remnant)	Frameshift in both
SBOV09691	+	976443 – 976559	-	Absent	-	Absent	-	Absent	?	No close blast match

SBOV10101	-	1017599 – 1017748	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV10221	-	1027048 - 1027191	-	Absent	-	Absent	-	Absent	Putative uncharacterized protein	Pointmutations
SBOV11031	-	1101060 – 1104244	STM_MW11931	Intact	SL1122	Intact	SDT1159	Intact	ribonuclease E	Frameshift
SBOV11491	-	1143869 – 1143991	-	Absent	-	Absent	SDT1200	Intact	spermidine/putrescine transport system permease	No close blast match
SBOV13391	-	1295140 – 1295295	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV14591	+	1412016 – 1412267	-	Absent	-	Absent	-	Absent	Maltose regulon regulatory protein	C-terminal missing compared to blast
SBOV14691	-	1422925 – 1423230	-	Absent	-	Absent	-	Absent		No close blast match
SBOV14941	-	1445673 – 1445852	-	Absent	-	Absent			?	No close blast match
SBOV15541	+	1499611 – 1499757	-	Absent	-	Absent	-	Absent		No close blast match
SBOV15591	+	150627 – 1502785	-	Absent	-	Absent	-	Absent	Putative uncharacterized protein	No start codon, N-terminal missing
SBOV15601	-	1502895 – 1503104	-	Absent	-	Absent	-	Absent	?	No function, absent from D23580
SBOV15781	-	1523400 – 1526447	STM_MW15651	Pseudo	SL1500	pseudo	SDT1536	Pseudo ?	Putative molybdopterin oxidoreductase	internal stop codon, Frameshift in SBOV and D23580
SBOV16941	+	1643613 –	-	Absent	-	Absent	-	Absent	?	internal stop codon

		1643750								
SBOV17771	+	1725790 – 17525903	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV18931	+	1830023 – 1830310	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV44131			STM_MW18521	pseudo	SL1790	pseudo	SDT1825	pseudo	transposase (pseudogene)	2 frameshifts, insertions and rearrangements compared to D23580
SBOV19141	-	1843514 – 1843735	-	Absent	-	Absent	-	Absent	?	N-terminal missing in SBOV
SBOV20031	-	1917526 – 1917857	-	Absent	SL1874A	Intact	SDT1952	Intact	Putative inner membrane protein	Frameshift at nt 205
SBOV22421	+	2155812 – 2156009	STM_MW22051	Intact	-	Absent	-	Absent	Putative uncharacterized protein	Pointmutation, GAT->AAT
SBOV23111	-	222322 – 2223624	STM_MW22781	Intact	SL2223	Intact	-	Absent	Cytochrome c-type biogenesis pro	frameshift
SBOV23251	+	2236871 – 2237113	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV25041	+	2421257 – 2421796	STM_MW24581	Intact	SL2402	Intact	SDT2488	Intact	Putative membrane carboxypeptidase	Internal stop codon
SBOV25051	-	2421702 – 2421881	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV25461	+	2461963 – 2462115	-	Absent	-	Absent	-	Absent	Plasmid transfer protein	Seq missing

SBOV25801	-	2506255 – 2513562	STM_MW25311	pseudo	SL2475	pseudo	SDT2562	Intact	putative outer membrane protein (RatB) (pseudogene)	Pointmutation at nt 130, GAC->GTC, no frameshift in SBOV
SBOV26011	-	2545736 – 2545858	-	Absent	-	Absent			Putative uncharacterized pro	Pointmutation at nt34 compared to blast
SBOV26031	-	2546569 – 2546761	-	Absent	-	Absent			Putative uncharacterized pro	N-terminal missing in SBOV, at least 1 frameshift, 1 pointmutation at nt 45
SBOV26251	-	2565598 – 2566871	STM_MW25721	Intact	SL2517	Intact	SDT2603	Intact	Serine hydroxymethyltransferase	Frameshift
SBOV26371	-	2580041 – 2584028	STM_MW25821	Intact	SL2527	Intact	SDT2613	Intact	Phosphoribosylformylglycinamidin	pointmutation
SBOV26951	-	2632059 – 2632661	STM_MW26951	Intact	SL2582	Intact	-	Absent		Substitution at aa 69 T->A
SBOV27101	+	2640278 – 2640433	-	Absent	SL2597	Intact	SDT2682	Intact	Gifsy-1 prophage RecE	Frameshift
SBOV27071	+	2640737 – 2643684	STM_MW26501	pseudo	SL2596	pseudo	SDT2679	pseudo	putative phage protein (pseudogene)	Same as D23580
SBOV27841	-	2717753 – 2717908	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	Longer than all blast matches, mutation in stop codon?
SBOV27861	-	2719293 – 2720429	STM_MW27381	Intact	SL2756	Intact	SDT2769	Intact	Flagellin	Large part of N-terminal missing, poor seq

SBOV28101	-	2747352 – 2747486	-	Absent	-	Absent	-	Absent	?	No N-terminal, no close blast match
SBOV28771		2808586 – 2808699	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	N-terminal missing in SBOV
SBOV29301	-	2857804 – 2857986	-	Absent	-	Absent	SDT2900	Intact	?	C-terminal and stop missing
SBOV29651		2887999 – 2888211			Repeat region CRISPR		-	Absent	?	Appear intact (i.e. Start and stop) but no close blast match. Rearrangements compared to D23580, missassembly?
SBOV29661	+	2888351 – 2888617	-	Absent						
SBOV29681		2888832 – 2888966								
SBOV29691		2888931 – 2889047								
SBOV29701		2889228 – 2889581								
SBOV29771	-	2893935 – 2895470	-	Absent	-	Absent	-	Absent	Crispr-associated protein, Cse1	Internal stop codon
SBOV29871	+	2906236 – 2906532	-	Absent	-	Absent	-	Absent	?	No close blast match, rearrangements
SBOV30361	+	2959860 – 2961442	STM_MW29531	Intact	SL2970	Intact	SDT2984	Intact	Amino-acid N-acetyltransferase	Frameshift, poor seq

SBOV30411	-	2969920 – 2970036	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV30701	-	2996626 – 2996796	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	Pointmutation in nt10, change in start codon
SBOV30891	+	3013927 – 3014061	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	Pointmutation at nt 55
SBOV31011	+	3022441 – 3022566	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV32491	-	3159314 – 3159427	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	Start codon missing
SBOV32731	+	3184690 – 3185931	STM_MW31651	Intact	SL3177	Intact	SDT3197	Intact	Multifunctional CCA protein	Pointmutation in nt1236 G->A
SBOV35121	-	3407224 – 3408493	STM_MW34371	Intact	SL3412	Intact	SDT3427	Intact	Translation elongation factor EF Tu	pointmutation
SBOV35431	+	341079 – 3432508	STM_MW34641	Intact	SL3440	Intact	SDT3454	Intact	Protein tsgA	pointmutation
SBOV35501	+	3438487 – 3444280	STM_MW34691	Intact	SL3445	Intact	SDT3459	Intact	Putative surface-exposed virulence	pointmutation
SBOV36221	+	3524682 – 3525677	-	Absent	-	Absent	-	Absent	?	pointmutation
SBOV36901	-	3596394 – 3599952	STM_MW36051	Intact	SL3581	Intact	SDT3595	Intact	Putative polysaccharide biosynthase	pointmutation
SBOV37341	-	3642824 – 3643240	-	Absent	-	Absent	-	Absent	Integrase/transposase	Internal stopcodon, N-terminal missing

SBOV37821	+	3686571 – 3691080	STM_MW36791	Intact	SL3656	Intact	SDT3670	Intact	Putative autotransporter, Haemagglutinin family	Frameshift, pointmutation
SBOV37881	-	3695500 – 3695619	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV39051	-	3813778 – 3814864	STM_MW38021	Intact	SL3779	Intact	SDT3793	Intact	Putative autotransporter	sequence missing
SBOV39201	-	3827940 – 3828056	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	Pointmutation
SBOV39351	-	3846910 – 3847032	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV40231	-	3942291 – 3942812	STM_MW39161	Intact	-	Absent	-	Absent	Possible exported protein	Two frameshifts
SBOV40651	+	3986823 – 3987038	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	C-terminal missing, no close blast match
SBOV40881	-	4010954 – 4011076	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV41121	-	4030092 – 4033142	STM_MW40001	Pseudo?	SL3984	Pseudo ?	SDT4037	Pseudo ?	Formate dehydrogenase-O, major subunit	internal stop codon
SBOV41941	+	4103478 – 4104189	STM_MW40701	Intact	SL4054	Intact	SDT4107	Intact	5'-Nucleotidase domain protein	Internal stop codon
SBOV42241	+	4140898 – 4141217	STM_MW40991	Intact	SL4085	Intact	SDT4136	Intact	Elongation factor Tu	Framshift and large deletion
SBOV42991	-	4213019 – 4213174	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV43471	+	4257554 –	STM_MW42121	Intact	SL4197	Intact	SDT4249	Intact	large repetitive	C-terminals missing,

SBOV43481 SBOV43491 SBOV43501		4261888 4261934 – 4262848 4262896 – 4265220 4265214 – 4274177							protein	rearrangements, deletions
SBOV43741	-	4299310 – 4299441	-	Absent	-	Absent	-	Absent	Putative uncharacterized pro	No close blast match
SBOV43761	-	4299434 – 4300097	STM_MW42351	Intact	SL4220	Intact	SDT4272	Intact	Putative exported protein	pointmutation
SBOV43781	-	4300274 – 4302421	STM_MW42361	Pseudo	SL4221	Pseudo ?	SDT4273	Pseudo ?	Formate dehydrogenase H	Both internal stop codons
SBOV44101	-	4336734 – 4337066	-	Absent	-	Absent	-	Absent	?	No close blast match
SBOV44371	-	4358924 – 4360535	STM_MW42911	Intact	SL4280	Intact	SDT4328	Intact	Fumarate reductase, flavoprotein	pointmutation
SBOV44471	+	437262 – 4371423	-	Absent	-	Absent	-	Absent	?	absent
SBOV44671	+	4391149 – 4391833	STM_MW43171	Intact	SL4306	Intact	SDT4354	Intact	Putative cytoplasmic protein	pointmutation
SBOV45131	+	4435643 – 4435759	-	Absent	-	Absent	-	Absent	?	No close blast match

5.2.6. Single Nucleotide Polymorphism (SNP) analysis of African bacteraemia and veterinary *S.* Bovismorbificans Illumina genomes

In order to look for variations between strains, 25 *S.* Bovismorbificans genomes (including 3114) were genome sequenced using Illumina, 7 of which have been discounted from further analysis due to their unusual MLST type (see section 5.3 and Table 5.2 for details). Reads for the 18 remaining genomes were mapped using ssaha (www.sanger.ac.uk/resources/software/ssaha/) with the 3114 454 genome sequence acting as a reference sequence (for further details see materials and methods section 2.12.2.) Using an in-house script (Sanger Institute, Dr Simon Harris) single nucleotide polymorphism (SNP) analysis was performed. Both an accelerated and a delayed transformation parsimony tree was constructed. There was no difference between the two trees and it was decided to use an accelerated transformation as there is no biological data to support the use of delayed transformation (Figure 5.11). The 454 3114 genome sequence without prophage regions was initially used to contruct the tree and was subsequently overlayed with the full genome sequence containing prophage regions, which accounts for the long branch length. The layout of the tree was edited in figtree (http://tree.bio.ed.ac.uk/software/figtree/). Using the Illumina reads for each of the 18 genomes, Artemis was used to contruct a heatmap showing the presence and absence of sequence compared to 3114 in each strain, highlighted in red are the 454 genome sequence and the Illumina genome sequence of 3114. 3114 carries 1178 SNPs and forms a group with the paediatric bacteraemia strains D4891 and A1104 as well as the adult bacteraemia strain A1104; strains within this group differ by 25 SNPs. The 3 veterinary strains 653308, 276608 and 499208 form another group, carrying between 945 and 950 SNPs As a group they share approximately 390 SNPs with the 3114 group. The location and identity of the SNPs are visualized in Figure 5.12.

The 18 sequences include a plasmid sequence at the end, which was absent from the two veterinary strains 653309 and 276608, and showed some degree of variation among all of the strains. The map shows three major regions of variation which correspond to prophage RODs 13, 14 and 34 (indicated by black arrows in Figure 5.11), which have been discussed in section 5.4. (Figure 5.11). Table 5.10 summarizes the presence and/or absence of ROD13, -14 and -34, all three of which are absent from the four veterinary strains, ROD13 is partially present in 499208,

653308 and 276608. Human isolates A1608 and A1668 are lacking ROD34, A31126 does not carry ROD14, while A5893 lacks ROD13 (Table 5.10 and Figure 5.11).

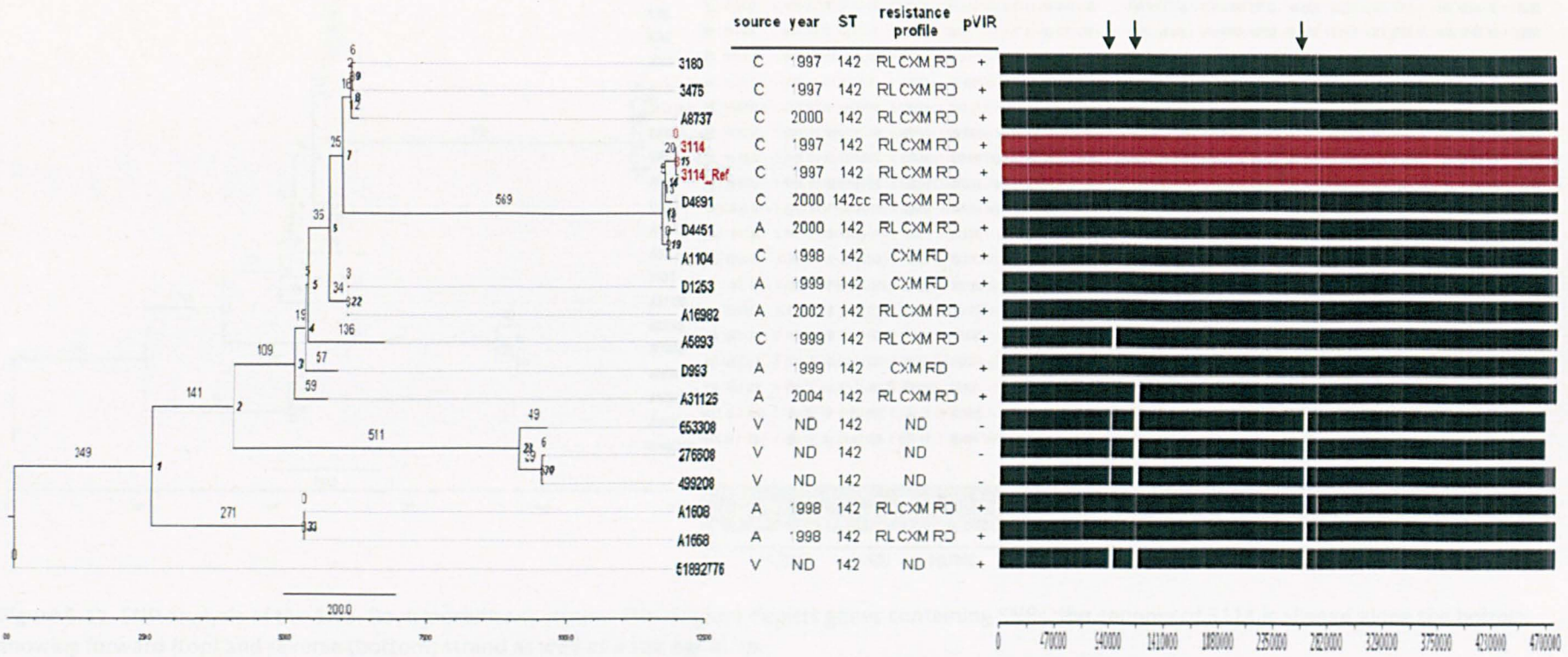


Figure 5.10. Accelerated parsimony tree of *S.* Bovismorbificans genomes, branch length corresponds to SNP number (raw data including prophage regions), node number is labelled in italics at each node, a table in the middle gives further information of each strain, including source (C=child, A=adult, V=veterinary), the year of isolation, the ST, antibiotic resistance profile (RL=Sulphamethoxazole, CXM= Cefuroxime, RD= Rifampicin) and virulence plasmid carriage, the heatmap was constructed from reads with dnaplotter, marked in red are 3114- 454 and Illumina genomes, arrows higlight RODs 13, 14 and 34.

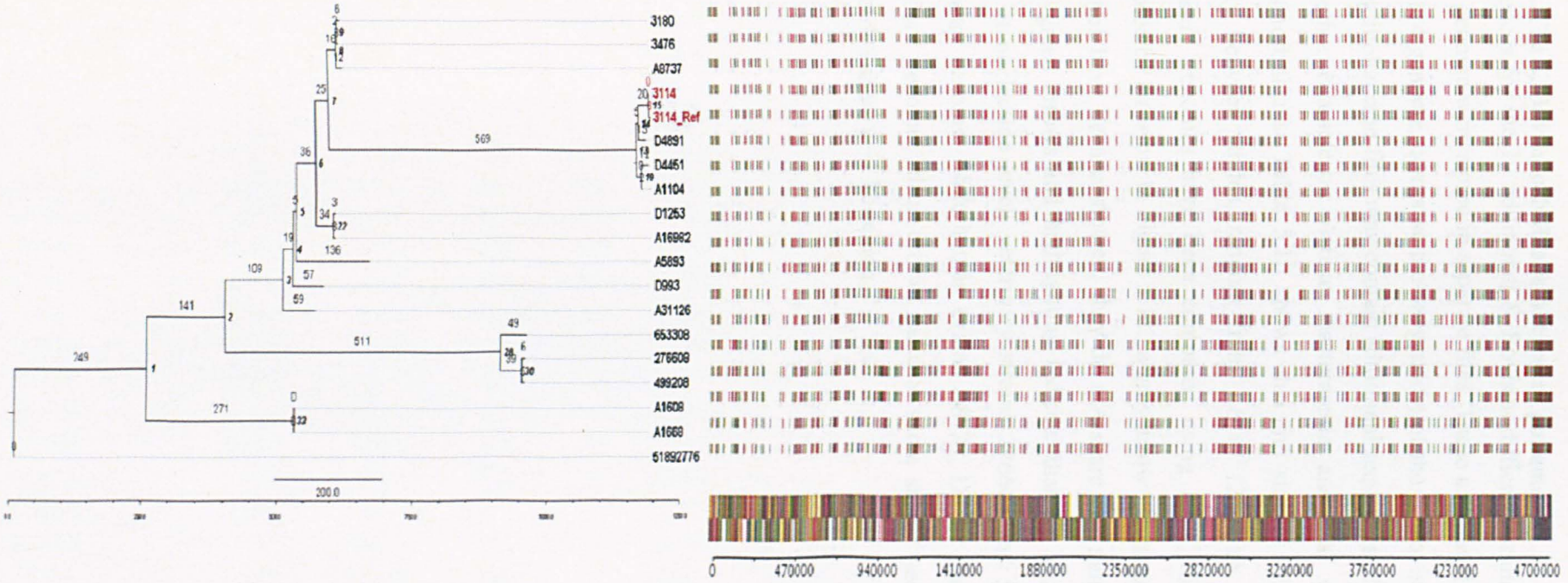


Figure 5.11. SNP analysis of the 18 *S*. Bovismorbificans strains. The diagram depicts genes containing SNPs, the genome of 3114 is aligned along the bottom showing forward (top) and reverse (bottom) strand as well as a size bar in bp.

5.2.7. The *S.* Bovismorbificans Accessory genomes

Sequences for the 17 additional *S.* Bovismorbificans strains that did not match the 3114 genome were stored in separate files. These additional sequences, together with the 3114 genome, comprise the pangenome of the *S.* Bovismorbificans serovar, and the additional read files may contain additional sequences of interest that will be used to further characterize African bacteraemia and UK veterinary strains of *S.* Bovismorbificans. Table 5.10 shows that the size of accessory genomes varies greatly between strains, ranging from 1.6 to 128.8 kb in size. A selection of accessory genomes have been compared using ACT alignments, comparison is indicated by an arrow in Figure 5.12., a grey arrow indicates that, although accessory genomes have been compared, they do not share any similarity at nucleotide (nt) level, green arrows and highlighting indicate that the accessory genomes are the same at nucleotide level, veterinary strains 276608 and 5189277 share the same accessory genome, while human strains A8737, D4451 and D4891 share the same accessory genome. Black arrows indicate some shared sequence identity. A16982 and A1608 share a ~1 kb region.

Table 5.10 Table summarizes the presence and absence of ROD13, -14 and -34 and the size of the Accessory genome in each of the 17 Illumina-sequenced *S. Bovismorbificans* genomes (ND = no data)

Strain	Isolation year	host	age	outcome	pSLT-like	Resistance profile	ROD13	ROD 14	ROD 34	Accessory genome size (bp)
human										
3114	1997	child	ND	ND	pVIR+	RL_CXM_RD	+	+	+	-
3180	1997	child	ND	ND	pVIR+	RL_CXM_RD	+	+	+	32482
3476	1997	child	ND	ND	pVIR+	RL_CXM_RD	+	+	+	128791
D993	1999	child	ND	ND	pVIR+	CXM_RD	+	+	+	95986
D1253	1999	child	4M	ND	pVIR+	CXM_RD	+	+	+	ND
D4451	2000	child	1Y11M	ND	pVIR+	RL_CXM_RD	+	+	+	1702
D4891	2000	child	11M	ND	pVIR+	RL_CXM_RD	+	+	+	1594
A1104	1998	adult	ND	3	pVIR+	CXM_RD	+	+	+	5752
A1608	Unk	adult	ND	ND	pVIR+	RL_CXM_RD	+	+	-	3160
A1668	1998	adult	ND	1	pVIR+	RL_CXM_RD	+	+	-	12637
A16982	2002	adult	ND	2	pVIR+	RL_CXM_RD	+	+	+	4942
A31126	2004	adult	23Y	1	pVIR+	RL_CXM_RD	+	-	+	55081
A5893	1999	adult	35Y	ND	pVIR+	RL_CXM_RD	-	+	+	42850
A8737	2000	adult	30Y	ND	pVIR+	RL_CXM_RD	+	+	+	1675
veterinary										
499208	Unk	alpaca	ND	ND	pVIR-	Unk	/	-	-	264628
653308	Unk	pig	ND	ND	pVIR-	Unk	/	-	-	97930
276608	Unk	pig	ND	ND	pVIR-	Unk	/	-	-	43984
51892776	Unk	pig	ND	ND	pVIR+	Unk	-	-	-	43903

+ present, - absent, / partial match

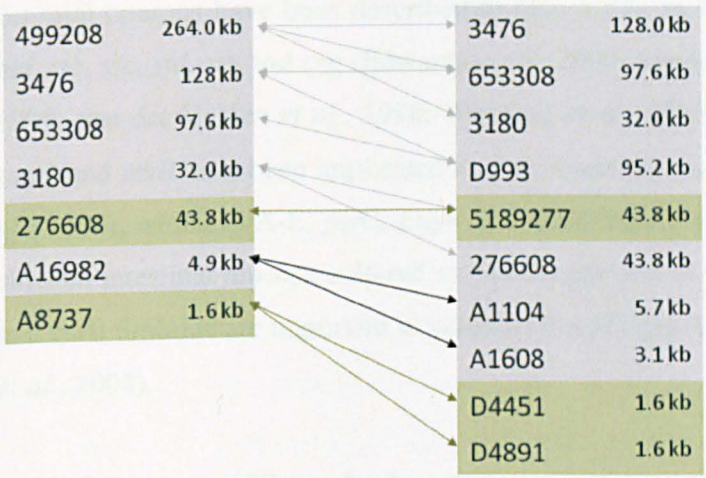


Figure 5.12 *S.* Bovismorbificans accessory genomes have been compared using ACT alignments. Sequences highlighted in green are the same at nucleotide (nt) level, grey arrow = accessory genomes share no similarity, green arrow = same nt sequence content, black arrow = some similarity at nt level

5.3. Discussion

5.3.1. MLST

The STs of 25 *S.* Bovismorbificans strains were determined from genome sequence and confirmed by PCR. Table 5.1. includes all STs identified through Illumina sequences, however a number of STs are ambiguous and it was decided to remove these from further analysis, due to the fact that they were not represented in the MLST database. Although these strains may still represent novel *S.* Bovismorbificans STs, investigation of O- and H-antigen would be necessary to confirm these findings.

5.3.3. The 3114 genome

The *S.* Bovismorbificans 3114 genome is 4.79 Mb in size, and contains a chromosome and a 93.6 kb virulence plasmid that contains the *spv* virulence gene cassette (Table 5.3.).

The genome of *S.* Bovismorbificans shares high similarity and co-linearity with that of *S.* Typhimurium LT2 and shares a number of common virulence-related features with Typhimurium and other *Salmonella* serovars, such as a repertoire of 13 fimbrial operons (*stf, saf, stb, fim, stc, std, lpf, stj, sth, bcf, sti, csg*; and *pef* which is located on

the Bovismorbificans virulence plasmid, Figure 5.3 and 5.6.). Eight of these 13 fimbrial operons have been described to play a role in virulence in mice: *lpf, fim, bcf, stb, stc, std, sth* and *csg* (Edwards *et al.*, 2000; Lawley *et al.*, 2006; Tsolis *et al.*, 1999; van der Velden *et al.*, 1998; Weening *et al.*, 2005). The fimbrial genes *stbC*, *csgD* and *sthB* have been implicated in the colonisation of avian gut cells (Morgan *et al.*, 2004), whilst *lpf*A-E, *pef*C, *csg*A and *fim*H have a role in biofilm formation on chicken intestinal mucosa cultured ex-vivo (Ledeboer *et al.*, 2006). Type I fimbriae and curli fimbriae are important in colonisation of eggs (Cogan *et al.*, 2004; De Buck *et al.*, 2004).

5.3.4. *S.* Bovismorbificans Pathogenicity Islands

Salmonella Pathogenicity Islands (SPIs) 1-5 and 9, as well as two partials of SPI-6, are present in *S.* Bovismorbificans 3114. SPI-1 is involved in host cell invasion and macrophage apoptosis. *Salmonella* carrying mutations in SPI-1 genes are unable to orally infect mice (Galan & Curtiss, 1989 reviewed by Collazo & Galan, 1997b). SPI-2 is located adjacent to $tRNA^{val}$. The SPI-2 genes are required for systemic infection and intra-macrophage survival (Shea *et al.*, 1996).

SPI-3 is inserted downstream from *selC* and contains the *mgtCB* operon necessary for intra-macrophage survival and survival in low Mg^{2+} environments (Blanc-Potard *et al.*, 1999). SPI-3 in 3114 is ~13 kb and shows a small deletion of 4.4 kb, compared to SPI-3 in *S.* Typhimurium LT2 which contains *sugR* and *rhuM*, as well as two additional CDS. The *sugR-rhuM* region of SPI-3 showed greater variation than any other SPI investigted, and within SPI-3 this region showed higher variability than the remainder on SPI-3. *S. enterica* I serovars Bovismorbificans, Derby, Hessarek, Infantis, Ohio, Ratchaburi, Virchow and Zanzibar all showed variations of this region, with *sugR* and *rhuM* absent from some of the isolates including *S.* Bovismorbificans (Amavisit *et al.*, 2003). The region is approximately 4200bp in size in LT2 compared to 220 bp in 3114, which is smaller than reported previously (~700bp) (Amavisit *et al.*, 2003). SPI-4 of *S.* Bovismorbificans 3114 is 25.5 kb in size and includes a variation, compared to SPI-4 of LT2, in a gene encoding the putative inner membrane protein (STM4261). In its place 3114 carries SBOV43471, a putative inner membrane protein, which shares an identical N-terminal with STM4261, two putative Ig domains (SBOV43481 and SBOV43491) and a large

repetitive protein (SBOV43501), the C-terminal region of which is identical to the C-terminal region of STM4261.

SPI-5 is 9.1 kb in size and is identical to SPI-5 of *S.* Typhimurium LT2, except for the insertion of one hypothetical protein. SPI-5 carries *pipA,B,D*, *sopB* and *copS.*, Amavisit *et al.* showed variation of restriction enzyme sites amongst different serovars but no difference in gene content (Amavisit *et al.*, 2003).

The two partials of SPI-6 are located at 239898-243437 and 242054-251907. Together they are ~10.6 kb in size, compared to the whole SPI-6 which is 47 kb in *S.* Typhimurium LT2. SPI-6 or *Salmonella* Chromosomal Island (SCI) of *S.* Typhimurium contains the *saf* gene cluster encoding fimbriae and the invasion encoding *pagN* gene, as well as several other genes with unknown functions. *pagN* was absent from the *S.* Bovismorbificans regions matching parts of SPI-6. These SPI-6 partials encompass a region encoding cytoplasmic proteins and part of the *saf* operon with *safA* appearing absent or incomplete. The variable presence of *safA* in *S. enterica* serovars has previously been described by Clayton and co-workers (Clayton *et al.*, 2008).

The *saf* operon is not required for *S.* Typhimurium virulence in mice (Folkesson *et al.*, 1999; Morgan *et al.*, 2004), however it appears that *saf* has a role in porcine intestinal infection by *S.* Typhimurium (Carnell *et al.*, 2007).

The *pagN* gene is a PhoP-activated (*pag*) gene the function of which is not completely determined. PagN shares 54% similarity with Tia and Hek adhesins/invasins of pathogenic *E. coli*, and it was shown that *pagN* was necessary for bacterial survival in BALB/c mice (Fagan & Smith, 2007; Fagan *et al.*, 2008; Fleckenstein *et al.*, 1996; Heithoff *et al.*, 1997; Heithoff *et al.*, 1999). Lambert and Smith showed that PagN agglutinates erythrocytes and mediates adhesion to and invasion of mammalian cells by *S.* Typhimurium. They further showed evidence that expression of multiple copies of *pagN* can compensate for loss of the SPI-1 T3SS (Lambert & Smith, 2008).

Deletion of the SPI-6 locus has no effect on *S.* Typhimurium's systemic pathogenicity in mice (Folkesson, *et al* 2002). Parts of SPI-6 were also detected in *Salmonella* subspecies IIIb, IV and VII. Homologs of SPI-6 have also been identified in enterohaemorrhagic *E. coli*, *Pseudomonas aeruginosa* and *Yersinia pestis* genome sequences, but the function of these homologues is unknown (Hensel, 2004).

SPI-9 is 16.9 kb in size in 3114 and shares 100% nucleotide identity to SPI-9 of *S.* Typhimurium. SPI-9 shares ~40% nucleotide identity with SPI-4. Three of the four genes are homologous to those required for the T1SSs, SPI-9 also contains a large repetitive protein (SBOV2765), in *S.* Typhimurium LT2 this gene carries a frameshift mutation and is therefore likely to have lost its function, however in *S.* Bovismorbificans the gene appears to be intact. It is thought that therefore SPI-9 may not be functional in *S.* Typhimurium, leaving only SPI-4 functional. Sequence analysis identified a highly conserved operon polarity suppressor (*ops*) motif, which is required for transcription elongation under control of the RfaH protein, and is associated with virulence gene clusters such as those required for the production of LPS in *S.* Typhimurium, Vi antigen in *S.* Typhi, and α-haemolysin and K5 capsule in pathogenic *E. coli* The large ORF STM 4261 of LT2 contains a stop codon in *S.* Typhi and is therefore organised into two ORFs. (Bailey *et al.*, 1997; Morgan *et al.*, 2004). In *S.* Bovismorbificans 3114 this same sequence is organised into four ORFs (SBOV43471 to SBOV43501) and are located on two different frames.

5.3.4. Prophage variation between *S.* Bovimorbificans 3114 and *S.* Typhimurium LT2

A number of deletions in the *S.* Bovismorbificans 3114 genome compared to *S.* Typhimurium LT2 are notable, these are mainly related to prophages. Fels-1 and Fels-2 are absent from the Bovismorbificans 3114 genome, while prophage Gifsy-1 and -2 appear to have been replaced by other Gifsy-like prophages. Fels-1 is a 47.9kb lambdoid Siphoviridae and contains two potential virulence genes, the neuraminidase *nanH* and the superoxide dismutase *sodC3* (Figueroa-Bossi *et al.*, 2001; Yamamoto, 1969). Fels-2 is 33.7 kb in size; it shares 33 homologs with PSP3 and 29 homologs with P2. Fels-2 can recombine with the morphologically unrelated phage P22 to form F22, which is serologically related to Fels-1 but carries the P22 *c* genes. The prophage integrates into the 3'end of the host *ssrA* genes. One of the unique features of this prophage is the presence of a DNA adenine methylase (DAM-STM2730 (Yamamoto, 1969; Yamamoto & McDonald, 1986).

Gifsy-1 in *S.* Typhimurium LT2 is 47.84 kb in size and is integrated into the 5'end of the host *lepA* gene, which encodes a ribosome-binding GTPase. Gifsy-1 is inducible through mitomycin C and UV light. Both of these techniques were tried with *S.* Bovismorbificans 3114 but we were unable to detect phage induction on the recipient

strain LT2, which confirms the findings that Gifsy-1-like prophage in *S.* Bovismorbificans is sufficiently different. Gifsy-1 carries a number of potential virulence genes including *gipA*, which is involved in colonization of the small intestine. Deletion of this gene results in reduced bacterial virulence (Stanley *et al.*, 2000), The Gifsy-1 prophage-related region in 3114 located at the insertion site of Gifsy-1 in LT2 has been designated ROD34. ROD34 is 45.8 kb in size. ROD34 carries Gifsy-1 like elements at the biginning and end, as well as one Fels-1 like element (Table 5.8). Comparsions with *S.* Typhimurium D23580, DT104 and SL1344 show different phages present in the location of ROD34 in each strain of the three *S.* Typhimurium strains. The ROD34 prophage partially matches Gifsy-2 SLP272 of SL1344 (Table 5.4), the whole prophage ROD34 is summarized in Table 5.4.

The Region of Difference labelled ROD13, matches the location of the Gifsy-2 prophage in the genome of *S.* Typhimurium LT2. Gifsy-2 is 45.5 kb in size in LT2 and integrates between *pncB* (nicotinate phosphoribosyl transferase) and *pepN* (aminopeptidase), which is probably defective in LT2. Gifsy-2 contains a number of virulence genes implicated in host pathogenesis, such as *gtgA*, deletion of which results in seven fold reduction in virulence, and *sodC1* (periplasmic superoxide dismutase) deletion of which attenuates virulence by five-fold in BALB/c mice (Ho *et al.*, 2002).

A 15.3 kb sequence of Gifsy-2 has been conserved at the C-terminal end of the ROD13 prophagein 3114, while the 30.2 kb of ROD13 differs between LT2 and 3114 (Table 5.6). ROD13 appears to be a variable region as it also different in *S.* Typhimurium D23850, where it has been labelled BTP2, and in DT104 and SL1344 (Table 5.4.).

A large insertion, 46.4 kb in size, downstream of SPI-5 has been labelled ROD14 (Figure 5.8 (A)), ROD14 appears to have inserted into a spermidine/putrescine transporter system (SBOV11451 and SBOV12111), a bacterial periplasmic transport system first described in *E. coli* by Furuchi and co-workers (Furuchi *et al.*, 1991).

ROD14 appears to be prophage-related, carrying putative genes for phage head and tail proteins as well as a number of genes matching hypothetical proteins and a *sifA* gene (SBOV11471). 14 of the 73 ROD14 CDS match *E. coli* strains more closely than *Salmonella* strains in the database. A 13.1 kb region of the 46.4 kb prophage

closely matches bacteriophage P27 (SBOV11891-SBOV12051, nt 1168610-1181704). Predicted CDSs within this region match a putative prohead protease (SBOV11891), a major capsid protein (SBOV11901), a sheath protein (SBOV11961) and a tail protein (SBOV12021) (Table 5.7.). ROD14 was also absent from *S.* Typhimurium D23580, DT104 and SL1344.

5.3.5. The 3114 virulence plasmid

Salmonella serovars belonging to subspecies I carry a large virulence plasmid. *S.* Bovismorbificans 3114 was found to carry a 93.65 kb virulence plasmid (Figure 5.6.) that contains the *spv* virulence gene cassette (nt55870-60514 separate plasmid file). Despite their low copy number (1-2 per chromosome) the plasmids are very stable. Their involvement in the enteric stage of infection is unclear. *Salmonella* virulence plasmids are heterogeneous in size (50-90 kb), but all of them share a 7.8 kb region encoding for *spv* (*Salmonella* plasmid virulence) gene cassette, which is necessary for multiplication of these serovars in the reticulo-endothelial system of warm-blooded vertebrates. It has been shown that cloning the *spv* locus into a low copy number vector was sufficient to restore virulence in a plasmid-cured strain (Gulig, 1990).

Other loci of the plasmid encode for PEF (plasmid-encoded fimbriae) (nt 78409-82825). The *pef* operon consists of *pefBACD*, orf5, orf6 and *pefI*. Pef mediate adhesion to murine intestinal epithelial cells, resulting in fluid accumulation (Baumler *et al.*, 1996a), regulation of these fimbriae is mediated through phase variation similar to the Pap system in *E. coli* (Nicholson & Low, 2000). *S.* Typhimurium mutants defective in Pef show a significant increase in biofilm formation (Ledeboer *et al.*, 2006). Located downstream of the *pef* operon is a member of the serum resistance group of genes: the *rcK* (resistance to complement killing) gene (SBOV47961). RcK is a 17kDa outer membrane protein that is homologous to a group of Gram negative bacterial outer membrane proteins involved in virulence phenotypes. Rck is most closely related to PagC (PhoP activated gene C), which is required by *S.* Typhimurium for survival in macrophages and for virulence in mice (Heffernan *et al.*, 1992). Another closely-related virulence-related OMP is Ail, a product of the *Yersinia enterocolitca* chromosome capable of mediating bacterial adherence to and invasion of epithelial cell lines (Heffernan *et*

al., 1992). Deletions and variations in orf5, orf6 of the *pef* operon and *rck* may be involved in host adaptation (Chiu *et al.*, 2004).

Salmonella virulence plasmids of *S.* Enteritidis, *S.* Typhimurium and *S.* Choleraesuis carry two independent replication genes, *repB* and *repC*, which are involved in maintaining the low copy number. Host adapted *Salmonella* serovars such as Dublin, Gallinarum and Pullorum appear to only carry *repB*. The RepA of *S.* Bovismorbificans is identical to that of the *S.* Choleraesuis plasmid, and almost identical (99%) to those encoded by the plasmids of *S.* Typhimurium and *S.* Dublin (Lopez *et al.*, 1991). The *repC* replicon may be associated with the control of serum resistance. RepC shows the typical organization of the RepFIB family, with a set of direct repeats (iterons) that exert stringent control of copy number. The first *repC* iterons (B, C, and D) correspond to the reported sequence of the *rsk* (resistance to serum killing) fragment, which has been associated with the control of serum resistance (Vandenbosch *et al.*, 1989). The *S.* Bovismorbificans plasmid does not appear to carry *rsk*. The third member of the serum resistance genes is *traT* (SBOV48111), The *traT* gene encodes a surface lipoprotein homologous to the product of the *traT* surface exclusion gene located on plasmid F and F-like conjugation systems (Rhen & Sukupolvi, 1988). *traT* is a member of the *tra* operon (transfer genes), some of which are necessary for non-sexual transfer of the plasmid. The presence of a complete or incomplete set of *tra* operon genes indicates the transfer by conjugation to the *Salmonella* ancestor and the subsequent divergence of the different serovars. *S.* Bovismorbificans lacks the *traS* gene and *traY* and *traJ* appear incomplete, compared to pSLT. The *traS* gene product appears to work in concert with *traT*, reducing DNA transfer in stable mating partners (Achtman *et al.*, 1977). *traY* together with *traD* is involved in conjugational metabolism, by nicking the origin of tranfer (Boyd & Hartl, 1997; Fowler *et al.*, 1983). *traJ* lies upstream of *traY* and *traJ* mutants fail to express the *tra* operon genes (Achtman *et al.*, 1972). *traJ* is located in the outer membrane of the bacterial envelope, the protein is a positive transcriptional regulator (Fowler *et al.*, 1983).

Contrary to high copy number plasmids which can rely on random distribution, low copy number plasmids need to be positioned in both dividing cells, ensuring that each progeny carries a copy of the plasmid. These observations have led to the discovery of the partitioning genes called *parA* and *parB*. *parA* encodes a membrane-

associated ATPase that is essential for this symmetric movement of the ParB foci, while *parB* encodes the DNA binding protein ParB.

ParA may facilitate ParB movement along the inner surface of the cytoplasmic membrane to encounter and become tethered to the next replication zone (Cerin & Hackett, 1993 reviewed by Bignell & Thomas, 2001).

The *samAB* operon (*Salmonella* mutagenesis genes) appears to have a role in UV mutagenesis. It is a homolog of the *umuDC* operon in *E. coli*. *S.* Typhimurium carries both *umuDC* on the chromosome and samAB on the virulence plasmid. It appears that *samAB* is recessive to *umuDC* (Nohmi *et al.*, 1992). The *S.* Bovismorbificans plasmid appears to lack *samA* but there are two copies of *samB* (SBOV48541 and SBOV48571) flanking the *par* operon, one of which appears to contain a frameshift mutation (SBOV48541). The *S.* Bovismorbificans chromosome contains intact *umuC* (SBOV20581) and *umuD* (SBOV20591) genes.

Salmonella serovars such as Typhi and Paratyphi lack virulence plasmids and therefore the *spv* region. Interestingly, plasmids are more frequently found in *S.* Typhimurium and Enteritidis strains isolated from blood and other extraintestinal sources than in strains isolated from faeces, however the importance of the virulence plasmid in bacteraemic disease can only be inferred from this indirect evidence (Rotger & Casadesus, 1999). Direct evidence for the importance of virulence plasmids of some serovars in systemic disease is lacking. The virulence plasmid affects intracellular growth in macrophages but not in non-phagocytic cells (Rotger & Casadesus, 1999). There was no antimicrobial resistance cassette located on the 3114 virulence plasmid. In *Salmonella* these resistance islands have been identified on both the chromosome and on plasmids (Miriagou *et al.*, 2006). 3114 does not appear to carry additional plasmids. Mobile genetic elements such as integrons, transposons and insertion sequences are exchanged between bacteria during horizontal gene transfer and/or through translocation, where genes move between locations such as the chromosome and plasmids. The tetracycline resistance conferring Tn*1721* has been identified on conjugative and mobilizable plasmids in various *Salmonella* isolates (Allmeier *et al.*, 1992). The *strA-strB* genes, which mediate streptomycin resistance, are often located on plasmids as part of the transposon Tn*5393* (Miriagou *et al.*, 2006). In 2008 Caveco and co-workers sequenced a 4270 bp plasmid (FJ228229) of *S.* Bovismorbificans from a MDR isolate from China, the plasmid carried a novel *qnrD* gene conferring reduced

susceptibility to fluroquinolones (Cavaco *et al.*, 2009). No such plasmid has been identified in *S.* Bovismorbificans 3114.

5.3.6. *S.* Bovismorbificans 3114 pseudogenes

In contrast to eukaryotic genomes bacterial genomes show a high coding density of over 80% (Kuo *et al.*, 2009), however recent large scale analysis of bacterial genomes has shown that eroded or disrupted genes are ubiquitous in bacterial genomes (Andersson & Andersson, 2001; Karro *et al.*, 2007; Lerat & Ochman, 2005; Liu *et al.*, 2004). These pseudogenes may arise through degradation, where N- and/or C-terminal ends have eroded away. Point mutations may disrupt genes through internal stop codons or prevent the protein from functioning properly, or through frameshift mutations within the gene. Pseudogenes are particularly prevalent in bacteria that undergo a shift towards or specialization towards a particular eukryotic host (Andersson *et al.*, 1998). In some extreme cases, such as rickettsia, pseudogenes can number in their thousands, affecting over half of the genome (Cole *et al.*, 2001; Toh *et al.*, 2006). Although the population structure of *S. enterica* is essentially clonal, *Salmonella* genomes possess considerable numbers of pseudogenes. The number of pseudogenes in even closely related servars such as Gallinarum and Enteritidis varies from 147 to 21 respectively (Kuo & Ochman). It appears that pseudogenes are serovar specific. Of the 147 pseudogenes in the genome of *S.* Gallinarum, only five are shared with Enteritidis, only three of which carry the same debilitating mutation. The majority of *Salmonella* pseudogenes are due to single mutations (91.5%) and amongst those, small deletions at the C-or N-terminal were the most common (41%). Deletions of C- or N-termini were found in 9% of 3114 pseudogenes (Figure 5.2.10). The majority of pseudogenes in the genome of *S.* Bovismorbificans 3114 were due to frameshift mutations (42%) or a combination of other mutations (32%), however some of these may be due to sequencing errors (Figure 5.2.10). A total of 113 pseudogenes were identified in the genome of *S.* Bovismorbificans 3114 using Artemis and ACT alignments with *S.* Typhimurium D23580, DT104 and SL1344. These are summarized in Table 5.9.

An example of a gene disabled through an internal stop codon is *pipD* (pathogenicity island-encoded protein). The *pip* operon is located at the end of SPI-5 immediatly upstream of *serT*, however this CDS is located at nt 29752-29844 outside of SPI-5. The PipD protein is a structural homologoue of *Lactobacillus* dipeptidases, which

may indicate that PipD is a secreted peptidase targeting the host cell (Wood *et al.*, 1998).

5.3.7. 17 *S.* Bovismorbificans Illumina-sequenced genomes

An accelerated parismony tree, based on SNPs, and a heatmap have been constructed from the 18 Illumina sequenced genomes (Figure 5.11.). Phage–related sequences are normally excluded from SNP analysis, however from the SSH data and the reciprocal fasta searches it was apparent that *S.* Bovismorbificans is similar to *S.* Typhimurium and other *Salmonella* serovars and the decision was made to include prophage regions, which accounts for the long branch length. Accelerated parismony which ascribes SNPs as soon as possible within the evolution is commonly used if there are no further data present to support the use of delayed parismony. Both trees have been drawn but did not differ significantly and therefore the decision was made to use accelerated parismony. The heatmap,constructed using ssaha, mapped reads for each strain which has been aligned using ACT. The 454 genome sequence of *S.* Bovismorbificans 3114 was used as a reference that all 17 *S.* Bovismorbificans genomes have been mapped onto and is included in the heatmap (shown in red, alongside the Illumina sequence of 3114). The 454 and Illumina genomes of 3114 only differ by approximately 3 SNPs, which indicates a good quality of the Illumina-derived genome sequence. The heatmap illustrates that the major regions of variation between *S.* Bovismorbificans genomes are the three prophage related regions ROD13, ROD14 and ROD34 (Figure 5.11), the contents of which are summarized in section 5.2.4. ROD14 and 34 were absent from UK veterinary isolates and ROD13 was only partially present in three out of four veterinary isolates and absent from the fourth. A number of adult bacteraemia isolates also lacked some of the prophage regions. ROD14 was absent from the human bacteraemia isolate A31126, while ROD34 was absent from A1608 and A1668, ROD13 was absent from A5893. All of the paediatric bacteraemia isolates (designated by D) carried the full set of prophage regions. The heatmap also includes the plasmid sequence and therefore shows that the virulence plasmid is absent from three (653308, 276608 and 499208) of the four veterinary strains. This is in keeping with the observation that virulence plasmids are more common in blood isolates than faecal isolates (Rotger & Casadesus, 1999). Figure 5.12 summarizes the SNPs present in each of the 18 *S.* Bovismorbificans strains.

5.4. Further Work

The Bovismorbificans "Accessory genome" remains to be investigated, the 17 Illumina genome sequences were mapped using the 3114 454 genome as a reference. Sequences not present in this genome were therefore not assembled into the genomes; these read files remain to be investigated for the presence of further virulence related sequences and possible plasmids such as the resistance plasmid described by Cavaco and co-workers (2009). The presence and homology of pseudogenes in the Illumina genomes remains to be established and SNPs will be mapped to pseudogenes and related to the stage in Bovismorbificans evolution. It also remains to be seen if any SNPs are associated with antimicrobial resistance genes. Three novel prophage regions account for major variation between African bacteraemia and UK veterinary strains, it would be important to compare these strains to human diarrhoea isolates to determine, the importance of these prophages in invasive infections and possibly their specificity to Africa.

5.5. Conclusions

Bacteriophages appear to be the driving force in African invasive NTS (Kingsley et al, 2009). We have demonstrated that three novel prophage regions account for major differences between African human bacteraemia and UK veterinary Bovismorbificans strains.

Evidence for genome degradation (pseudogene formation) appears to support development towards human adapted invasive NTS serovars (Holt *et al* , 2008). An increase in virulence plasmid carriage was clearly demonstrated and has been related previously to invasive NTS infections (Rotger & Casadesus, 1999).

5.6. Summary

- Three prophage-related RODs, ROD13, ROD14 and ROD34, have been identified, that define major genetic variation between between *S.* Bovismorbificans strains.
- Gifsy-1 and -2 are prophages normally present in the *S.* Typhimuirum genome and have a known role in virulence; hence the novel Gifsy-like prophages in *S.* Bovismorbificans may play a similar role,
- Fels prophages are absent from the *S.* Bovismorbificans genome

- Table 5.9. summarizes 113 pseudogenes have been identfied in the 3114 genome, compared to 25 in *S.* Typhimutium LT2, which may indicate genomic degradation occuring in *S.* Bovismorbificans 3114.
- *S.* Bovismorbificans strain 3114 carries a 94kb virluence plasmid, containing the *spv* virulence cassette, similar to pSLT of S. Typhimurium LT2; this plasmid was present in 15 out of 18 isolates

CHAPTER 6: GENERAL DISCUSSION

6.1. General Discussion and Limitations

NTS have long been a common but neglected cause of invasive disease in young children from sub-Saharan Africa. Research and public health efforts in developing countries tend to focus around HIV, malaria and tuberculosis (TB), with the cause of febrile illness often being mistaken for malaria. NTS are the commonest or second commonest cause of bacteraemia in children and HIV infected adults and a common cause of focal infections in children in sub-Saharan Africa. *S.* Typhimurium and Entertidis are the most frequently isolated serovars, while *S.* Typhi infections are generally rare in these countries. MDR is a great problem in sub-Sahran Africa. Data from Malawi showed MDR to commonly used antimicrobials in *S.* Enteritidis increasing sharply in 1998 and MDR in *S.* Typhimurium arising in 2001 (Gordon *et al.*, 2008). A better understanding of the pathogenesis of NTS would help to improve treatment and outcome. Evidence has been mounting that African bacteraemia NTS strains differ from common gastroenteritis strains. Using basic typing techniques Kariuki and co-workers demonstrated that human bacteraemia isolates differed from isolates found in the environment of patients, in domestic and food animals. This has led to the conclusion that strains may be distributed by human to human transmission (Kariuki *et al.*, 2002; Kariuki *et al.*, 2006a). Host-adapted serovars such as Gallinarum are continuously introduced into humans from their reservoir species, but do not circulate within humans. It is possible that Bovismorbificans and other host specific serovars actually circulate within the human population.

MLST carried out on Malawian and Kenyan adult *S.* Typhimurium bacteraemia isolates showed that the majority belonged to a novel sequence type ST313, that had previously been reported only twice outside of Africa, interestingly both of which came from bacteraemia cases (Kingsley *et al.*, 2009).We were able to confirm these findings in that paediatric and adult bacteraemia isolates tested from Malawi, Kenya and the DRC also matched ST313.

We applied SSH to four pediatric Malawian bacteraemia isolates of four different serovars, *S.* Typhimurium, *S.* Entertidis, *S.* Heidelberg and *S.* Bovismorbificans, and subsequently screened a panel of NTS isolates from Malawi, Uganda, Kenya, the DRC and the UK for the presence or absence of a total of 33 SSH sequences using PCR assays and in some cases DNA-DNA-hybridisation. SSH is a technique capable

of identifying sequences that differ between species, serovars or strains of bacteria, although it was originally developed to identify differences in expression of eukaryotic genes (Diatchenko *et al.*, 1996).

The choice of driver strain is important. We have taken two different approaches: (1) using driver strains of the same and (2) of a different serovar. One of the rationales behind choosing *S.* Typhimurium LT2 was the fact that it had previously been genome sequenced and is well characterised (McClelland *et al.*, 2001). It is therefore possible to test each SSH sequence for tester specificity using BLAST searches of the driver genome, thereby excluding any unsubtracted sequences from future analysis. Similarly the genome sequence of *S.* Enteritidis NCTC13349 has been published on the Sanger Institute website and is available for BLAST searches there. By subtracting the genome of LT2 from the genome of *S.* Heidelberg and *S.* Bovismorbificans, we were able to identify a wider range of different types of sequences, some of which are potentially specific to those serovars. The genome of *S.* Typhimurium LT2 was also subtracted from that of the Malawian isolate *S.* Typhimurium D26104, resulting in a number of bacteriophage and antimicrobial resistance-related sequences, illustrating the point of how same-serovar subtractions will select for recently acquired genetic elements. We also subtracted the genome of *S.* Enteritidis NCTC13349 from that of *S.* Enteritidis D21685, which resulted in a very limited number of sequences, suggesting that there is little variation between the *S.* Enteritidis tester and driver strains. It was possible to show some geographical distribution of a number of *S.* Typhimurium and *S.* Bovismorbificans SSH sequences through PCR assays.

We were able to show that *S.* Heidelberg carries two components of a fimbrial operon, *tcf,* that was previously described in *S.* Typhi and Paratyphi. *S.* Heidelberg further contains the *tcf* transcriptional regulator *tinR.* Expression patterns of *tcf* related genes and other fimbriae related genes appears to vary between *S.* Heidelberg strains of different origins. We went on to carry out non-phagocytic cell invasion assays. Three strains with significantly higher levels of invasiveness were identified, which were PCR positive for all 14 SSH sequences with the exception of D23734-B10, which was absent from the veterinary isolate 20070502 (Table S1) However the expression profile of KMS1977 and 845 differed. Our findings have led us to the conclusion that fimbrial genes play an important part in *S.* Heidelberg diversity and

may contribute to an increase in virulence. We further showed that although genes may be present they may not be expressed.

SSH of *S.* Bovismorbificans and subsequent distribution PCR assays and BLAST searches identified a number of phage related sequences. Distribution analysis showed that Bovismorbificans SSH sequences were completely absent from *S.* Typhimurium isolates according to PCR assays and BLAST searches. Given the fact that *S.* Typhimurium LT2 was used as a driver strain and that five SSH sequences correlated with serotype, these findings predicted a high similarity between *S.* Typhimurium and *S.* Bovismorbificans, which was subsequently confirmed through whole-genome comparison.

We genome sequenced the first *S.* Bovismorbificans strain 3114, a paediatric bacteraemia isolate from Malawi, using 454 sequencing technology and used Illumina next generation sequencing on 17 further *S.* Bovismorbificans Malawian and UK veterinary strains, for SNP analysis and comparison purposes.

S. Bovismorbificans strains varied through the presence or absence of three novel prophage regions predicted by SSH and a virulence plasmid highly homologous to pSLT of S. Typhimurium LT2. The four veterinary *S.* Bovismorbificans isolates lacked the three prophage regions and three out of the four lacked the virulence plasmid. We also derived the ST from the genome sequence of all 18 *S.* Bovismorbificans strains, which were either ST142 or ST142cc.

MLST is a useful and widely used typing technique. However these findings highlight the fact that due to the nature of the technique it is not suitable for determining variations in the accessory genome of a strain.

Genome sequencing is evolving rapidly and is becoming very cost efficient, with 3rd generation sequencing on the horizon, however annotation tools are evolving at the same rate, rather than being available ahead of obtaining the data and therefore data analysis is often time consuming. 454 pyrosequencing was used to sequence the *S.* Bovismorbificans reference strain 3114, which was then assembled de novo. Typically the read length resulting from 454 sequencing technology is 200-300 bp. The Illumina system was used to genome sequence a further 17 *S.* Bovismorbificans strains. Illumina sequencing technology is based on the sequencing by synthesis concept. Reads are typically very short, ~32-40bp, and assembly of Illumina reads requires a reference genome. *S.* Bovismorbificans 3114 was also Illumina sequenced and 454 and Illumina sequence data were overlayed to improve coverage and the

overall quality of what subsequently became the reference genome for the other 17 Illumina-sequenced *S.* Bovismorbificans genomes. The added complication for this work was that the reference genome, 3114, was still constructed and improved as the Illumina sequences were mapped onto the reference, drawing out the process and requiring some steps of analysis to be repeated. We have encountered a number of limitations when it came to analysing the genome sequence data, even though the reference genome was of good quality with high coverage. We carried out pseudogene analysis by manually scanning the genome for the presence of mutations within CDSs that may lead to loss of function of a gene, such as a missing start or stop codon or frameshift mutations.

This was carried out on the original 454 assembly, where we initially identified in the region of 145 pseudogenes. However the corrected version of the 3114 genome contains ca 113 pseudogenes, with a high number of the original pseudogenes being due to frameshifts that have been resolved through the new assembly. During 454 genome assembly a string of Ns is inserted between scaffolds that did not show overlap. These regions inevitably lead to frameshifts within putative CDSs. Nevertheless we do believe that genome degradation through pseudogenes formation is an important step towards host specialization and that in NTS from sub-Saharan Africa this specialization is directed towards the human host, similarly to what historically occurred in the human adapted invasive *Salmonella* serovar Typhi (Andersson & Andersson, 2001; Holt *et al.*, 2009; Kingsley *et al.*, 2009; Kuo & Ochman; Thomson *et al.*, 2008).

In order to ascribe a putative function to SSH sequences we have relied on the NCBI database. However the rapid influx of unfinished genomes, made possible through next-generation sequencing, has led to a number of problems.

BLAST searches of SSH sequences regularly return a long list of hypothetical proteins, while at the beginning of this work only finished genomes were published reducing the number of hypothetical functions and listing a number of putative functions from homologs in related species such as *E.coli.* Further problems occur due to mis-annotations. This is illustrated through the example of *S.* Enteritidis SSH sequences D21685-F7 and SE3 being mis-annotated as related to *lasA* (Chapter 3).

6.2. General Conclusions

It has been demonstrated here and by others that NTS causing invasive disease from sub-Sahran Africa differ from common gastroenteritis and veterinary strains in a number of ways. Prophage regions are a major source of genetic variation between strains of the same serovar. We were able to demonstrate that this is true for isolates of *S.* Bovismorbificans from Malawi and the UK, while Kingsley and co-workers (2009) described the presence of novel prophages in Malawian *S.* Typhimurium.

We demonstrated that not merely the presence of novel sequences can account for differences in virulence, but possibly differences in expression of virulence-related genes. Genome sequencing enabled us to identfy a number of pseudogenes in the genome of *S.* Bovismorbificans, indicating genome degradation and possible host specialization of bacteraemia isolates.

We confirmed the presence of a virulence plasmid in *S.* Bovismorbificans isolates that showed variable distribution.

6.3. Future work

- It would be of interest to determine the MLST ST of NTS taken from the environment and livestock of bacteramia patients from sub-Sahran Africa to further confirm the original findings, that human bacteraemia isolates differ from those of the patients environment (Kariuki *et al.*, 2002). However as discussed above MLST will not show any differences in the accessory genome, such as the presence of antimicrobial resistance genes.
- We were able to demontrate differences in phage carriage between human African bacteraemia and UK veterinary isolates. It would be interesting to test a number of human diarrhoea isolates from developed countries for the presence of prophages and virulence plasmids to determine the difference between human and veterinary and human bacteraemia and human diarrhoea isolates.
- The accessory genome of the *Salmonella* serovar Bovismorbificans remains to be determined in full. During the time of this work it was only possible to describe the accessory genome of a single strain (3114) in full. It remains for us to anaylse the accessory genome of the remaining 17 genome sequenced *S.* Bovismorbificans strains, particularly those of veterinary origin.

- It will also be important to confirm the presence of pseudogenes identfied in S. Bovismorbificans 3114 in the 17 strains and assign SNPs to the pseudogenes identified.
- It would be benefital to further investigate the expression profile of a greater number of SSH sequences, possibly under different growth conditions to determine the importance of preliminary observations made for *S.* Heidelberg virulence related SSH sequences.

S. SUPPLEMENTARY DATA

Table S1 complete set of SSH sequence PCR distribution assays. strains used as tester strains in SSH are highlighted in red.Every PCR reaction was carried out twice.

	NTS strains				
	Uganda				
	146	666	812	868	4448
	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium
D23734-1 hypothetical protein	-	-	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-
D23734-A1 putative fimbrial protein	-	-	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-
D23734-D2 ferrichrome-iron receptor	-	-	-	-	-
D23734-D4 transcriptional regulator	-	-	-	-	-
D23734-G3 fimbrial subunit	-	-	-	-	-
D21685 unknown	-	-	-	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	+	+	+	+	+
D26104-ST2 Tnp [Enterobacteria phage P7]	+	+	+	-	-
D26104-H11 G protein [Enterobacteria phage 186]	-	+	+	-	-
D26104-11 Tum [Enterobacteria phage 186]	+	+	+	+	+
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	+	+	+	+	+
D26104-9 colicin E1 protein [Plasmid ColE1]	+	+	+	-	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	+	+	+	+	+
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	+	+	+
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+	+	+	+
D26104-E11 AadA1 [E. coli]	-	+	+	+	+
3114-11 putative phosphotransferase system component	-	-	-	-	-
3114-18 peptidase S14, ClpP	-	-	-	-	-
3114-A9 phage terminase GpA	-	-	-	-	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-
3114-C12 putative outer membrane usher protein	-	-	-	-	-
3114-F10 Gifsy-1 prophage terminase large chaing gp2	-	-	-	-	-
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-

NTS strains	884	4234	1060	1190	81	4002	93	4016	204	213	263	4068
	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis
D23734-1 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-	-	-	-	-	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-	-	-	-	-	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-	-	-	-	-	-	-	-
D23734-A1 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-D2 ferrichrome-iron receptor	-	-	-	-	+	+	+	+	+	+	+	+
D23734-D4 transcriptional regulator	-	-	-	-	-	-	-	-	-	-	-	-
D23734-G3 fimbrial subunit	-	-	-	-	+	+	+	+	+	+	+	+
D21685 unknown	-	-	-	-	-	-	-	-	-	-	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	-	-	-	-	-	-	-	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	+	+	+	+	-	-	-	-	-	-	-	-
D26104-ST2 Tnp [Enterobacteria phage P7]	-	+	-	-	+	-	-	-	+	+	+	+
D26104-H11 G protein [Enterobacteria phage 186]	-	-	-	-	-	-	-	-	-	-	-	-
D26104-11 Tum [Enterobacteria phage 186]	+	+	+	+	-	-	-	-	-	-	-	-
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	+	+	+	+	-	-	-	-	-	-	-	-
D26104-9 colicin E1 protein [Plasmid ColE1]	-	+	-	-	-	-	-	-	-	-	-	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	+	+	+	+	+	-	-	-	+	+	+	+
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	+	+	+	-	-	-	+	+	+	+
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+	+	+	+	-	-	-	+	+	+	+
D26104-E11 AadA1 [E. coli]	+	+	+	+	-	-	-	-	-	-	-	-
3114-11 putative phosphotransferase system component	-	-	-	-	+	+	+	+	+	+	+	+
3114-18 peptidase S14, ClpP	-	-	-	-	-	-	-	-	-	-	-	-
3114-A9 phage terminase GpA	-	-	-	-	-	-	-	-	-	-	-	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-	-	-	-	-	-	-	-
3114-C12 putative outer membrane usher protein	-	-	-	-	-	-	-	-	-	-	-	-
3114-F10 Gifsy-1 prophage terminase large chaing gp2	-	-	-	-	+	+	-	-	+	+	+	+
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-	-	-	-	-	-	-	-

	NTS strains							Malawi			
	298 S. Enteritidis	300 S. Enteritidis	791 S. Enteritidis	1181 S. Enteritidis	1305 S. Enteritidis	256 S. Stanleyville	845 S. Heidelberg	D23734 S. Heidelberg	D21685 S. Enteritidis	D26104 S. Typhimurium	3114 S. Bovismorbificans
D23734-1 hypothetical protein	-	-	-	-	-	+	+	+	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-	-	+	+	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-	-	+	+	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-	-	+	+	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-	-	+	+	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-	-	+	+	-	-	+
D23734-A1 putative fimbrial protein	-	-	-	-	-	-	+	+	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-	+	+	+	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-	-	+	+	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-	-	+	+	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-	+	+	+	-	-	+
D23734-D2 ferrichrome-iron receptor	+	+	+	+	+	+	+	+	+	-	-
D23734-D4 transcriptional regulator	-	-	-	-	-	+	+	+	-	-	-
D23734-G3 fimbrial subunit	+	+	+	+	+	-	+	+	+	-	+
D21685 unknown	-	-	-	+	-	-	-	-	+	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	+	-	-	-	-	+	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	-	-	-	-	-	-	-	-	-	+	-
D26104-ST2 Tnp [Enterobacteria phage P7]	-	+	-	-	+	-	-	-	-	+	-
D26104-H11 G protein [Enterobacteria phage 186]	-	-	-	-	-	-	-	-	-	+	-
D26104-11 Tum [Enterobacteria phage 186]	-	-	-	-	-	-	-	-	-	+	-
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	-	-	-	-	-	-	-	-	-	+	-
D26104-9 colicin E1 protein [Plasmid ColE1]	-	-	-	-	-	-	-	-	-	+	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	+	+	-	-	+	-	-	-	-	+	-
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	-	-	+	-	-	-	-	+	-
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+	-	-	+	-	-	-	-	+	-
D26104-E11 AadA1 [E. coli]	-	-	-	-	-	-	-	-	-	+	-
3114-11 putative phosphotransferase system component	+	+	+	+	+	+	-	-	+	-	+
3114-18 peptidase S14, ClpP	-	-	-	-	-	-	-	-	-	-	+
3114-A9 phage terminase GpA	-	-	-	-	-	-	-	-	-	-	+
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-	-	-	-	-	-	+
3114-C12 putative outer membrane usher protein	-	-	-	-	-	-	-	-	-	-	+
3114-F10 Gifsy-1 prophage terminase large chaing gp2	+	+	-	-	+	-	-	-	-	-	+
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-	-	-	-	-	-	+

NTS strains														
D20941	S. Typhimurium													
D20890	S. Typhimurium													
D22076	S. Typhimurium													
D22209	S. Typhimurium													
D22219	S. Typhimurium													
D22337	S. Typhimurium													
D22404	S. Typhimurium													
D22988	S. Typhimurium													
D23002	S. Typhimurium													
D23424	S. Typhimurium													
D23674	S. Typhimurium													
D24627	S. Typhimurium													
														D23734-1 hypothetical protein
														D23734-2 hypothetical protein
														D23734-4 2-nitropropane dioxygenase NPD
														D23734-5 hypothetical protein
														D23734-10 short-chain dehydrogenase/reductase SDR
														D23734-14 putative autotransporter/pertactin
														D23734-A1 putative fimbrial protein
														D23734-A5 putative fimbrial protein
														D23734-B2 hypothetical protein
														D23734-B9 conserved hypothetical protein
														D23734-B10 probable lipoprotein
														D23734-D2 ferrichrome-iron receptor
														D23734-D4 transcriptional regulator
														D23734-G3 fimbrial subunit
														D21685 unknown
														D21685-3 Rom-like protein [E. coli]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-ST1** EaA [Enterobacteria phage P22]
														D26104-ST2 Tnp [Enterobacteria phage P7]
														D26104-H11 G protein [Enterobacteria phage 186]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-11** Tum [Enterobacteria phage 186]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
														D26104-9 colicin E1 protein [Plasmid ColE1]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
+	+	+	+	+	+	+	+	+	+	+	+	+	+	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
+		+	+	+	+	+	+	+	+	+	+	+	+	**D26104-E11** AadA1 [E. coli]
														3114-11 putative phosphotransferase system component
														3114-18 peptidase S14, ClpP
														3114-A9 phage terminase GpA
														3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]
														3114-C12 putative outer membrane usher protein
														3114-F10 Gifsy-1 prophage terminase large chaing gp2
														3114-G4 hypothetical protein [Stx2 converting bacteriophage]

NTS strains	D25352	D25834	D25840	D25907	D25991	D23145	D23682	D8993.1	D8993.2	D993	811	2291
	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Bovismorbificans	S. Bovismorbificans	S. Bovismorbificans
D23734-1 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-	-	-	-	-	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-	-	-	-	-	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-	-	-	-	-	+	+	+
D23734-A1 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-	-	-	-	-	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-	-	-	-	-	+	+	+
D23734-D2 ferrichrome-iron receptor	-	-	-	-	-	+	+	+	+	-	-	-
D23734-D4 transcriptional regulator	-	-	-	-	-	-	-	-	-	-	-	-
D23734-G3 fimbrial subunit	-	-	-	-	-	+	+	+	+	+	+	+
D21685 unknown	-	-	-	-	-	+	+	-	-	-	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	-	-	+	+	-	-	-	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	+	+	+	+	+	-	-	-	-	-	-	-
D26104-ST2 Tnp [Enterobacteria phage P7]	-	-	-	-	-	-	-	-	-	-	-	-
D26104-H11 G protein [Enterobacteria phage 186]	-	-	-	-	-	-	-	-	-	-	-	-
D26104-11 Tum [Enterobacteria phage 186]	+	+	+	+	+	-	-	-	-	-	-	-
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	+	+	+	+	+	-	-	-	-	-	-	-
D26104-9 colicin E1 protein [Plasmid ColE1]	-	-	-	-	-	-	-	-	-	-	-	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	+	+	+	+	+	-	-	-	+	-	-	-
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	+	+	+	-	-	-	+	-	-	-
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+	+	+	+	-	-	-	+	-	-	-
D26104-E11 AadA1 [E. coli]	+	+	+	+	-	-	-	-	-	-	-	-
3114-11 putative phosphotransferase system component	-	-	-	-	-	+	+	+	+	+	+	+
3114-18 peptidase S14, ClpP	-	-	-	-	-	-	-	-	-	+	+	+
3114-A9 phage terminase GpA	-	-	-	-	-	-	-	-	-	+	+	+
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-	-	-	-	-	+	+	+
3114-C12 putative outer membrane usher protein	-	-	-	-	-	-	-	-	-	+	+	+
3114-F10 Gifsy-1 prophage terminase large chaing gp2	-	-	-	-	-	-	-	+	+	+	+	+
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-	-	-	-	-	+	+	+

NTS strains

3476 S.Bovismorbificans	170 S.Bovismorbificans	3064 S.Bovismorbificans	3160 S.Bovismorbificans	3180 S.Bovismorbificans	2819 S.Bovismorbificans	5104 S.Bovismorbificans	6571 S.Bovismorbificans	A7360 S.Bovismorbificans	D4891 S.Bovismorbificans	D4896 S.Bovismorbificans	D1253 S.Bovismorbificans	
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-1** hypothetical protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-2** hypothetical protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-4** 2-nitropropane dioxygenase NPD
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-5** hypothetical protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-10** short-chain dehydrogenase/reductase SDR
+	X	X	X	X	X	X	X	X	X	X	X	**D23734-14** putative autotransporter/pertactin
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-A1** putative fimbrial protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-A5** putative fimbrial protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-B2** hypothetical protein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-B9** conserved hypothetical protein
+	X	X	X	X	X	X	X	X	X	X	X	**D23734-B10** probable lipoprotein
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-D2** ferrichrome-iron receptor
-	X	X	X	X	X	X	X	X	X	X	X	**D23734-D4** transcriptional regulator
+	X	X	X	X	X	X	X	X	X	X	X	**D23734-G3** fimbrial subunit
-	X	X	X	X	X	X	X	X	X	X	X	**D21685** unknown
-	X	X	X	X	X	X	X	X	X	X	X	**D21685-3** Rom-like protein [E. coli]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-ST1** EaA [Enterobacteria phage P22]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-ST2** Tnp [Enterobacteria phage P7]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-H11** G protein [Enterobacteria phage 186]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-11** Tum [Enterobacteria phage 186]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-9** colicin E1 protein [Plasmid ColE1]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
-	X	X	X	X	X	X	X	X	X	X	X	**D26104-E11** AadA1 [E. coli]
+	+	-	+	+	+	+	+	+	+	+	+	**3114-11** putative phosphotransferase system component
+	-	-	+	+	+	+	+	+	+	+	+	**3114-18** peptidase S14, ClpP
-	+	+	+	+	+	+	+	+	+	+	+	**3114-A9** phage terminase GpA
-	-	+	+	+	+	+	+	+	+	+	+	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]
+	-	-	+	+	+	+	+	+	+	+	+	**3114-C12** putative outer membrane usher protein
-	-	+	+	+	+	+	+	+	+	+	+	**3114-F10** Gifsy-1 prophage terminase large chaing gp2
+	+	+	+	+	+	+	+	+	+	+	+	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]

NTS strains	D2877	A5893	D9582	A10387	D4551	7822	A9112	A8737	12855	A1104	A1668	A9362
	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans
D23734-1 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-2 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-4 2-nitropropane dioxygenase NPD	X	X	X	X	X	X	X	X	X	X	X	X
D23734-5 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-10 short-chain dehydrogenase/reductase SDR	X	X	X	X	X	X	X	X	X	X	X	X
D23734-14 putative autotransporter/pertactin	X	X	X	X	X	X	X	X	X	X	X	X
D23734-A1 putative fimbrial protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-A5 putative fimbrial protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B2 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B9 conserved hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B10 probable lipoprotein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-D2 ferrichrome-iron receptor	X	X	X	X	X	X	X	X	X	X	X	X
D23734-D4 transcriptional regulator	X	X	X	X	X	X	X	X	X	X	X	X
D23734-G3 fimbrial subunit	X	X	X	X	X	X	X	X	X	X	X	X
D21685 unknown	X	X	X	X	X	X	X	X	X	X	X	X
D21685-3 Rom-like protein [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-ST1 EaA [Enterobacteria phage P22]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-ST2 Tnp [Enterobacteria phage P7]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-H11 G protein [Enterobacteria phage 186]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-11 Tum [Enterobacteria phage 186]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-9 colicin E1 protein [Plasmid ColE1]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-E11 AadA1 [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
3114-11 putative phosphotransferase system component	+	+	+	-	+	+	+	+	+	+	+	+
3114-18 peptidase S14, ClpP	+	+	+	-	+	+	+	+	+	+	-	-
3114-A9 phage terminase GpA	+	+	+	-	+	+	+	+	+	+	-	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	+	+	+	-	+	+	+	+	+	+	-	-
3114-C12 putative outer membrane usher protein	+	+	+	-	+	+	+	+	+	+	+	+
3114-F10 Gifsy-1 prophage terminase large chaing gp2	+	-	+	+	+	+	+	+	+	+	+	+
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	+	+	+	-	+	+	+	+	+	+	+	-

NTS strains	A1608	A15523	A16934	A16982	D15823	A15363	A18317	A19308	D18327	D19094	D19206	A22622
	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans	S.Bovismorbificans
D23734-1 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-2 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-4 2-nitropropane dioxygenase NPD	X	X	X	X	X	X	X	X	X	X	X	X
D23734-5 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-10 short-chain dehydrogenase/reductase SDR	X	X	X	X	X	X	X	X	X	X	X	X
D23734-14 putative autotransporter/pertactin	X	X	X	X	X	X	X	X	X	X	X	X
D23734-A1 putative fimbrial protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-A5 putative fimbrial protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B2 hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B9 conserved hypothetical protein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-B10 probable lipoprotein	X	X	X	X	X	X	X	X	X	X	X	X
D23734-D2 ferrichrome-iron receptor	X	X	X	X	X	X	X	X	X	X	X	X
D23734-D4 transcriptional regulator	X	X	X	X	X	X	X	X	X	X	X	X
D23734-G3 fimbrial subunit	X	X	X	X	X	X	X	X	X	X	X	X
D21685 unknown	X	X	X	X	X	X	X	X	X	X	X	X
D21685-3 Rom-like protein [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-ST1 EaA [Enterobacteria phage P22]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-ST2 Tnp [Enterobacteria phage P7]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-H11 G protein [Enterobacteria phage 186]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-11 Tum [Enterobacteria phage 186]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-9 colicin E1 protein [Plasmid ColE1]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	X	X	X	X	X	X	X	X	X	X	X	X
D26104-E11 AadA1 [E. coli]	X	X	X	X	X	X	X	X	X	X	X	X
3114-11 putative phosphotransferase system component	+	+	+	+	+	+	+	+	+	-	+	-
3114-18 peptidase S14, ClpP	'	+	+	+	+	+	+	+	+	-	+	-
3114-A9 phage terminase GpA	'	+	+	+	+	+	+	+	+	-	+	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	'	+	+	+	+	+	+	+	+	-	+	-
3114-C12 putative outer membrane usher protein	+	+	+	+	+	+	+	+	+	-	+	-
3114-F10 Gifsy-1 prophage terminase large chaing gp2	+	+	+	+	+	+	+	+	+	-	+	-
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	+	+	+	+	+	+	+	+	+	-	+	-

NTS strains

S. Sundsvall 22402	S. Bukavu 3064	S.Bovismorbificans D9375	S.Bovismorbificans D28655	S.Bovismorbificans D28312	S.Bovismorbificans A31126	S.Bovismorbificans A30633	S.Bovismorbificans A24091	S.Bovismorbificans A23856	S.Bovismorbificans D19953	S.Bovismorbificans A22920	S.Bovismorbificans A22921	
+	-	X	X	X	X	X	X	X	X	X	X	**D23734-1** hypothetical protein
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-2** hypothetical protein
-	+	X	X	X	X	X	X	X	X	X	X	**D23734-4** 2-nitropropane dioxygenase NPD
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-5** hypothetical protein
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-10** short-chain dehydrogenase/reductase SDR
+	-	X	X	X	X	X	X	X	X	X	X	**D23734-14** putative autotransporter/pertactin
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-A1** putative fimbrial protein
-	+	X	X	X	X	X	X	X	X	X	X	**D23734-A5** putative fimbrial protein
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-B2** hypothetical protein
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-B9** conserved hypothetical protein
-	+	X	X	X	X	X	X	X	X	X	X	**D23734-B10** probable lipoprotein
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-D2** ferrichrome-iron receptor
+	+	X	X	X	X	X	X	X	X	X	X	**D23734-D4** transcriptional regulator
-	-	X	X	X	X	X	X	X	X	X	X	**D23734-G3** fimbrial subunit
-	-	X	X	X	X	X	X	X	X	X	X	**D21685** unknown
-	-	X	X	X	X	X	X	X	X	X	X	**D21685-3** Rom-like protein [E. coli]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-ST1** EaA [Enterobacteria phage P22]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-ST2** Tnp [Enterobacteria phage P7]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-H11** G protein [Enterobacteria phage 186]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-11** Tum [Enterobacteria phage 186]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-9** colicin E1 protein [Plasmid ColE1]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
-	-	X	X	X	X	X	X	X	X	X	X	**D26104-E11** AadA1 [E. coli]
+	+	+	+	+	+	+	+	+	+	-	-	**3114-11** putative phosphotransferase system component
-	-	+	+	+	+	+	-	+	+	-	-	**3114-18** peptidase S14, ClpP
-	-	+	+	+	+	+	-	+	+	-	-	**3114-A9** phage terminase GpA
-	-	+	+	+	-	+	-	+	+	-	-	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]
-	-	+	+	+	+	+	-	+	+	-	-	**3114-C12** putative outer membrane usher protein
-	-	+	+	+	+	+	-	+	+	-	-	**3114-F10** Gifsy-1 prophage terminase large chaing gp2
-	-	+	+	+	-	+	-	+	+	-	-	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]

NTS strains	DRC (Zaire)											
	1	2	3	4	5	6	6,2 previous 6SE	1	2	3	4
	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis
D23734-1 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-	-	-	-	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-	-	-	-	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-	-	-	-	-	-	-
D23734-A1 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-	-	-	-	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-	-	-	-	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-	-	-	-	-	-	-
D23734-D2 ferrichrome-iron receptor	-	-	-	-	-	-	-	+	+	+	+
D23734-D4 transcriptional regulator	-	-	-	-	-	-	-	-	-	-	-
D23734-G3 fimbrial subunit	-	-	-	-	-	-	-	+	+	+	+
D21685 unknown	-	-	-	-	-	-	-	-	-	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	-	-	-	-	-	-	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	+	+	+	+	+	-	+	-	-	-	-
D26104-ST2 Tnp [Enterobacteria phage P7]	-	+	-	+	+	-	+	+	-	-	-
D26104-H11 G protein [Enterobacteria phage 186]	-	+	-	+	+	-	+	-	-	-	-
D26104-11 Tum [Enterobacteria phage 186]	+	+	+	+	+	-	+	-	-	-	-
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	+	+	+	+	+	-	+	-	-	-	-
D26104-9 colicin E1 protein [Plasmid ColE1]	-	-	-	+	+	-	-	-	-	-	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	-	+	+	+	+	-	+	+	-	-	-
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	+	+	+	-	+	+	-	-	-
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+	+	+	+	-	+	+	-	-	-
D26104-E11 AadA1 [E. coli]	-	+	+	+	+	-	+	-	-	-	-
3114-11 putative phosphotransferase system component	-	-	-	-	-	-	-	+	+	+	+
3114-18 peptidase S14, ClpP	-	-	-	-	-	-	-	-	-	-	-
3114-A9 phage terminase GpA	-	-	-	-	-	-	-	-	-	-	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-	-	-	-	-	-	-
3114-C12 putative outer membrane usher protein	-	-	-	-	-	-	-	-	-	-	-
3114-F10 Gifsy-1 prophage terminase large chaing gp2	-	-	-	-	-	-	-	+	-	-	-
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-	-	-	-	-	-	-

NTS strains	5 S. Enteritidis	Kenya 9663 S. Typhimurium	9664 S. Typhimurium	9665 S. Typhimurium	9812 S. Typhimurium	417 S. Enteritidis	506 S. Enteritidis	705 S. Enteritidis	767 S. Enteritidis	UK human diarrhoea 37002 S. Typhimurium
D23734-1 hypothetical protein	-	-	-	-	-	-	-	-	-	-
D23734-2 hypothetical protein	-	-	-	-	-	-	-	-	-	-
D23734-4 2-nitropropane dioxygenase NPD	-	-	-	-	-	-	-	-	-	-
D23734-5 hypothetical protein	-	-	-	-	-	-	-	-	-	-
D23734-10 short-chain dehydrogenase/reductase SDR	-	-	-	-	-	-	-	-	-	-
D23734-14 putative autotransporter/pertactin	-	-	-	-	-	-	-	-	-	-
D23734-A1 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-
D23734-A5 putative fimbrial protein	-	-	-	-	-	-	-	-	-	-
D23734-B2 hypothetical protein	-	-	-	-	-	-	-	-	-	-
D23734-B9 conserved hypothetical protein	-	-	-	-	-	-	-	-	-	-
D23734-B10 probable lipoprotein	-	-	-	-	-	-	-	-	-	-
D23734-D2 ferrichrome-iron receptor	+	-	-	-	-	+	+	+	+	-
D23734-D4 transcriptional regulator	-	-	-	-	-	-	-	-	-	-
D23734-G3 fimbrial subunit	+	-	-	-	-	+	+	+	+	-
D21685 unknown	-	-	-	-	-	-	-	-	-	-
D21685-3 Rom-like protein [E. coli]	-	-	-	-	-	-	-	-	-	-
D26104-ST1 EaA [Enterobacteria phage P22]	-	-	-	-	-	-	-	-	-	-
D26104-ST2 Tnp [Enterobacteria phage P7]	+	+	+	-	-	-	+	+	-	-
D26104-H11 G protein [Enterobacteria phage 186]	-	+	+	-	+	-	-	-	-	-
D26104-11 Tum [Enterobacteria phage 186]	-	+	+	-	+	-	-	-	-	-
D26104-B6 gp10 [S. typhimurium bacteriophage ST104]	-	+	+	+	+	-	-	-	-	+
D26104-9 colicin E1 protein [Plasmid ColE1]	-	-	+	-	+	-	-	-	-	-
D26104-C6 TEM extended-spectrum beta-lactamase [E. coli]	+	+	+	-	-	-	+	+	-	+
D26104-C11 beta-lactamase TEM-160 [Proteus mirabilis]	+	+	+	-	-	-	+	+	-	+
D26104-D7 extended-spectrum beta-lactamase TEM-138 [S. enterica]	+	+?	-	-	-	-	+	+	-	+
D26104-E11 AadA1 [E. coli]	-	+	+	-	+	-	-	-	-	-
3114-11 putative phosphotransferase system component	+	-	-	-	-	+	+	+	+	-
3114-18 peptidase S14, ClpP	-	-	-	-	-	-	-	-	-	-
3114-A9 phage terminase GpA	-	-	-	-	-	-	-	-	-	-
3114-B6 putative tail fiber protein [Enterobacteria phage phiP27]	-	-	-	-	-	-	-	-	-	-
3114-C12 putative outer membrane usher protein	-	-	-	-	-	-	-	-	-	-
3114-F10 Gifsy-1 prophage terminase large chaing gp2	+	-	-	-	-	-	+	+	-	-
3114-G4 hypothetical protein [Stx2 converting bacteriophage]	-	-	-	-	-	-	-	-	-	-

various Tropical locations	NTS strains	67155	57285	57255	57190	57126	57115	57193	57315	57021	47014	37069	
		S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Enteritidis	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	S. Typhimurium	
		-	-	-	-	-	-	-	-	-	-	-	**D23734-1** hypothetical protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-2** hypothetical protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-4** 2-nitropropane dioxygenase NPD
		-	-	-	-	-	-	-	-	-	-	-	**D23734-5** hypothetical protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-10** short-chain dehydrogenase/reductase SDR
		-	-	-	-	-	-	-	-	-	-	-	**D23734-14** putative autotransporter/pertactin
		-	-	-	-	-	-	-	-	-	-	-	**D23734-A1** putative fimbrial protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-A5** putative fimbrial protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-B2** hypothetical protein
		-	-	-	-	-	-	-	-	-	-	-	**D23734-B9** conserved hypothetical protein
		+	-	-	-	-	-	-	-	-	-	-	**D23734-B10** probable lipoprotein
		+	+	+	+	+	+	-	-	-	-	-	**D23734-D2** ferrichrome-iron receptor
		+	-	-	-	-	-	-	-	-	-	-	**D23734-D4** transcriptional regulator
		+	+	+	+	+	+	-	-	-	-	-	**D23734-G3** fimbrial subunit
		-	-	-	-	-	-	-	-	-	-	-	**D21685** unknown
		-	-	-	-	-	-	-	-	-	-	-	**D21685-3** Rom-like protein [E. coli]
		-	-	-	-	-	-	-	-	-	-	-	**D26104-ST1** EaA [Enterobacteria phage P22]
		-	-	-	-	-	-	-	-	-	-	-	**D26104-ST2** Tnp [Enterobacteria phage P7]
		-	-	-	-	-	-	-	-	-	-	-	**D26104-H11** G protein [Enterobacteria phage 186]
		-	-	-	-	-	-	-	-	-	-	-	**D26104-11** Tum [Enterobacteria phage 186]
		-	-	-	-	-	-	-	-	-	+	+	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
		-	-	-	-	-	-	-	-	-	-	-	**D26104-9** colicin E1 protein [Plasmid ColE1]
		-	-	-	-	-	-	-	+	-	-	-	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
		-	-	-	-	-	-	-	+	-	-	-	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
		-	-	-	-	-	-	-	+	-	-	-	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
		-	-	-	-	-	-	-	-	+	-	+	**D26104-E11** AadA1 [E. coli]
		+	+	+	+	+	+	-	-	-	-	-	**3114-11** putative phosphotransferase system component
		-	-	-	-	-	-	-	-	-	-	-	**3114-18** peptidase S14, ClpP
		-	-	-	-	-	-	-	-	-	-	-	**3114-A9** phage terminase GpA
		-	-	-	-	-	-	-	-	-	-	-	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]
		-	-	-	-	-	-	-	-	-	-	-	**3114-C12** putative outer membrane usher protein
		-	-	-	-	-	-	-	-	-	-	-	**3114-F10** Gifsy-1 prophage terminase large chaing gp2
		-	-	-	-	-	-	-	-	-	-	-	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]

UK veterinary							NTS strains				
20070502	S.Heidelberg						20031619	S.Heidelberg			
+	+	+	+	+	-	+	+	+	+	**D23734-1** hypothetical protein	
+	+	+	+	+	+	+	+	+	+	**D23734-2** hypothetical protein	
+	+	+	+	+	+	+	+	+	+	**D23734-4** 2-nitropropane dioxygenase NPD	
+	+	+	+	+	+	+	+	+	+	**D23734-5** hypothetical protein	
+	+	+	+	+	+	+	+	+	+	**D23734-10** short-chain dehydrogenase/reductase SDR	
+	+	+	+	+	+	+	-	+	+	**D23734-14** putative autotransporter/pertactin	
+	-	+	+	+	+	-	+	-	-	**D23734-A1** putative fimbrial protein	
+	+	+	+	+	+	+	+	+	+	**D23734-A5** putative fimbrial protein	
+	+	+	+	+	+	+	+	+	+	**D23734-B2** hypothetical protein	
+	+	+	+	+	+	+	+	+	+	**D23734-B9** conserved hypothetical protein	
-	+	-	-	+	-	+	-	+	+	**D23734-B10** probable lipoprotein	
+	+	+	+	-	+	+	-	+	+	**D23734-D2** ferrichrome-iron receptor	
+	+	+	+	+	+	+	+	+	+	**D23734-D4** transcriptional regulator	
+	+	+	+	-	+	+	-	+	+	**D23734-G3** fimbrial subunit	
-	-	-	-	-	-	-	-	-	-	**D21685** unknown	
-	-	-	-	-	-	-	-	-	-	**D21685-3** Rom-like protein [E. coli]	
-	-	-	-	-	-	-	-	-	-	**D26104-ST1** EaA [Enterobacteria phage P22]	
-	-	-	-	-	-	-	-	-	-	**D26104-ST2** Tnp [Enterobacteria phage P7]	
-	-	-	-	-	-	-	-	-	-	**D26104-H11** G protein [Enterobacteria phage 186]	
-	-	-	-	-	-	-	-	-	-	**D26104-11** Tum [Enterobacteria phage 186]	
+	-	+	+	+	-	-	-	-	-	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]	
-	-	-	-	-	-	-	-	-	-	**D26104-9** colicin E1 protein [Plasmid ColE1]	
+	-	+	-	-	+	-	-	-	-	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]	
+	-	+	-	-	+	-	-	-	-	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]	
+	-	+	-	-	+	-	-	-	-	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]	
+	-	+	-	-	+	-	-	-	-	**D26104-E11** AadA1 [E. coli]	
-	-	-	-	-	+	-	-	-	-	**3114-11** putative phosphotransferase system component	
-	-	-	-	+	-	-	-	-	+	**3114-18** peptidase S14, ClpP	
-	-	-	-	+	-	-	-	-	+	**3114-A9** phage terminase GpA	
-	-	-	-	-	-	-	-	-	-	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]	
-	-	-	-	-	-	-	-	-	-	**3114-C12** putative outer membrane usher protein	
-	-	-	-	-	-	-	-	-	-	**3114-F10** Gifsy-1 prophage terminase large chaing gp2	
-	-	-	-	-	-	-	-	-	-	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]	

NTS strains							
S. Bovismorbificans NCTC5754	S. Bovismorbificans 5840/80	S. Bovismorbificans 518927/76	S. Bovismorbificans 544086/79	S. Heidelberg 17705	S. Heidelberg 55/366	S. Heidelberg KMS1977	
+	+	+	+	+	+	+	**D23734-1** hypothetical protein
-	-	-	-	+	+	+	**D23734-2** hypothetical protein
-	-	-	-	+	+	+	**D23734-4** 2-nitropropane dioxygenase NPD
-	-	-	-	+	+	+	**D23734-5** hypothetical protein
-	-	-	-	+	+	+	**D23734-10** short-chain dehydrogenase/reductase SDR
-	-	-	-	+	+	+	**D23734-14** putative autotransporter/pertactin
-	-	-	-	-	+	+	**D23734-A1** putative fimbrial protein
-	-	-	-	+	+	+	**D23734-A5** putative fimbrial protein
-	-	-	-	+	+	+	**D23734-B2** hypothetical protein
-	-	-	-	+	+	+	**D23734-B9** conserved hypothetical protein
+	+	+	+	+	+	+	**D23734-B10** probable lipoprotein
-	-	-	-	+	+	+	**D23734-D2** ferrichrome-iron receptor
-	-	-	+	+	+	+	**D23734-D4** transcriptional regulator
+	+	+	+	+	+	+	**D23734-G3** fimbrial subunit
-	-	-	-	-	-	-	**D21685** unknown
-	-	-	-	-	-	-	**D21685-3** Rom-like protein [E. coli]
-	-	-	-	-	-	+	**D26104-ST1** EaA [Enterobacteria phage P22]
-	-	-	-	-	-	-	**D26104-ST2** Tnp [Enterobacteria phage P7]
-	-	-	-	-	-	-	**D26104-H11** G protein [Enterobacteria phage 186]
-	-	-	-	-	-	-	**D26104-11** Tum [Enterobacteria phage 186]
-	-	-	-	-	-	+	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
-	-	-	-	-	-	-	**D26104-9** colicin E1 protein [Plasmid ColE1]
-	-	-	-	-	-	-	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
-	-	-	-	-	-	-	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
-	-	-	-	-	-	-	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
-	-	-	-	-	-	-	**D26104-E11** AadA1 [E. coli]
-	-	-	-	-	-	-	**3114-11** putative phosphotransferase system component
-	-	-	-	-	-	-	**3114-18** peptidase S14, ClpP
-	-	-	-	-	-	-	**3114-A9** phage terminase GpA
-	-	-	-	-	-	-	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]
+	+	+	+	-	-	-	**3114-C12** putative outer membrane usher protein
+	+	-	+	-	-	-	**3114-F10** Gifsy-1 prophage terminase large chaing gp2
-	-	-	-	-	-	-	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]

Table S2 Complete set of SSH distrubition BLAST searches of the NCBI database (September 2008)

genome sequenced NTS strains published on NCBI (2008)

SARA23	S. Saintpaul	CVM29188	S. Kentucky	CDC191	S. Kentucky	GA_MM04042433	S. Javiana	SL483	S. Agona	CT18	S. Typhi	Ty2	S. Typhi	SC-B67	S. Choleraesuis	SL486	S. Heidelberg	SL476	S. Heidelberg	
-	-	+	+	+	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	**D23734-1** hypothetical protein
-	-	+	+	+	+	-	-	+	+	-	-	-	-	-	-	+	+	+	+	**D23734-2** hypothetical protein
-	-	+	+	+	+	-	-	+	+	-	-	-	-	-	-	+	+	+	+	**D23734-4** 2-nitropropane dioxygenase NPD
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-5** hypothetical protein
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-10** short-chain dehydrogenase/reductase SDR
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-14** putative autotransporter/pertactin
-	-	+	+	+	+	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-A1** putative fimbrial protein
-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	**D23734-A5** putative fimbrial protein
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-B2** hypothetical protein
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-B9** conserved hypothetical protein
-	-	-	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	**D23734-B10** probable lipoprotein
-	-	+	+	+	+	-	-	+	+	-	-	-	-	-	-	+	+	+	+	**D23734-D2** ferrichrome-iron receptor
-	-	-	-	-	-	+	+	-	-	+	+	+	+	+	+	+	+	+	+	**D23734-D4** transcriptional regulator
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+	**D23734-G3** fimbrial subunit
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D21685** unknown
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D21685-3** Rom-like protein [E. coli]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-ST1** EaA [Enterobacteria phage P22]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-ST2** Tnp [Enterobacteria phage P7]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-H11** G protein [Enterobacteria phage 186]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-11** Tum [Enterobacteria phage 186]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-B6** gp10 [S. typhimurium bacteriophage ST104]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**D26104-9** colicin E1 protein [Plasmid ColE1]
-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-	-	**D26104-C6** TEM extended-spectrum beta-lactamase [E. coli]
-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	+	+	**D26104-C11** beta-lactamase TEM-160 [Proteus mirabilis]
-	-	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	-	+	+	**D26104-D7** extended-spectrum beta-lactamase TEM-138 [S. enterica]
-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	**D26104-E11** AadA1 [E. coli]
-	-	-	-	-	-	+	+	-	-	+	+	+	+	-	-	-	-	-	-	**3114-11** putative phosphotransferase system component
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-18** peptidase S14, ClpP
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-A9** phage terminase GpA
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-B6** putative tail fiber protein [Enterobacteria phage phiP27]
-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-C12** putative outer membrane usher protein
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-F10** Gifsy-1 prophage terminase large chaing gp2
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**3114-G4** hypothetical protein [Stx2 converting bacteriophage]

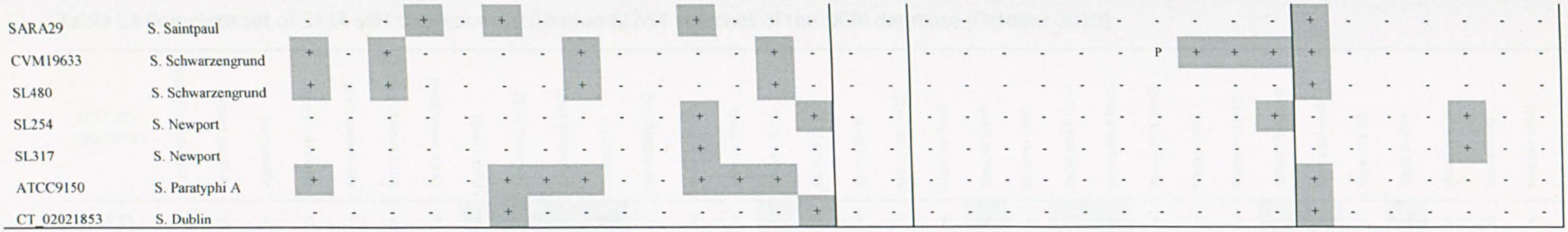

SARA29 S. Saintpaul
CVM19633 S. Schwarzengrund
SL480 S. Schwarzengrund
SL254 S. Newport
SL317 S. Newport
ATCC9150 S. Paratyphi A
CT_02021853 S. Dublin
P

Table S3 Complete set of 3114 SSH sequence distribution BLAST searches of the NCBI database (October 2010)

3114 SSH sequences	Typhimurium (n=10)	Arizonae (N=1)	Agona (n=1)	Diarizonae (n=2)	Abortusovis (n=1)	Brandenburg (N=1)	Choleraesuis (n=2)	Dublin (n=3)	Enteritidis (n=5)	Gallinarum (n=2)	Hadar (n=1)	Heidelberg (n=2)	Indiana (n=1)	Infantis (n=1)	Javiana (n=1)	Kentucky (n=2)	Miami (n=1)	München (n=1)	Münster (n=1)	Newport (n=4)	Panama (n=1)	Paratyphi A (n=2)	Paratyphi B (n=3)	Paratyphi C (n=2)	Poona (n=1)	Pullorum (n=1)	Schwarzengrund (n=3)	Saintpaul (n=2)	Tenessee (1)	Typhi (n=14)	Houtenae (1)	indica (n=2)	salamae (n=2)	Virchow (n=1)	Weltevreden (n=1)
3114-11	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	0	0	0	2	0	2	1	0	0	0	2	1	0	12	0	0	0	0	1
3114-18–ROD34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3114-A9 - ROD34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3114-B6 – ROD13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3114-G4-ROD13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3114-F10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3114-C12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table S4 Pseudogenes identfied in *S.* Typhimurium strains intact or absent from *S.* Bovismorbificans 3114

CDS in 3114		strand	location	D23580		SL1344		DT104		putative function	diabeling mutation
SBOV00191	Intact	-	18795 – 20066	STM_MW00741	Intact	SL0073	pseudo	SDT0074	Intact	L-carnitine dehydratase	
SBOV00581	Intact	+	63165 – 63287	STM_MW03251	Pseudo	SL0310	pseudo	SDT0316	pseudo	release factor H-coupled RctB family protein (pseudogene)	No frameshift, insertion has extra GCG GGG = AG where frameshift is in STM
SBOV01141	Intact	-	120357 – 121982	STM_MW01631	pseudo	SL0157	pseudo	SDT0163	Intact	putative exported protein (pseudogene)	Has HH present were frameshift is in STM
SBOV02581	Intact	+	2959481 – 260620	STM_MW03251	pseudo	SL0310	pseudo	SDT0316	pseudo	release factor H-coupled RctB family protein (pseudogene)	No frameshift in SBOV
SBOV03901	Intact	+	389189 – 390691	STM_MW05081	Intact	SL0431	Intact	SDT0479	pseudo	putative exported protein	
SBOV04551	Intact	-	456083 – 456910	STM_MW05671	Intact	SL0490	Intact	SDT0536	pseudo	probable secreted protein	Intact in SBOV
SBOV04761	Intact	+	476587 – 478368	STM_MW05871	Intact	SL0510	Intact	SDT0556	pseudo	Glyoxylate carboligase	Intact in SBOV
SBOV04801	Intact?	+	481545 – 482999	STM_MW05911	pseudo	SL0515	Intact	NP	NP	putative allantoin permease (pseudogene)	Appears intact in SBOV, check SL1344
SBOV06061	Intact	-	608539 – 609534	STM_MW07191	Intact	SL0642	Intact	SDT0678	pseudo	conserved hypothetical protein	
SBOV06091	Intact	+	611808 – 612503	STM_MW07221	pseudo	SL0645	pseudo	SDT0682	Intact	putative conserved hypothetical protein (pseudogene)	No frameshift in SBOV
SBOV07671	Intact	-	776596 – 778176	STM_MW08851	pseudo	SL0809	Intact	SDT0844	Intact	putative membrane protein (pseudogene)	No frameshift in SBOV, extra S=GCT at location of STM frameshift
SBOV07721	Intact	+	779225 – 780337	STM_MW08871	Intact	SL0811	Intact	SDT0846	pseudo	putative membrane protein	
SBOV08181	Intact	+	822414 – 823136	STM_MW09251	Intact	SL0850	pseudo	SDT0883	Intact	oxygen-insensitive NADPH nitroreductase	SBOV sequence is longer ?
SBOV08501	Intact	+	853635 – 855581	STM_MW09551	pseudo	SL0881	Intact	SDT0913	Intact	conserved hypothetical ABC transporter (pseudogene)	
SBOV09571	Intact	+	964583 – 965320	STM_MW10571	Intact	SL0985	Intact	SDT1018	pseudo	Minor phage tail protein (pseudogene)	Intact in SBOV
SBOV09591	Intact	+	965994 – 969344	STM_MW10591	Intact	SL0987	Intact	SDT1021	pseudo	host specificity tail protein J	Intact in SBOV
SBOV09641	Intact?	+	973379 – 974143	STM_MW10631	pseudo	SL0991	Intact	SDT1028	Intact	putative type III secreted protein (pseudogene)	Possibly intact, start missing due to poor seq
SBOV10071	Intact	-	1013841 – 1015403	STM_MW11041	pseudo	SL1033	Intact	SDT1071	Intact	putative secreted peptidase (pseudogene)	Appears intact in SBOV
SBOV12131	Intact	+	1188640 – 1189869	STM_MW12361	Intact	SL1164	Intact	SDT1202	pseudo	Peptidase T	Intact in both
SBOV12141	Intact	+	1189873 – 1191207	STM_MW12371	pseudo	SL1165	pseudo	SDT1204	Intact	conserved hypothetical protein (pseudogene)	Intact in SBOV
SBOV12591	Intact	+	1226651 – 1227013	STM_MW12771	Intact	SL1206	pseudo	SDT1245	Intact	conserved hypothetical protein	
SBOV13041	Intact	-	1264410 – 1265318	STM_MW13161	Intact	SL1244	Intact	SDT1281	pseudo	conserved hypothetical protein	Intact in both

SBOV13471	Intact	-	1302925 – 1304565	STM_MW13571	Intact	SL1284	Intact	SDT1323	pseudo	putative putative ligase/synthetase	
SBOV13591	Intact	-	1314087 – 1315322	STM_MW13671	Intact	SL1294	pseudo	SDT1334	Intact	Putative MFS-family transport protein	
SBOV15081	Intact	+	1457585 – 1458964	STM_MW15071	pseudo	SL1437	pseudo	SDT1473	pseudo	putative membrane transport protein (pseudogene)	
SBOV15201	Intact	-	1467452 – 1468639	STM_MW15161	pseudo	SL1446	Intact	SDT1482	Intact	putative membrane protein (pseudogene)	
SBOV1552	Intact	+	1497911 – 1498990	STM_MW15471	pseudo	SL1477	Intact	SDT1513	Intact	putative isomerase (pseudogene)	
SBOV15581	Intact	-	1502140 – 1502523	STM_MW15481	pseudo	SL1483	pseudo	SDT1519	pseudo	transposase (pseudogene)	
SBOV15611	Intact	-	1503364 – 1504830	STM_MW15491	Intact	SL1484	Intact	SDT1520	pseudo	conserved hypothetical protein	
SBOV15621	Intact	+	1505088 – 1506095	STM_MW15501	Intact	SL1485	Intact	SDT1521	pseudo	putative putative LacI-family transcriptional regulator	
SBOV16481	Intact	+	1599756 – 1601306	STM_MW16321	pseudo	SL1567	Intact	SDT1604	Intact	putative membrane protein (pseudogene)	
SBOV16861	Intact	+	1637254 – 1638105	STM_MW16661	pseudo	SL1601	pseudo	SDT1637	Intact	putative transcriptional regulator (pseudogene)	
SBOV19051	Intact?	+	1838141 – 1838509	STM_MW18531	pseudo	SL1791	pseudo	SDT1826	pseudo	phage integrase (pseudogene)	SBOV appears to be complete but is shorter than than D23580, fasta search 100% match
SBOV20011	Intact	-	1915873 – 1917306	STM_MW19221	pseudo	SL1873	Intact	SDT1951	Intact	putative exported protein (pseudogene)	
SBOV23061	Intact	-	2219849 – 2220640	STM_MW22741	pseudo	SL2219	pseudo	SDT2298	pseudo	Putative phage tail fibre protein (pseudogene)	Seems intact but is longer?
SBOV27551	Intact	+	2685350 – 2685955	STM_MW26941	pseudo	SL2653	Intact	SDT2724	Intact	conserved hypothetical protein (pseudogene)	
SBOV27881	Intact	-	2722035 – 2722658	STM_MW27312	pseudo	SL2747A	pseudo	SDT2762	pseudo	putative transposase (pseudogene)	
SBOV29601	Intact	-	2883377-2883655	STM_MW28951	pseudo	SL2911	Intact	SDT2926	Intact	conserved hypothetical protein (pseudogene)	
SBOV31331	Intact	+	3051270 – 3051926	STM_MW30361	pseudo	SL3051	Intact	SDT3068	Intact	possible ABC-cobalt transport protein, ATP-binding component (pseudogene)	
SBOV34201	Intact	-	3332700 – 3333599	STM_MW33531	pseudo	SL3327	Intact	SDT3343	Intact	tartrate dehydratase (pseudogene)	
SBOV35901	Intact	?	3488231 – 3488989	STM_MW35121	Intact	SL3489	pseudo	SDT3502	pseudo	glycerol-3-phosphate regulon repressor	
SBOV37001	Intact	+	3608936 – 3610615	STM_MW36131	pseudo	SL3589	Intact	SDT3603	Intact	putative endoglucanase (pseudogene)	
SBOV37151	Intact	-	3625703 – 3626782	STM_MW36271	pseudo	SL3602	pseudo	SDT3618	Intact	fimbrial protein (LpfD) (pseudogene)	
SBOV38581	Intact	-	3770140-3771249	STM_MW37571	pseudo	SL3733	Intact	SDT3748	Intact	putative transferase (pseudogene)	
SBOV38661	Intact	?	3778843 – 3780036	STM_MW37651	Intact	SL3741	pseudo	SDT3756	Intact	putative inner membrane transport protein	
SBOV39821	Intact	-	3894648 – 3896075	STM_MW38711	Intact	SL3854	pseudo	SDT3907	Intact	putative transmembrane efflux protein	
SBOV43791	Intact		4302776 – 4303684	STM_MW42371	pseudo	SL4223	Intact	SDT4274	Intact	putative membrane-bound beta-hydroxylase (pseudogene)	

SBOV44461	Intact?	-	4370167 – 4370886	STM_MW42991	pseudo	SL4288	Intact	SDT4336	pseudo	putative arginine-binding periplasmic protein	SBOV inatct but different start codon
SBOV44651	Intact	+	439408 – 4390106	STM_MW43151	Intact	SL4304	pseudo	SDT4352	Intact	conserved hypothetical protein	
SBOV45501	Intact	-	4471446 – 4472864	STM_MW43991	pseudo	SL4385	pseudo	SDT4436	pseudo	PTS system, trehalose-specific IIBC component (pseudogene)	
SBOV46521	Intact	+	4571784 – 4572557	STM_MW45061	pseudo	SL4491	Intact	SDT4543	pseudo	conserved hypothetical protein	
SBOV46601	Intact	-	4580271 – 4582679	STM_MW45141	Intact	SL4499	Intact	SDT4552	pseudo	fimbrial usher proteins stjB	
SBOV46861	Intact	?	4608593 – 4608907	STM_MW45401	pseudo	SL4523A	pseudo	SDT4577	pseudo	putative secreted protein (pseudogene)	

Pseudogenes absent from 3114

-	Absent		STM_MW00661	Pseudo	SL0066A	pseudo	SDT0066	pseudo	putative viral protein (pseudogene)	Absent from SBOV
	Absent		STM_MW01041	pseudo	SL0099A	pseudo	SDT0104	pseudo	putative inner membrane protein (pseudogene)	Absent from SBOV some sequence bits present, inverted
	Absent		STM_MW02981	pseudo	SL0286	pseudo	SDT0292	pseudo	Rhs-family protein fragment	Absent from SBOV
	Absent		STM_MW03021	pseudo	SL0289	pseudo	SDT0294	pseudo	putative phosphotriesterase (pseudogene)	Absent from SBOV
-	Absent		STM_MW05071	pseudo	SL0430A	pseudo	SDT0478	pseudo	putative transposase (fragment)	Absent from SBOV, very short match at N-and C-terminal
-	Absent		STM_MW07761	Intact	SL0700	pseudo	SDT0736	Intact	putative udp-galactopyranose mutase	Absent from SBOV
-	Absent		STM_MW10251	pseudo	SL0952	Intact	SDT0985	Intact	Gifsy-2 prophage replication Protein O (pseudogene)	Absent from SBOV
-	Absent		STM_MW10351	pseudo	SL0962	Intact	SDT0995	Intact	conserved prophage Protein (pseudogene)	Absent from SBOV
-	Absent		STM_MW10621	pseudo	SL0990	pseudo	SDT1026	pseudo	putative tail fibre assembly protein (remnant)	Absent from SBOV
-	Absent		STM_MW18491	Intact	SL1788	pseudo	SDT1822	Intact	conserved hypothetical protein	
-	Absent		STM_MW18511	Intact	SL1789	pseudo	SDT1823	Intact	conserved hypothetical protein	
-	Absent		STM_MW18771	pseudo	SL1830	pseudo	SDT1907	Intact	conserved hypothetical protein (pseudogene)	
-	Absent		STM_MW18791	pseudo	SL1830A	pseudo	SDT1908	pseudo	conserved hypothetical protein (pseudogene)	
-	Absent		STM_MW19232	pseudo	SL1874A	pseudo	SDT1952A	pseudo	conserved hypothetical protein (pseudogene)	
-	Absent		STM_MW19811	pseudo	SL1927B	pseudo	NP	Intact	phage DNA invertase	
-	Absent		STM_MW20961	pseudo	SL2042	pseudo	SDT2118A	NP	conserved hypothetical protein (pseudogene)	
-	Absent		STM_MW22681	pseudo	SL2214	Intact	SDT2292	Intact	putative phage protein (pseudogene)	
-	Absent		STM_MW27071	pseudo	SL2723	pseudo	SDT2737	pseudo	prophage integrase (pseudogene)	

-	Absent	STM_MW27131	Intact	SL2730	Intact	SDT2743	pseudo	putative ATPase
-	Absent	STM_MW27311	pseudo	SL2747	pseudo	SDT2761	pseudo	phage integrase protein (pseudogene)
-	Absent	STM_MW37341	pseudo	SL3711	Intact	SDT3725	Intact	conserved hypothetical protein (pseudogene)
-	Absent	STM_MW37411	pseudo	SL3718	pseudo	SDT3732	pseudo	putative ATP binding protein (pseudogene)
-	Absent	STM_MW38291	pseudo	SL3811	pseudo	SDT3865	pseudo	phage integrase (pseudogene)
-	Absent	STM_MW41451	pseudo	SL4131	Intact	SDT4182	Intact	hypothetical protein (pseudogene)
-	Absent	STM_MW43821	pseudo	SL4369A	pseudo	SDT4419	pseudo	conserved hypothetical protein (pseudogene)
-	Absent	STM_MW44331	pseudo	SL4417	pseudo	SDT4470	pseudo	putative integrase (fragment)
-	Absent	STM_MW44611	pseudo	SL4445A	pseudo	SDT4498	pseudo	putative membrane protein (pseudogene)

8. References

Achtman, M., Willetts, N. & Clark, A. J. (1972). Conjugational complementation analysis of transfer-deficient mutants of Flac in Escherichia coli. *J Bacteriol* **110**, 831-842.

Achtman, M., Kennedy, N. & Skurray, R. (1977). Cell--cell interactions in conjugating Escherichia coli: role of traT protein in surface exclusion. *Proc Natl Acad Sci U S A* **74**, 5104-5108.

Adejuyigbe, E. A., Ako-Nai, A. K. & Adisa, B. (2004). Bacterial isolates in the sick young infant in Ile-Ife, Nigeria. *Journal of tropical pediatrics* **50**, 323-327.

Afifi, S., Earhart, K., Azab, M. A. & other authors (2005). Hospital-based surveillance for acute febrile illness in Egypt: a focus on community-acquired bloodstream infections. *The American journal of tropical medicine and hygiene* **73**, 392-399.

Agron, P. G., Walker, R. L., Kinde, H., Sawyer, S. J., Hayes, D. C., Wollard, J. & Andersen, G. L. (2001). Identification by subtractive hybridization of sequences specific for Salmonella enterica serovar enteritidis. *Appl Environ Microbiol* **67**, 4984-4991.

Ahmed, A. M., Nakano, H. & Shimamoto, T. (2005). Molecular characterization of integrons in non-typhoid Salmonella serovars isolated in Japan: description of an unusual class 2 integron. *The Journal of antimicrobial chemotherapy* **55**, 371-374.

Akopyants, N. S., Fradkov, A., Diatchenko, L., Hill, J. E., Siebert, P. D., Lukyanov, S. A., Sverdlov, E. D. & Berg, D. E. (1998). PCR-based subtractive hybridization and differences in gene content among strains of Helicobacter pylori. *Proc Natl Acad Sci U S A* **95**, 13108-13113.

Allmeier, H., Cresnar, B., Greck, M. & Schmitt, R. (1992). Complete nucleotide sequence of Tn1721: gene organization and a novel gene product with features of a chemotaxis protein. *Gene* **111**, 11-20.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* **215**, 403-410.

Amavisit, P., Lightfoot, D., Browning, G. F. & Markham, P. F. (2003). Variation between pathogenic serovars within Salmonella pathogenicity islands. *J Bacteriol* **185**, 3624-3635.

Andersson, J. O. & Andersson, S. G. (2001). Pseudogenes, junk DNA, and the dynamics of Rickettsia genomes. *Molecular biology and evolution* **18**, 829-839.

Andersson, S. G., Zomorodipour, A., Andersson, J. O. & other authors (1998). The genome sequence of Rickettsia prowazekii and the origin of mitochondria. *Nature* **396**, 133-140.

Angrist, A. & Mollov, M. (1946). Bacteriologic, clinical and pathologic experience with 86 sporadic cases of Salmonella infection. *American Journal of Medical Sciences* **212**, 10.

Archibald, S. C., Sunderland, D., Ashford, R. W. & Hart, C. A. (1989). Bacterial Infections in patients with the hyperinfection syndrome: interactions between strongyloides larvae and bacteria. *Biochemical Society Transactions* **17**, 2.

Atkinson, N., Woodroofe, G. M. & Macbeth, A. M. (1944). The occurence of Salmonella types in Australia 2. *Australian Journal of Experimental Biology and Medical Science* **22**, 3.

Atkinson, N., Woodroofe, G. M. & Macbeth, A. M. (1947). The occurence of Salmonella types in Australia 3. *Australian Journal of Experimental Biology and Medical Science* **25**, 6.

Bachou, H., Tylleskar, T., Kaddu-Mulindwa, D. H. & Tumwine, J. K. (2006). Bacteraemia among severely malnourished children infected and uninfected with the human immunodeficiency virus-1 in Kampala, Uganda. *BMC infectious diseases* **6**, 160.

Bae, J. W., Rhee, S. K., Nam, Y. D. & Park, Y. H. (2005). Generation of subspecies level-specific microbial diagnostic microarrays using genes amplified from subtractive suppression hybridization as microarray probes. *Nucleic acids research* **33**, e113.

Bailey, M. J., Hughes, C. & Koronakis, V. (1997). RfaH and the ops element, components of a novel system controlling bacterial transcription elongation. *Molecular microbiology* **26**, 845-851.

Baird, G. D., Manning, E. J. & Jones, P. W. (1985). Evidence for related virulence sequences in plasmids of Salmonella dublin and Salmonella typhimurium. *Journal of general microbiology* **131**, 1815-1823.

Barrow, P. A. & Lovell, M. A. (1989). Invasion of Vero cells by Salmonella species. *J Med Microbiol* **28**, 59-67.

Barrow, P. A., Huggins, M. B. & Lovell, M. A. (1994). Host specificity of Salmonella infection in chickens and mice is expressed in vivo primarily at the level of the reticuloendothelial system. *Infect Immun* **62**, 4602-4610.

Baumler, A. J., Tsolis, R. M., Bowe, F. A., Kusters, J. G., Hoffmann, S. & Heffron, F. (1996a). The pef fimbrial operon of Salmonella typhimurium mediates adhesion to murine small intestine and is necessary for fluid accumulation in the infant mouse. *Infect Immun* **64**, 61-68.

Baumler, A. J., Tsolis, R. M. & Heffron, F. (1996b). The lpf fimbrial operon mediates adhesion of Salmonella typhimurium to murine Peyer's patches. *Proc Natl Acad Sci U S A* **93**, 279-283.

Baumler, A. J. (1997). The record of horizontal gene transfer in Salmonella. *Trends in microbiology* **5**, 318-322.

Baumler, A. J., Hargis, B. M. & Tsolis, R. M. (2000). Tracing the origins of Salmonella outbreaks. *Science (New York, NY* **287**, 50-52.

Bautz, E. K. & Reilly, E. (1966). Gene-specific messenger RNA: isolation by the deletion method. *Science (New York, NY* **151**, 328-330.

Bell, M., Archibald, L. K., Nwanyanwu, O., Dobbie, H., Tokars, J., Kazembe, P. N., Reller, L. B. & Jarvis, W. R. (2001). Seasonal variation in the etiology of bloodstream infections in a febrile inpatient population in a developing country. *Int J Infect Dis* **5**, 63-69.

Berkley, J. A., Lowe, B. S., Mwangi, I. & other authors (2005). Bacteremia among children admitted to a rural hospital in Kenya. *N Engl J Med* **352**, 39-47.

Beuzon, C. R., Banks, G., Deiwick, J., Hensel, M. & Holden, D. W. (1999). pH-dependent secretion of SseB, a product of the SPI-2 type III secretion system of Salmonella typhimurium. *Molecular microbiology* **33**, 806-816.

Beuzon, C. R., Meresse, S., Unsworth, K. E., Ruiz-Albert, J., Garvis, S., Waterman, S. R., Ryder, T. A., Boucrot, E. & Holden, D. W. (2000). Salmonella maintains the integrity of its intracellular vacuole through the action of SifA. *The EMBO journal* **19**, 3235-3249.

Bignell, C. & Thomas, C. M. (2001). The bacterial ParA-ParB partitioning proteins. *Journal of biotechnology* **91**, 1-34.

Blanc-Potard, A. B., Solomon, F., Kayser, J. & Groisman, E. A. (1999). The SPI-3 pathogenicity island of Salmonella enterica. *J Bacteriol* **181**, 998-1004.

Bogush, M. L., Velikodvorskaya, T. V., Lebedev, Y. B. & other authors (1999). Identification and localization of differences between Escherichia coli and Salmonella typhimurium genomes by suppressive subtractive hybridization. *Mol Gen Genet* **262**, 721-729.

Boyd, E. F., Wang, F. S., Whittam, T. S. & Selander, R. K. (1996). Molecular genetic relationships of the salmonellae. *Appl Environ Microbiol* **62**, 804-808.

Boyd, E. F. & Hartl, D. L. (1997). Recent horizontal transmission of plasmids between natural populations of Escherichia coli and Salmonella enterica. *J Bacteriol* **179**, 1622-1627.

Boyd, E. F. & Hartl, D. L. (1998). Salmonella virulence plasmid. Modular acquisition of the spv virulence region by an F-plasmid in Salmonella enterica subspecies I and insertion into the chromosome of subspecies II, IIIa, IV and VII isolates. *Genetics* **149**, 1183-1190.

Boyd, J. S. (1950). The symbiotic bacteriophages of Salmonella typhi-murium. *The Journal of pathology and bacteriology* **62**, 501-517.

Brennan, M. A. & Cookson, B. T. (2000). Salmonella induces macrophage death by caspase-1-dependent necrosis. *Molecular microbiology* **38**, 31-40.

Brenner, F. W., Villar, R. G., Angulo, F. J., Tauxe, R. & Swaminathan, B. (2000). Salmonella nomenclature. *J Clin Microbiol* **38**, 2465-2467.

Brent, A. J., Oundo, J. O., Mwangi, I., Ochola, L., Lowe, B. & Berkley, J. A. (2006). Salmonella bacteremia in Kenyan children. *Pediatr Infect Dis J* **25**, 230-236.

Bronowski, C., Smith, S. L., Yokota, K., Corkill, J. E., Martin, H. M., Campbell, B. J., Rhodes, J. M., Hart, C. A. & Winstanley, C. (2008). A subset of mucosa-associated Escherichia coli isolates from patients with colon cancer, but not Crohn's disease, share pathogenicity islands with urinary pathogenic E. coli. *Microbiology* **154**, 571-583.

Caldon, C. E., Yoong, P. & March, P. E. (2001). Evolution of a molecular switch: universal bacterial GTPases regulate ribosome function. *Molecular microbiology* **41**, 289-297.

Calis, J. C., Phiri, K. S., Faragher, E. B. & other authors (2008). Severe anemia in Malawian children. *N Engl J Med* **358**, 888-899.

Campbell, A., Schneider, S. J. & Song, B. (1992). Lambdoid phages as elements of bacterial genomes (integrase/phage21/Escherichia coli K-12/icd gene). *Genetica* **86**, 259-267.

Carlsen, E., Flatmark, A. & Prydz, H. (1988). Cytokine-induced procoagulant activity in monocytes and endothelial cells. Further enhancement by cyclosporine. *Transplantation* **46**, 575-580.

Carnell, S. C., Bowen, A., Morgan, E., Maskell, D. J., Wallis, T. S. & Stevens, M. P. (2007). Role in virulence and protective efficacy in pigs of Salmonella enterica serovar Typhimurium secreted components identified by signature-tagged mutagenesis. *Microbiology* **153**, 1940-1952.

Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics (Oxford, England)* **25**, 119-120.

Cavaco, L. M., Hasman, H., Xia, S. & Aarestrup, F. M. (2009). qnrD, a novel gene conferring transferable quinolone resistance in Salmonella enterica serovar Kentucky and Bovismorbificans strains of human origin. *Antimicrobial agents and chemotherapy* **53**, 603-608.

CDC (1985). Revision of the case definition of AIDS for national reporting - United States. *MMWR CDC Surveil Summ* **34**, 2.

Cerin, H. & Hackett, J. (1993). The parVP region of the Salmonella typhimurium virulence plasmid pSLT contains four loci required for incompatibility and partition. *Plasmid* **30**, 30-38.

Chappell, L., Kaiser, P., Barrow, P., Jones, M. A., Johnston, C. & Wigley, P. (2009). The immunobiology of avian systemic salmonellosis. *Vet Immunol Immunopathol* **128**, 53-59.

Charles, I. G., Dougan, G., Pickard, D., Chatfield, S., Smith, M., Novotny, P., Morrissey, P. & Fairweather, N. F. (1989). Molecular cloning and characterization of protective outer membrane protein P.69 from Bordetella pertussis. *Proc Natl Acad Sci U S A* **86**, 3554-3558.

Cheesbrough, J. S., Taxman, B. C., Green, S. D., Mewa, F. I. & Numbi, A. (1997). Clinical definition for invasive Salmonella infection in African children. *Pediatr Infect Dis J* **16**, 277-283.

Chen, C. Y., Nace, G. W., Solow, B. & Fratamico, P. (2007). Complete nucleotide sequences of 84.5- and 3.2-kb plasmids in the multi-antibiotic resistant Salmonella enterica serovar Typhimurium U302 strain G8430. *Plasmid* **57**, 29-43.

Chiou, C. S., Huang, J. F., Tsai, L. H., Hsu, K. M., Liao, C. S. & Chang, H. L. (2006). A simple and low-cost paper-bridged method for Salmonella phase reversal. *Diagn Microbiol Infect Dis* **54**, 315-317.

Chiu, C. H., Su, L. H. & Chu, C. (2004). Salmonella enterica serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. *Clinical microbiology reviews* **17**, 311-322.

Chiu, C. H., Tang, P., Chu, C., Hu, S., Bao, Q., Yu, J., Chou, Y. Y., Wang, H. S. & Lee, Y. S. (2005). The genome sequence of Salmonella enterica serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic acids research* **33**, 1690-1698.

Cirillo, D. M., Valdivia, R. H., Monack, D. M. & Falkow, S. (1998). Macrophage-dependent induction of the Salmonella pathogenicity island 2 type III secretion system and its role in intracellular survival. *Molecular microbiology* **30**, 175-188.

Clayton, D. J., Bowen, A. J., Hulme, S. D. & other authors (2008). Analysis of the role of 13 major fimbrial subunits in colonisation of the chicken intestines by Salmonella enterica serovar Enteritidis reveals a role for a novel locus. *BMC Microbiol* **8**, 228.

Cogan, T. A., Jorgensen, F., Lappin-Scott, H. M., Benson, C. E., Woodward, M. J. & Humphrey, T. J. (2004). Flagella and curli fimbriae are important for the growth of Salmonella enterica serovars in hen eggs. *Microbiology* **150**, 1063-1071.

Cole, S. T., Eiglmeier, K., Parkhill, J. & other authors (2001). Massive gene decay in the leprosy bacillus. *Nature* **409**, 1007-1011.

Collazo, C. M. & Galan, J. E. (1997a). The invasion-associated type III system of Salmonella typhimurium directs the translocation of Sip proteins into the host cell. *Molecular microbiology* **24**, 747-756.

Collazo, C. M. & Galan, J. E. (1997b). The invasion-associated type-III protein secretion system in Salmonella--a review. *Gene* **192**, 51-59.

Collinson, S. K., Emody, L., Muller, K. H., Trust, T. J. & Kay, W. W. (1991). Purification and characterization of thin, aggregative fimbriae from Salmonella enteritidis. *J Bacteriol* **173**, 4773-4781.

Collinson, S. K., Clouthier, S. C., Doran, J. L., Banser, P. A. & Kay, W. W. (1996). Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae. *J Bacteriol* **178**, 662-667.

Crosa, J. H., Brenner, D. J., Ewing, W. H. & Falkow, S. (1973). Molecular relationships among the Salmonelleae. *J Bacteriol* **115**, 307-315.

Crump, J. A., Luby, S. P. & Mintz, E. D. (2004). The global burden of typhoid fever. *Bulletin of the World Health Organization* **82**, 346-353.

Darwin, K. H. & Miller, V. L. (1999). Molecular basis of the interaction of Salmonella with the intestinal mucosa. *Clinical microbiology reviews* **12**, 405-428.

De Buck, J., Van Immerseel, F., Haesebrouck, F. & Ducatelle, R. (2004). Effect of type 1 fimbriae of Salmonella enterica serotype Enteritidis on bacteraemia and reproductive tract infection in laying hens. *Avian Pathol* **33**, 314-320.

Deiwick, J. & Hensel, M. (1999). Regulation of virulence genes by environmental signals in Salmonella typhimurium. *Electrophoresis* **20**, 813-817.

Deiwick, J., Nikolaus, T., Erdogan, S. & Hensel, M. (1999). Environmental regulation of Salmonella pathogenicity island 2 gene expression. *Molecular microbiology* **31**, 1759-1773.

Dhanoa, A. & Fatt, Q. K. (2009). Non-typhoidal Salmonella bacteraemia: epidemiology, clinical characteristics and its' association with severe immunosuppression. *Annals of clinical microbiology and antimicrobials* **8**, 15.

Diatchenko, L., Lau, Y. F., Campbell, A. P. & other authors (1996). Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci U S A* **93**, 6025-6030.

Donnenberg, M. S. (2000). Pathogenic strategies of enteric bacteria. *Nature* **406**, 768-774.

Dougle, M. L., Hendriks, E. R., Sanders, E. J. & Dorigo-Zetsma, J. W. (1997). Laboratory investigations in the diagnosis of septicaemia and malaria. *East African medical journal* **74**, 353-356.

Dowling, J. J., Whitty, C. J., Chaponda, M., Munthali, C., Zijlstra, E. E., Gilks, C. F., Squire, S. B. & Gordon, M. A. (2002). Are intestinal helminths a risk factor for non-typhoidal Salmonella bacteraemia in adults in Africa who are seropositive for HIV? A case-control study. *Annals of tropical medicine and parasitology* **96**, 203-208.

Edwards, R. A., Schifferli, D. M. & Maloy, S. R. (2000). A role for Salmonella fimbriae in intraperitoneal infections. *Proc Natl Acad Sci U S A* **97**, 1258-1262.

Edwards, R. A., Olsen, G. J. & Maloy, S. R. (2002). Comparative genomics of closely related salmonellae. *Trends in microbiology* **10**, 94-99.

Emmerth, M., Goebel, W., Miller, S. I. & Hueck, C. J. (1999). Genomic subtraction identifies Salmonella typhimurium prophages, F-related plasmid sequences, and a novel fimbrial operon, stf, which are absent in Salmonella typhi. *J Bacteriol* **181**, 5652-5661.

Engelberts, I., von Asmuth, E. J., van der Linden, C. J. & Buurman, W. A. (1991). The interrelation between TNF, IL-6, and PAF secretion induced by LPS in an in vivo and in vitro murine model. *Lymphokine Cytokine Res* **10**, 127-131.

Enwere, G., Biney, E., Cheung, Y. B. & other authors (2006). Epidemiologic and clinical characteristics of community-acquired invasive bacterial infections in children aged 2-29 months in The Gambia. *Pediatr Infect Dis J* **25**, 700-705.

Ernst, R. K., Guina, T. & Miller, S. I. (2001). Salmonella typhimurium outer membrane remodeling: role in resistance to host innate immunity. *Microbes Infect* **3**, 1327-1334.

Fagan, R. P. & Smith, S. G. (2007). The Hek outer membrane protein of Escherichia coli is an auto-aggregating adhesin and invasin. *FEMS Microbiol Lett* **269**, 248-255.

Fagan, R. P., Lambert, M. A. & Smith, S. G. (2008). The hek outer membrane protein of Escherichia coli strain RS218 binds to proteoglycan and utilizes a single extracellular loop for adherence, invasion, and autoaggregation. *Infect Immun* **76**, 1135-1142.

Fang, F. C., Libby, S. J., Buchmeier, N. A., Loewen, P. C., Switala, J., Harwood, J. & Guiney, D. G. (1992). The alternative sigma factor katF (rpoS) regulates Salmonella virulence. *Proc Natl Acad Sci U S A* **89**, 11978-11982.

Feulner, G., Gray, J. A., Kirschman, J. A., Lehner, A. F., Sadosky, A. B., Vlazny, D. A., Zhang, J., Zhao, S. & Hill, C. W. (1990). Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhsA with other members of the rhs multigene family. *J Bacteriol* **172**, 446-456.

Fierer, J., Krause, M., Tauxe, R. & Guiney, D. (1992). Salmonella typhimurium bacteremia: association with the virulence plasmid. *J Infect Dis* **166**, 639-642.

Figueroa-Bossi, N., Coissac, E., Netter, P. & Bossi, L. (1997). Unsuspected prophage-like elements in Salmonella typhimurium. *Molecular microbiology* **25**, 161-173.

Figueroa-Bossi, N., Uzzau, S., Maloriol, D. & Bossi, L. (2001). Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in Salmonella. *Molecular microbiology* **39**, 260-271.

Finley, R., Reid-Smith, R., Ribble, C., Popa, M., Vandermeer, M. & Aramini, J. (2008). The occurrence and anti-microbial susceptibility of Salmonellae isolated from commercially available pig ear pet treats. *Zoonoses and public health* **55**, 455-461.

Fisher, I. S. (2004). International trends in salmonella serotypes 1998-2003--a surveillance report from the Enter-net international surveillance network. *Euro Surveill* **9**, 45-47.

Fleckenstein, J. M., Kopecko, D. J., Warren, R. L. & Elsinghorst, E. A. (1996). Molecular characterization of the tia invasion locus from enterotoxigenic Escherichia coli. *Infect Immun* **64**, 2256-2265.

Folkesson, A., Advani, A., Sukupolvi, S., Pfeifer, J. D., Normark, S. & Lofdahl, S. (1999). Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease. *Molecular microbiology* **33**, 612-622.

Foultier, B., Troisfontaines, P., Muller, S., Opperdoes, F. R. & Cornelis, G. R. (2002). Characterization of the ysa pathogenicity locus in the chromosome of Yersinia enterocolitica and phylogeny analysis of type III secretion systems. *Journal of molecular evolution* **55**, 37-51.

Fowler, T., Taylor, L. & Thompson, R. (1983). The control region of the F plasmid transfer operon: DNA sequence of the traJ and traY genes and characterisation of the traY leads to Z promoter. *Gene* **26**, 79-89.

Friedrich, M. J., Kinsey, N. E., Vila, J. & Kadner, R. J. (1993). Nucleotide sequence of a 13.9 kb segment of the 90 kb virulence plasmid of Salmonella typhimurium: the presence of fimbrial biosynthetic genes. *Molecular microbiology* **8**, 543-558.

Furuchi, T., Kashiwagi, K., Kobayashi, H. & Igarashi, K. (1991). Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the Escherichia coli chromosome. *The Journal of biological chemistry* **266**, 20928-20933.

Galan, J. E. & Curtiss, R., 3rd (1989). Cloning and molecular characterization of genes whose products allow Salmonella typhimurium to penetrate tissue culture cells. *Proc Natl Acad Sci U S A* **86**, 6383-6387.

Galofre, J., Moreno, A., Mensa, J. & other authors (1994). Analysis of factors influencing the outcome and development of septic metastasis or relapse in Salmonella bacteremia. *Clin Infect Dis* **18**, 873-878.

Garcia Vescovi, E., Soncini, F. C. & Groisman, E. A. (1996). Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* **84**, 165-174.

Garvis, S. G., Beuzon, C. R. & Holden, D. W. (2001). A role for the PhoP/Q regulon in inhibition of fusion between lysosomes and Salmonella-containing vacuoles in macrophages. *Cellular microbiology* **3**, 731-744.

Gasem, M. H., Dolmans, W. M., Keuter, M. M. & Djokomoeljanto, R. R. (2001). Poor food hygiene and housing as risk factors for typhoid fever in Semarang, Indonesia. *Trop Med Int Health* **6**, 484-490.

Gendrel, D., Kombila, M., Beaudoin-Leblevec, G. & Richard-Lenoble, D. (1994). Nontyphoidal salmonellal septicemia in Gabonese children infected with Schistosoma intercalatum. *Clin Infect Dis* **18**, 103-105.

Giannella, R. A., Broitman, S. A. & Zamcheck, N. (1971). Salmonella enteritis. I. Role of reduced gastric secretion in pathogenesis. *Am J Dig Dis* **16**, 1000-1006.

Gibson, E. A. (1961). Salmonellosis in calves. *The Veterinary record* **73**, 1284-1296.

Gilks, C. F., Brindle, R. J., Otieno, L. S. & other authors (1990). Life-threatening bacteraemia in HIV-1 seropositive adults admitted to hospital in Nairobi, Kenya. *Lancet* **336**, 545-549.

Gilsdorf, A., Jansen, A., Alpers, K. & other authors (2005). A nationwide outbreak of Salmonella Bovismorbificans PT24, Germany, December 2004-March 2005. *Euro Surveill* **10**, E050324 050321.

Ginocchio, C. C., Olmsted, S. B., Wells, C. L. & Galan, J. E. (1994). Contact with epithelial cells induces the formation of surface appendages on Salmonella typhimurium. *Cell* **76**, 717-724.

Gordon, M. A., Walsh, A. L., Chaponda, M., Soko, D., Mbvwinji, M., Molyneux, M. E. & Gordon, S. B. (2001). Bacteraemia and mortality among adult medical admissions in Malawi--predominance of non-typhi salmonellae and Streptococcus pneumoniae. *J Infect* **42**, 44-49.

Gordon, M. A., Banda, H. T., Gondwe, M. & other authors (2002). Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. *AIDS (London, England)* **16**, 1633-1641.

Gordon, M. A. (2008). Salmonella infections in immunocompromised adults. *J Infect* **56**, 413-422.

Gordon, M. A. & Graham, S. M. (2008). Invasive salmonellosis in Malawi. *Journal of infection in developing countries* **2**, 438-442.

Gordon, M. A., Graham, S. M., Walsh, A. L. & other authors (2008). Epidemics of invasive Salmonella enterica serovar enteritidis and S. enterica Serovar typhimurium infection associated with multidrug resistance among adults and children in Malawi. *Clin Infect Dis* **46**, 963-969.

Graham, J. P., Boland, J. J. & Silbergeld, E. (2007). Growth promoting antibiotics in food animal production: an economic analysis. *Public Health Rep* **122**, 79-87.

Graham, S. M., Molyneux, E. M., Walsh, A. L., Cheesbrough, J. S., Molyneux, M. E. & Hart, C. A. (2000). Nontyphoidal Salmonella infections of children in tropical Africa. *Pediatr Infect Dis J* **19**, 1189-1196.

Graham, S. M. (2002). Salmonellosis in children in developing and developed countries and populations. *Curr Opin Infect Dis* **15**, 507-512.

Green, S. D. & Cheesbrough, J. S. (1993). Salmonella bacteraemia among young children at a rural hospital in western Zaire. *Ann Trop Paediatr* **13**, 45-53.

Gregorova, D., Matiasovicova, J., Sebkova, A., Faldynova, M. & Rychlik, I. (2004). Salmonella enterica subsp. enterica serovar Enteritidis harbours ColE1, ColE2, and rolling-circle-like replicating plasmids. *Canadian journal of microbiology* **50**, 107-112.

Groisman, E. A. & Ochman, H. (1993). Cognate gene clusters govern invasion of host epithelial cells by Salmonella typhimurium and Shigella flexneri. *The EMBO journal* **12**, 3779-3787.

Groisman, E. A. & Ochman, H. (1996). Pathogenicity islands: bacterial evolution in quantum leaps. *Cell* **87**, 791-794.

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

Gulig, P. A. (1990). Virulence plasmids of Salmonella typhimurium and other salmonellae. *Microb Pathog* **8**, 3-11.

Gulig, P. A., Danbara, H., Guiney, D. G., Lax, A. J., Norel, F. & Rhen, M. (1993). Molecular analysis of spv virulence genes of the Salmonella virulence plasmids. *Molecular microbiology* **7**, 825-830.

Haines & Wilson (1957). The bacteriology of spray-dried egg, with particular reference to food poisoning. *Special Report Series Medical Research Council, London* **260**.

Hale, C., Bowe, F., Pickard, D., Clare, S., Haeuw, J. F., Powers, U., Menager, N., Mastroeni, P. & Dougan, G. (2006). Evaluation of a novel Vi conjugate vaccine in a murine model of salmonellosis. *Vaccine* **24**, 4312-4320.

Hancox, L. S., Yeh, K. S. & Clegg, S. (1997). Construction and characterization of type 1 non-fimbriate and non-adhesive mutants of Salmonella typhimurium. *FEMS Immunol Med Microbiol* **19**, 289-296.

Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galan, J. E. (1998). S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815-826.

Hayes, W. & Freeman, J. F. (1945). The incidence, type and bacteriology of Salmonella infections in the Army in India. *Indian Journal of Medical Research* **33**, 16.

Hayward, R. D. & Koronakis, V. (1999). Direct nucleation and bundling of actin by the SipC protein of invasive Salmonella. *The EMBO journal* **18**, 4926-4934.

Heffernan, E. J., Harwood, J., Fierer, J. & Guiney, D. (1992). The Salmonella typhimurium virulence plasmid complement resistance gene rck is homologous to a family of virulence-related outer membrane protein genes, including pagC and ail. *J Bacteriol* **174**, 84-91.

Heithoff, D. M., Conner, C. P., Hanna, P. C., Julio, S. M., Hentschel, U. & Mahan, M. J. (1997). Bacterial infection as assessed by in vivo gene expression. *Proc Natl Acad Sci U S A* **94**, 934-939.

Heithoff, D. M., Conner, C. P., Hentschel, U., Govantes, F., Hanna, P. C. & Mahan, M. J. (1999). Coordinate intracellular expression of Salmonella genes induced during infection. *J Bacteriol* **181**, 799-807.

Henning, M. W. & Greenfield, C. G. (1937). Salmonella bovis-morbificans (Basenau) from an outbreak of food-poisoning in Cape province. *Ondersteport Journal of Veterinary Science* **8**, 5.

Hensel, M., Shea, J. E., Waterman, S. R. & other authors (1998). Genes encoding putative effector proteins of the type III secretion system of Salmonella pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. *Molecular microbiology* **30**, 163-174.

Hensel, M., Hinsley, A. P., Nikolaus, T., Sawers, G. & Berks, B. C. (1999). The genetic basis of tetrathionate respiration in Salmonella typhimurium. *Molecular microbiology* **32**, 275-287.

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

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

Hickey, M. M. & Shanson, D. C. (1993). Septicaemia in patients with and without AIDS at Westminster Hospital, London. *J Infect* **27**, 243-250.

Hill, C. W., Sandt, C. H. & Vlazny, D. A. (1994). Rhs elements of Escherichia coli: a family of genetic composites each encoding a large mosaic protein. *Molecular microbiology* **12**, 865-871.

Hill, C. W., Feulner, G., Brody, M. S., Zhao, S., Sadosky, A. B. & Sandt, C. H. (1995). Correlation of Rhs elements with Escherichia coli population structure. *Genetics* **141**, 15-24.

Hill, C. W. (1999). Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. *Res Microbiol* **150**, 665-674.

Hinton, M. (1971). Salmonella dublin abortion in cattle: preliminary observations on the serological response. *The Veterinary record* **88**, 481.

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**, 5234-5239.

Hoffman, S. L., Punjabi, N. H., Kumala, S., Moechtar, M. A., Pulungsih, S. P., Rivai, A. R., Rockhill, R. C., Woodward, T. E. & Loedin, A. A. (1984). Reduction of mortality in chloramphenicol-treated severe typhoid fever by high-dose dexamethasone. *N Engl J Med* **310**, 82-88.

Hohmann, E. L. (2001). Nontyphoidal salmonellosis. *Clin Infect Dis* **32**, 263-269.

Holden, D. W. (2002). Trafficking of the Salmonella vacuole in macrophages. *Traffic (Copenhagen, Denmark)* **3**, 161-169.

Holt, K. E., Thomson, N. R., Wain, J. & other authors (2009). Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC genomics* **10**, 36.

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

Hsu, R. B. & Lin, F. Y. (2005). Risk factors for bacteraemia and endovascular infection due to non-typhoid salmonella: a reappraisal. *Qjm* **98**, 821-827.

Humphries, A., Deridder, S. & Baumler, A. J. (2005). Salmonella enterica serotype Typhimurium fimbrial proteins serve as antigens during infection of mice. *Infect Immun* **73**, 5329-5338.

Humphries, A. D., Raffatellu, M., Winter, S. & other authors (2003). The use of flow cytometry to detect expression of subunits encoded by 11 Salmonella enterica serotype Typhimurium fimbrial operons. *Molecular microbiology* **48**, 1357-1376.

Hung, C. C., Hung, M. N., Hsueh, P. R. & other authors (2007). Risk of recurrent nontyphoid Salmonella bacteremia in HIV-infected patients in the era of highly active antiretroviral therapy and an increasing trend of fluoroquinolone resistance. *Clin Infect Dis* **45**, e60-67.

Hyams, K. C., Oldfield, E. C., Scott, R. M. & other authors (1986). Evaluation of febrile patients in Port Sudan, Sudan: isolation of dengue virus. *The American journal of tropical medicine and hygiene* **35**, 860-865.

Igra-Siegman, Y., Kapila, R., Sen, P., Kaminski, Z. C. & Louria, D. B. (1981). Syndrome of hyperinfection with Strongyloides stercoralis. *Reviews of infectious diseases* **3**, 397-407.

Jain, R., Rivera, M. C., Moore, J. E. & Lake, J. A. (2002). Horizontal gene transfer in microbial genome evolution. *Theoretical population biology* **61**, 489-495.

Jegathesan, M. (1984). Salmonella serotypes isolated from man in Malaysia over the 10-year period 1973-1982. *The Journal of hygiene* **92**, 395-399.

Kang, M. S., Besser, T. E., Hancock, D. D., Porwollik, S., McClelland, M. & Call, D. R. (2006). Identification of specific gene sequences conserved in contemporary epidemic strains of Salmonella enterica. *Appl Environ Microbiol* **72**, 6938-6947.

Kariuki, S., Revathi, G., Gakuya, F., Yamo, V., Muyodi, J. & Hart, C. A. (2002). Lack of clonal relationship between non-typhi Salmonella strain types from humans and those isolated from animals living in close contact. *FEMS Immunol Med Microbiol* **33**, 165-171.

Kariuki, S., Revathi, G., Kariuki, N., Muyodi, J., Mwituria, J., Munyalo, A., Kagendo, D., Murungi, L. & Anthony Hart, C. (2005). Increasing prevalence of multidrug-resistant non-typhoidal salmonellae, Kenya, 1994-2003. *International journal of antimicrobial agents* **25**, 38-43.

Kariuki, S., Revathi, G., Kariuki, N., Kiiru, J., Mwituria, J. & Hart, C. A. (2006a). Characterisation of community acquired non-typhoidal Salmonella from bacteraemia and diarrhoeal infections in children admitted to hospital in Nairobi, Kenya. *BMC Microbiol* **6**, 101.

Kariuki, S., Revathi, G., Kariuki, N. & other authors (2006b). Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol* **55**, 585-591.

Karro, J. E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrrison, P. & Gerstein, M. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic acids research* **35**, D55-60.

Kassa-Kelembho, E., Mbolidi, C. D., Service, Y. B., Morvan, J. & Minssart, P. (2003). Bacteremia in adults admitted to the Department of Medicine of Bangui Community Hospital (Central African Republic). *Acta tropica* **89**, 67-72.

Kilbourn, R. G. & Belloni, P. (1990). Endothelial cell production of nitrogen oxides in response to interferon gamma in combination with tumor necrosis factor, interleukin-1, or endotoxin. *J Natl Cancer Inst* **82**, 772-776.

Kingsley, R. A. & Baumler, A. J. (2000). Host adaptation and the emergence of infectious disease: the Salmonella paradigm. *Mol Microbiol* **36**, 1006-1014.

Kingsley, R. A., Santos, R. L., Keestra, A. M., Adams, L. G. & Baumler, A. J. (2002). Salmonella enterica serotype Typhimurium ShdA is an outer membrane fibronectin-binding protein that is expressed in the intestine. *Molecular microbiology* **43**, 895-905.

Kingsley, R. A., Msefula, C. L., Thomson, N. R. & other authors (2009). Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Res* **19**, 2279-2287.

Kisiela, D., Sapeta, A., Kuczkowski, M., Stefaniak, T., Wieliczko, A. & Ugorski, M. (2005). Characterization of FimH adhesins expressed by Salmonella enterica serovar Gallinarum biovars Gallinarum and Pullorum: reconstitution of mannose-binding properties by single amino acid substitution. *Infect Immun* **73**, 6187-6190.

Kisiela, D., Laskowska, A., Sapeta, A., Kuczkowski, M., Wieliczko, A. & Ugorski, M. (2006). Functional characterization of the FimH adhesin from Salmonella enterica serovar Enteritidis. *Microbiology* **152**, 1337-1346.

Knodler, L. A., Celli, J., Hardt, W. D., Vallance, B. A., Yip, C. & Finlay, B. B. (2002). Salmonella effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. *Molecular microbiology* **43**, 1089-1103.

Knodler, L. A. & Steele-Mortimer, O. (2003). Taking possession: biogenesis of the Salmonella-containing vacuole. *Traffic (Copenhagen, Denmark)* **4**, 587-599.

Kropinski, A. M., Sulakvelidze, A., Konczy, P. & Poppe, C. (2007). Salmonella phages and prophages--genomics and practical aspects. *Methods in molecular biology (Clifton, NJ* **394**, 133-175.

Kross, I. & Schiff, F. (1940). Pseudo-surgical syndromes produced by salmonella organisms. *American Journal of Digestive Diseases* **7**, 2.

Kubori, T., Matsushima, Y., Nakamura, D., Uralil, J., Lara-Tejero, M., Sukhan, A., Galan, J. E. & Aizawa, S. I. (1998). Supramolecular structure of the Salmonella typhimurium type III protein secretion system. *Science (New York, NY* **280**, 602-605.

Kunz, L. J. & Waddell, W. R. (1956). Association of Salmonella enteritis with operations on the stomach. *N Engl J Med* **255**, 555-559.

Kuo, C. H. & Ochman, H. The extinction dynamics of bacterial pseudogenes. *PLoS genetics* **6**.

Kuo, C. H., Moran, N. A. & Ochman, H. (2009). The consequences of genetic drift for bacterial genome complexity. *Genome research* **19**, 1450-1454.

Lambert, M. A. & Smith, S. G. (2008). The PagN protein of Salmonella enterica serovar Typhimurium is an adhesin and invasin. *BMC Microbiol* **8**, 142.

Lan, R., Reeves, P. R. & Octavia, S. (2009). Population structure, origins and evolution of major Salmonella enterica clones. *Infect Genet Evol* **9**, 996-1005.

Lavy, C. B., Peek, A. C. & Manjolo, G. (2005). The incidence of septic arthritis in Malawian children. *Int Orthop* **29**, 195-196.

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 pathogens* **2**, e11.

Ledeboer, N. A., Frye, J. G., McClelland, M. & Jones, B. D. (2006). Salmonella enterica serovar Typhimurium requires the Lpf, Pef, and Tafi fimbriae for biofilm formation on HEp-2 tissue culture cells and chicken intestinal epithelium. *Infect Immun* **74**, 3156-3169.

Lee, A. K., Detweiler, C. S. & Falkow, S. (2000). OmpR regulates the two-component system SsrA-ssrB in Salmonella pathogenicity island 2. *J Bacteriol* **182**, 771-781.

Lee, W. S., Puthucheary, S. D. & Boey, C. C. (1998). Non-typhoid Salmonella gastroenteritis. *Journal of paediatrics and child health* **34**, 387-390.

Leininger, E., Ewanowich, C. A., Bhargava, A., Peppler, M. S., Kenimer, J. G. & Brennan, M. J. (1992). Comparative roles of the Arg-Gly-Asp sequence present in the Bordetella pertussis adhesins pertactin and filamentous hemagglutinin. *Infect Immun* **60**, 2380-2385.

Lepage, P., Bogaerts, J., Van Goethem, C., Ntahorutaba, M., Nsengumuremyi, F., Hitimana, D. G., Vandepitte, J., Butzler, J. P. & Levy, J. (1987). Community-acquired bacteraemia in African children. *Lancet* **1**, 1458-1461.

Lerat, E. & Ochman, H. (2005). Recognizing the pseudogenes in bacterial genomes. *Nucleic acids research* **33**, 3125-3132.

Lesnick, M. L., Reiner, N. E., Fierer, J. & Guiney, D. G. (2001). The Salmonella spvB virulence gene encodes an enzyme that ADP-ribosylates actin and destabilizes the cytoskeleton of eukaryotic cells. *Molecular microbiology* **39**, 1464-1470.

Libby, S. J., Adams, L. G., Ficht, T. A., Allen, C., Whitford, H. A., Buchmeier, N. A., Bossie, S. & Guiney, D. G. (1997). The spv genes on the Salmonella dublin

virulence plasmid are required for severe enteritis and systemic infection in the natural host. *Infect Immun* **65**, 1786-1792.

Lin, R. J., Capage, M. & Hill, C. W. (1984). A repetitive DNA sequence, rhs, responsible for duplications within the Escherichia coli K-12 chromosome. *J Mol Biol* **177**, 1-18.

Lindquist, B. L., Lebenthal, E., Lee, P. C., Stinson, M. W. & Merrick, J. M. (1987). Adherence of Salmonella typhimurium to small-intestinal enterocytes of the rat. *Infect Immun* **55**, 3044-3050.

Lisitsyn, N., Lisitsyn, N. & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science (New York, NY* **259**, 946-951.

Liu, Y., Harrison, P. M., Kunin, V. & Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome biology* **5**, R64.

Lopez, J., Delgado, D., Andres, I., Ortiz, J. M. & Rodriguez, J. C. (1991). Isolation and evolutionary analysis of a RepFVIB replicon of the plasmid pSU212. *Journal of general microbiology* **137**, 1093-1099.

Luk'ianov, K. A., Gurskaia, I. G., Matts, M. V., Khaspekov, G. L., D'Iachenko L, B., Chenchik, A. A., Il'evich-Stuchkov, S. G. & Luk'ianov, S. A. (1996). [A method for obtaining the normalized cDNA libraries based on the effect of suppression of polymerase chain reaction]. *Bioorganicheskaia khimiia* **22**, 686-690.

Lule, J. R., Mermin, J., Awor, A. & other authors (2009). Aetiology of dirrhoea among persons with HIV & their family members in rural Uganda: A community-based study. *East African medical journal* **86**, 7.

MacLennan, C., Fieschi, C., Lammas, D. A. & other authors (2004). Interleukin (IL)-12 and IL-23 are key cytokines for immunity against Salmonella in humans. *J Infect Dis* **190**, 1755-1757.

Magistrali, C., Dionisi, A. M., De Curtis, P., Cucco, L., Vischi, O., Scuota, S., Zicavo, A. & Pezzotti, G. (2008). Contamination of Salmonella spp. in a pig finishing herd, from the arrival of the animals to the slaughterhouse. *Research in veterinary science* **85**, 204-207.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends Genet* **24**, 133-141.

Mariathasan, S., Newton, K., Monack, D. M., Vucic, D., French, D. M., Lee, W. P., Roose-Girma, M., Erickson, S. & Dixit, V. M. (2004). Differential activation of the inflammasome by caspase-1 adaptors ASC and Ipaf. *Nature* **430**, 213-218.

McClelland, M., Sanderson, K. E., Spieth, J. & other authors (2001). Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* **413**, 852-856.

McClelland, M., Sanderson, K. E., Clifton, S. W. & other authors (2004). Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nat Genet* **36**, 1268-1274.

Meresse, S., Unsworth, K. E., Habermann, A., Griffiths, G., Fang, F., Martinez-Lorenzo, M. J., Waterman, S. R., Gorvel, J. P. & Holden, D. W. (2001). Remodelling of the actin cytoskeleton is essential for replication of intravacuolar Salmonella. *Cellular microbiology* **3**, 567-577.

Meyer, C. N., Skinhoj, P. & Prag, J. (1994). Bacteremia in HIV-positive and AIDS patients: incidence, species distribution, risk-factors, outcome, and influence of long-term prophylactic antibiotic treatment. *Scandinavian journal of infectious diseases* **26**, 635-642.

Miao, E. A. & Miller, S. I. (2000). A conserved amino acid sequence directing intracellular type III secretion by Salmonella typhimurium. *Proc Natl Acad Sci U S A* **97**, 7539-7544.

Milledge, J., Calis, J. C., Graham, S. M. & other authors (2005). Aetiology of neonatal sepsis in Blantyre, Malawi: 1996-2001. *Ann Trop Paediatr* **25**, 101-110.

Miller, S. I., Kukral, A. M. & Mekalanos, J. J. (1989). A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. *Proc Natl Acad Sci U S A* **86**, 5054-5058.

Miller, S. I. & Mekalanos, J. J. (1990). Constitutive expression of the phoP regulon attenuates Salmonella virulence and survival within macrophages. *J Bacteriol* **172**, 2485-2490.

Miriagou, V., Carattoli, A. & Fanning, S. (2006). Antimicrobial resistance islands: resistance gene clusters in Salmonella chromosome and plasmids. *Microbes Infect* **8**, 1923-1930.

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 U S A* **96**, 9845-9850.

Monack, D. M., Hersh, D., Ghori, N., Bouley, D., Zychlinsky, A. & Falkow, S. (2000). Salmonella exploits caspase-1 to colonize Peyer's patches in a murine typhoid model. *The Journal of experimental medicine* **192**, 249-258.

Monno, R., Rizzo, C., De Vito, D., Nucleo, E., Migliavacca, R., Pagani, L. & Rizzo, G. (2007). Prevalence, antimicrobial resistance, and extended-spectrum beta-lactamases characterization of Salmonella isolates in Apulia, southern Italy (2001-2005). *Microbial drug resistance (Larchmont, NY* **13**, 124-129.

Morgan, E., Campbell, J. D., Rowe, S. C., Bispham, J., Stevens, M. P., Bowen, A. J., Barrow, P. A., Maskell, D. J. & Wallis, T. S. (2004). Identification of host-

specific colonization factors of Salmonella enterica serovar Typhimurium. *Molecular microbiology* **54**, 994-1010.

Morgan, E., Bowen, A. J., Carnell, S. C., Wallis, T. S. & Stevens, M. P. (2007). SiiE is secreted by the Salmonella enterica serovar Typhimurium pathogenicity island 4-encoded secretion system and contributes to intestinal colonization in cattle. *Infect Immun* **75**, 1524-1533.

Morpeth, S. C., Ramadhani, H. O. & Crump, J. A. (2009). Invasive non-Typhi Salmonella disease in Africa. *Clin Infect Dis* **49**, 606-611.

Morrow, B. J., Graham, J. E. & Curtiss, R., 3rd (1999). Genomic subtractive hybridization and selective capture of transcribed sequences identify a novel Salmonella typhimurium fimbrial operon and putative transcriptional regulator that are absent from the Salmonella typhi genome. *Infect Immun* **67**, 5106-5116.

Mweu, E. & English, M. (2008). Typhoid fever in children in Africa. *Trop Med Int Health* **13**, 532-540.

Nastasi, A., Mammina, C. & Aleo, A. (1994). Epidemic dissemination of Salmonella enterica spp. enterica serovar Bovismorbificans in southern Italy in the years 1989-1991. *European journal of epidemiology* **10**, 81-84.

Neves, J., Marinho, R. P., Martins, N. R. & de Araujo, P. K. (1969). Lucciola J: Prolonged septicaemic salmonellosis: treatment of intercurrent schistosomiasis with niridazole. *Trans R Soc Trop Med Hyg* **63**, 79-84.

Nicholson, B. & Low, D. (2000). DNA methylation-dependent regulation of pef expression in Salmonella typhimurium. *Molecular microbiology* **35**, 728-742.

Nohmi, T., Yamada, M., Watanabe, M., Murayama, S. Y. & Sofuni, T. (1992). Roles of Salmonella typhimurium umuDC and samAB in UV mutagenesis and UV sensitivity. *J Bacteriol* **174**, 6948-6955.

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

Oelschlaeger, T. A., Zhang, D., Schubert, S., Carniel, E., Rabsch, W., Karch, H. & Hacker, J. (2003). The high-pathogenicity island is absent in human pathogens of Salmonella enterica subspecies I but present in isolates of subspecies III and VI. *J Bacteriol* **185**, 1107-1111.

Ohl, M. E. & Miller, S. I. (2001). Salmonella: a model for bacterial pathogenesis. *Annu Rev Med* **52**, 259-274.

Olsiewski, P. J., Kaczorowski, G. J. & Walsh, C. (1980). Purification and properties of D-amino acid dehydrogenase, an inducible membrane-bound iron-sulfur flavoenzyme from Escherichia coli B. *The Journal of biological chemistry* **255**, 4487-4494.

Paranchych, W. & Frost, L. S. (1988). The physiology and biochemistry of pili. *Advances in microbial physiology* **29**, 53-114.

Parkhill, J., Dougan, G., James, K. D. & other authors (2001). Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413**, 848-852.

Parry, C. M., Hien, T. T., Dougan, G., White, N. J. & Farrar, J. J. (2002). Typhoid fever. *N Engl J Med* **347**, 1770-1782.

Pasquali, F., Kehrenberg, C., Manfreda, G. & Schwarz, S. (2005). Physical linkage of Tn3 and part of Tn1721 in a tetracycline and ampicillin resistance plasmid from Salmonella Typhimurium. *The Journal of antimicrobial chemotherapy* **55**, 562-565.

Petit, P. L., Haarlem, J. V., Poelman, M., Haverkamp, M. C. & Wamola, I. A. (1995). Bacteraemia in patients presenting with fever. *East African medical journal* **72**, 116-120.

Petros, A., Bennett, D. & Vallance, P. (1991). Effect of nitric oxide synthase inhibitors on hypotension in patients with septic shock. *Lancet* **338**, 1557-1558.

Pfeifer, C. G., Marcus, S. L., Steele-Mortimer, O., Knodler, L. A. & Finlay, B. B. (1999). Salmonella typhimurium virulence genes are induced upon bacterial invasion into phagocytic and nonphagocytic cells. *Infect Immun* **67**, 5690-5698.

Pickard, D., Wain, J., Baker, S. & other authors (2003). Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding Salmonella enterica pathogenicity island SPI-7. *J Bacteriol* **185**, 5055-5065.

Poh, J., Odendall, C., Spanos, A., Boyle, C., Liu, M., Freemont, P. & Holden, D. W. (2008). SteC is a Salmonella kinase required for SPI-2-dependent F-actin remodelling. *Cellular microbiology* **10**, 20-30.

Popiel, I. & Turnbull, P. C. (1985). Passage of Salmonella enteritidis and Salmonella thompson through chick ileocecal mucosa. *Infect Immun* **47**, 786-792.

Popoff, M. Y., Bockemuhl, J. & Gheesling, L. L. (2003). Supplement 2001 (no. 45) to the Kauffmann-White scheme. *Res Microbiol* **154**, 173-174.

Porwollik, S., Boyd, E. F., Choy, C., Cheng, P., Florea, L., Proctor, E. & McClelland, M. (2004). Characterization of Salmonella enterica subspecies I genovars by use of microarrays. *J Bacteriol* **186**, 5883-5898.

Pullinger, G. D., Dziva, F., Charleston, B., Wallis, T. S. & Stevens, M. P. (2008). Identification of Salmonella enterica serovar Dublin-specific sequences by subtractive hybridization and analysis of their role in intestinal colonization and systemic translocation in cattle. *Infect Immun* **76**, 5310-5321.

Punjabi, N. H., Hoffman, S. L., Edman, D. C. & other authors (1988). Treatment of severe typhoid fever in children with high dose dexamethasone. *Pediatr Infect Dis J* 7, 598-600.

Puohiniemi, R., Heiskanen, T. & Siitonen, A. (1997). Molecular epidemiology of two international sprout-borne Salmonella outbreaks. *J Clin Microbiol* **35**, 2487-2491.

Rathman, M., Sjaastad, M. D. & Falkow, S. (1996). Acidification of phagosomes containing Salmonella typhimurium in murine macrophages. *Infect Immun* **64**, 2765-2773.

Reddy, E. A., Shaw, A. V. & Crump, J. A. (2010). Community-acquired bloodstream infections in Africa: a systematic review and meta-analysis. *The Lancet infectious diseases* **10**, 417-432.

Rhen, M. & Sukupolvi, S. (1988). The role of the traT gene of the Salmonella typhimurium virulence plasmid for serum resistance and growth within liver macrophages. *Microb Pathog* **5**, 275-285.

Rode, C. K., Melkerson-Watson, L. J., Johnson, A. T. & Bloch, C. A. (1999). Type-specific contributions to chromosome size differences in Escherichia coli. *Infect Immun* **67**, 230-236.

Rogerson, S. J., Spooner, V. J., Smith, T. A. & Richens, J. (1991). Hydrocortisone in chloramphenicol-treated severe typhoid fever in Papua New Guinea. *Trans R Soc Trop Med Hyg* **85**, 113-116.

Rotger, R. & Casadesus, J. (1999). The virulence plasmids of Salmonella. *Int Microbiol* **2**, 177-184.

Sadosky, A. B., Gray, J. A. & Hill, C. W. (1991). The RhsD-E subfamily of Escherichia coli K-12. *Nucleic acids research* **19**, 7177-7183.

Salcedo, S. P., Noursadeghi, M., Cohen, J. & Holden, D. W. (2001). Intracellular replication of Salmonella typhimurium strains in specific subsets of splenic macrophages in vivo. *Cellular microbiology* **3**, 587-597.

Salit, I. E. & Gotschlich, E. C. (1977). Hemagglutination by purified type I Escherichia coli pili. *The Journal of experimental medicine* **146**, 1169-1181.

Saphra, I. & Winter, J. W. (1957). Clinical manifestations of salmonellosis in man; an evaluation of 7779 human infections identified at the New York Salmonella Center. *N Engl J Med* **256**, 1128-1134.

Schiellerup, P., Neimann, J., Hald, T. M. & Ethelberg, S. (2001). [Outbreak of Salmonella Bovismorbificans infection]. *Ugeskrift for laeger* **163**, 5683.

Selander, R. K., Beltran, P., Smith, N. H. & other authors (1990). Evolutionary genetic relationships of clones of Salmonella serovars that cause human typhoid and other enteric fevers. *Infect Immun* **58**, 2262-2275.

Seligmann, E., Saphra, I. & Wassermann, M. (1943). Salmonella infections in man. An analysis of 1000 cases bacteriologically identified by the New York Salmonella center. *American Journal of Hygiene* **38**, 33.

Seligmann, E., Saphra, I. & Wassermann, M. (1946). Salmonella infections in the U.S.A. A second series of 2000 infections recorded by the N.Y Salmonella Center. *Journal of Immunology* **54**, 18.

Shea, J. E., Hensel, M., Gleeson, C. & Holden, D. W. (1996). Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. *Proc Natl Acad Sci U S A* **93**, 2593-2597.

Shimeles, D. & Lulseged, S. (1994). Clinical profile and pattern of infection in Ethiopian children with severe protein-energy malnutrition. *East African medical journal* **71**, 264-267.

Sigauque, B., Roca, A., Mandomando, I. & other authors (2009). Community-acquired bacteremia among children admitted to a rural hospital in Mozambique. *Pediatr Infect Dis J* **28**, 108-113.

Silverman, M., Zieg, J., Hilmen, M. & Simon, M. (1979). Phase variation in Salmonella: genetic analysis of a recombinational switch. *Proc Natl Acad Sci U S A* **76**, 391-395.

Sinnott, C. R. & Teall, A. J. (1987). Persistent gallbladder carriage of Salmonella typhi. *Lancet* **1**, 976.

Smart, C. H., Walshaw, M. J., Hart, C. A. & Winstanley, C. (2006). Use of suppression subtractive hybridization to examine the accessory genome of the Liverpool cystic fibrosis epidemic strain of Pseudomonas aeruginosa. *J Med Microbiol* **55**, 677-688.

Stafford, R. J., McCall, B. J., Neill, A. S., Leon, D. S., Dorricott, G. J., Towner, C. D. & Micalizzi, G. R. (2002). A statewide outbreak of Salmonella bovismorbificans phage type 32 infection in Queensland. *Communicable diseases intelligence* **26**, 568-573.

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**, 4406-4413.

Stewart, D. F. (1940a). Two mortalities among sheep due to infection with species of the Salmonella group of bacteria. *Australian Veterinary Journal* **16**, 3.

Stewart, D. F. (1940b). Further Salmonella infections in pigs and acute paratyphoid septicaemia in horses. *Australian Veterinary Journal* **16**, 2.

Straus, D. & Ausubel, F. M. (1990). Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc Natl Acad Sci U S A* **87**, 1889-1893.

Suvarnapunya, A. E. & Stein, M. A. (2005). DNA base excision repair potentiates the protective effect of Salmonella Pathogenicity Island 2 within macrophages. *Microbiology* **151**, 557-567.

Taira, S., Baumann, M., Riikonen, P., Sukupolvi, S. & Rhen, M. (1991). Amino-terminal sequence analysis of four plasmid-encoded virulence-associated proteins of Salmonella typhimurium. *FEMS Microbiol Lett* **61**, 319-323.

Thamlikitkul, V., Dhiraputra, C., Paisarnsinsup, T. & Chareandee, C. (1996). Non-typhoidal Salmonella bacteraemia: clinical features and risk factors. *Trop Med Int Health* **1**, 443-448.

Thomson, N., Baker, S., Pickard, D. & other authors (2004). The role of prophage-like elements in the diversity of Salmonella enterica serovars. *J Mol Biol* **339**, 279-300.

Thomson, N. R., Clayton, D. J., Windhorst, D. & other authors (2008). Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**, 1624-1637.

Threlfall, E. J., Hall, M. L. & Rowe, B. (1992). Salmonella bacteraemia in England and Wales, 1981-1990. *J Clin Pathol* **45**, 34-36.

Tinge, S. A. & Curtiss, R., 3rd (1990). Conservation of Salmonella typhimurium virulence plasmid maintenance regions among Salmonella serovars as a basis for plasmid curing. *Infect Immun* **58**, 3084-3092.

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

Townsend, S. M., Kramer, N. E., Edwards, R. & other authors (2001). Salmonella enterica serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect Immun* **69**, 2894-2901.

Trevejo, R. T., Courtney, J. G., Starr, M. & Vugia, D. J. (2003). Epidemiology of salmonellosis in California, 1990-1999: morbidity, mortality, and hospitalization costs. *Am J Epidemiol* **157**, 48-57.

Tsolis, R. M., Townsend, S. M., Miao, E. A., Miller, S. I., Ficht, T. A., Adams, L. G. & Baumler, A. J. (1999). Identification of a putative Salmonella enterica

serotype typhimurium host range factor with homology to IpaH and YopM by signature-tagged mutagenesis. *Infect Immun* **67**, 6385-6393.

Tsolis, R. M., Young, G. M., Solnick, J. V. & Baumler, A. J. (2008). From bench to bedside: stealth of enteroinvasive pathogens. *Nat Rev Microbiol* **6**, 883-892.

Uchiya, K., Barbieri, M. A., Funato, K., Shah, A. H., Stahl, P. D. & Groisman, E. A. (1999). A Salmonella virulence protein that inhibits cellular trafficking. *The EMBO journal* **18**, 3924-3933.

Uthe, J. J., Stabel, T. J., Zhao, S. H., Tuggle, C. K. & Bearson, S. M. (2006). Analysis of porcine differential gene expression following challenge with Salmonella enterica serovar Choleraesuis using suppression subtractive hybridization. *Veterinary microbiology* **114**, 60-71.

Vaagland, H., Blomberg, B., Kruger, C., Naman, N., Jureen, R. & Langeland, N. (2004). Nosocomial outbreak of neonatal Salmonella enterica serotype Enteritidis meningitis in a rural hospital in northern Tanzania. *BMC infectious diseases* **4**, 35.

Van Asten, F. J., Hendriks, H. G., Koninkx, J. F., Van der Zeijst, B. A. & Gaastra, W. (2000). Inactivation of the flagellin gene of Salmonella enterica serotype enteritidis strongly reduces invasion into differentiated Caco-2 cells. *FEMS Microbiol Lett* **185**, 175-179.

van der Velden, A. W., Baumler, A. J., Tsolis, R. M. & Heffron, F. (1998). Multiple fimbrial adhesins are required for full virulence of Salmonella typhimurium in mice. *Infect Immun* **66**, 2803-2808.

van der Velden, A. W., Lindgren, S. W., Worley, M. J. & Heffron, F. (2000). Salmonella pathogenicity island 1-independent induction of apoptosis in infected macrophages by Salmonella enterica serotype typhimurium. *Infect Immun* **68**, 5702-5709.

Vandenbosch, J. L., Kurlandsky, D. R., Urdangaray, R. & Jones, G. W. (1989). Evidence of coordinate regulation of virulence in Salmonella typhimurium involving the rsk element of the 95-kilobase plasmid. *Infect Immun* **57**, 2566-2568.

Vugia, D. J., Kiehlbauch, J. A., Yeboue, K. & other authors (1993). Pathogens and predictors of fatal septicemia associated with human immunodeficiency virus infection in Ivory Coast, west Africa. *J Infect Dis* **168**, 564-570.

Vugia, D. J., Samuel, M., Farley, M. M., Marcus, R., Shiferaw, B., Shallow, S., Smith, K. & Angulo, F. J. (2004). Invasive Salmonella infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. *Clin Infect Dis* **38 Suppl 3**, S149-156.

Waage, A., Brandtzaeg, P., Espevik, T. & Halstensen, A. (1991). Current understanding of the pathogenesis of gram-negative shock. *Infect Dis Clin North Am* **5**, 781-791.

Walsh, A. L., Phiri, A. J., Graham, S. M., Molyneux, E. M. & Molyneux, M. E. (2000). Bacteremia in febrile Malawian children: clinical and microbiologic features. *Pediatr Infect Dis J* **19**, 312-318.

Weening, E. H., Barker, J. D., Laarakker, M. C., Humphries, A. D., Tsolis, R. M. & Baumler, A. J. (2005). The Salmonella enterica serotype Typhimurium lpf, bcf, stb, stc, std, and sth fimbrial operons are required for intestinal persistence in mice. *Infect Immun* **73**, 3358-3366.

Weinberger, M., Andorn, N., Agmon, V., Cohen, D., Shohat, T. & Pitlik, S. D. (2004). Blood invasiveness of Salmonella enterica as a function of age and serotype. *Epidemiol Infect* **132**, 1023-1028.

White, D. G., Zhao, S., Singh, R. & McDermott, P. F. (2004). Antimicrobial resistance among gram-negative foodborne bacterial pathogens associated with foods of animal origin. *Foodborne pathogens and disease* **1**, 137-152.

Williamson, C. M., Baird, G. D. & Manning, E. J. (1988). A common virulence region on plasmids from eleven serotypes of Salmonella. *Journal of general microbiology* **134**, 975-982.

Wilson, R. P., Raffatellu, M., Chessa, D., Winter, S. E., Tukel, C. & Baumler, A. J. (2008). The Vi-capsule prevents Toll-like receptor 4 recognition of Salmonella. *Cell Microbiol* **10**, 876-890.

Winstanley, C. (2002). Spot the difference: applications of subtractive hybridisation to the study of bacterial pathogens. *J Med Microbiol* **51**, 459-467.

Wood, M. W., Jones, M. A., Watson, P. R., Hedges, S., Wallis, T. S. & Galyov, E. E. (1998). Identification of a pathogenicity island required for Salmonella enteropathogenicity. *Molecular microbiology* **29**, 883-891.

Woodward, T. E., Smadel, J. E. & et al. (1948). Preliminary report on the beneficial effect of chloromycetin in the treatment of typhoid fever. *Ann Intern Med* **29**, 131-134.

Woodward, T. E. (1949). Chloromycetin and aureomycin; therapeutic results. *Ann Intern Med* **31**, 53-82.

Wright, R. C. (1986). The seasonality of bacterial quality of water in a tropical developing country (Sierra Leone). *The Journal of hygiene* **96**, 75-82.

Yamamoto, N. (1969). Genetic evolution of bacteriophage. I. Hybrids between unrelated bacteriophages P22 and Fels 2. *Proc Natl Acad Sci U S A* **62**, 63-69.

Yamamoto, N. & McDonald, R. J. (1986). Genomic structure of phage F22, a hybrid between serologically and morphologically unrelated Salmonella typhimurium bacteriophages P22 and Fels 2. *Genetical research* **48**, 139-143.

Zhang, S., Kingsley, R. A., Santos, R. L. & other authors (2003). Molecular pathogenesis of Salmonella enterica serotype typhimurium-induced diarrhea. *Infect Immun* **71**, 1-12.

Zhang, X. L., Morris, C. & Hackett, J. (1997). Molecular cloning, nucleotide sequence, and function of a site-specific recombinase encoded in the major 'pathogenicity island' of Salmonella typhi. *Gene* **202**, 139-146.

Zhao, S. & Hill, C. W. (1995). Reshuffling of Rhs components to create a new element. *J Bacteriol* **177**, 1393-1398.

Zhou, D., Mooseker, M. S. & Galan, J. E. (1999). Role of the S. typhimurium actin-binding protein SipA in bacterial internalization. *Science (New York, NY* **283**, 2092-2095.

RESEARCH LETTER

Identification and distribution of accessory genome DNA sequences from an invasive African isolate of *Salmonella* Heidelberg

Christina Bronowski & Craig Winstanley

Division of Medical Microbiology, University of Liverpool, Liverpool, UK

Correspondence: Craig Winstanley, Division of Medical Microbiology, University of Liverpool, Duncan Building, Daulby Street, Liverpool L69 3GA, UK. Tel.: +44 151 706 4388; fax: +44 151 706 5805; e-mail: c.winstanley@liv.ac.uk

Received 30 April 2009; accepted 8 June 2009. Final version published online 6 July 2009.

DOI:10.1111/j.1574-6968.2009.01697.x

Editor: Stephen Smith

Keywords
Salmonella; Heidelberg; fimbriae; subtractive hybridization.

Abstract

Nontyphoidal salmonellae (NTS) are a leading cause of invasive disease in young children in sub-Saharan Africa. We used suppression subtractive hybridization (SSH) to identify 41 sequences within the accessory genome of an invasive strain of *Salmonella* Heidelberg from Malawi. PCR assays and database searches, used to determine the distribution of 14 SSH sequences among a panel of African and UK NTS isolates and published genomes, indicated that two were specific for *S.* Heidelberg. However, we found no evidence for major differences in the accessory genome content between African invasive and gastrointestinal isolates of *S.* Heidelberg. Six of the SSH sequences were within fimbrial operons. The *tcf* operon, associated with the host specificity of *Salmonella* Typhi, and the *stk* operon, reported previously in *Salmonella* Paratyphi, were both present in either all (*tcf*) or most (*stk*) isolates of *S.* Heidelberg, but had restricted distributions among the other serovars tested. Reverse transcription PCR analysis of seven SSH sequences indicated variable expression of the *stk* operon among isolates of *S.* Heidelberg. Three of the seven targeted genes were not expressed in a UK veterinary isolate of *S.* Heidelberg, suggesting that although genome content *per se* may not explain the different pathogenicity of the invasive isolates, it is possible that variations in gene expression may play a role.

Introduction

Nontyphoidal salmonellae (NTS) are a common cause of meningitis and septic arthritis in children (Graham, 2002; Lavy *et al.*, 2005), and of bacteraemia in both children and HIV-infected adults (Bahwere *et al.*, 2001; Gordon *et al.*, 2001; Berkley *et al.*, 2005; Gordon, 2008) in sub-Saharan Africa. However, the factors determining this invasive pathogenicity are not well understood. There is some evidence that clinical NTS isolates from Kenya differ from isolates taken from livestock and the environment of patients in both their genetic profile, as determined by molecular typing, and their antibiotic susceptibility patterns (Kariuki *et al.*, 2002, 2006). This implies an alternative route of transmission of African NTS, compared with North America and Europe, where domestic animals represent the major reservoir of NTS and foodstuffs of animal origin are the vehicle of human infection (Threlfall *et al.*, 2000).

Salmonella Heidelberg is one of the more virulent and invasive *Salmonella* serovars, causing extraintestinal infections associated with severe disease symptoms, such as myocarditis and bacteraemia (Rice *et al.*, 1976; Burt *et al.*, 1990; Wilmshurst & Sutcliffe, 1995). As well as being common among gastrointestinal isolates, *S.* Heidelberg has been reported as the most common *Salmonella* serovar among blood culture isolates in Canada (Demczuk *et al.*, 2003), and the third most common serovar causing invasive NTS infection, after *Salmonella* Typhimurium and *Salmonella* Enteritidis, in the United States (Vugia *et al.*, 2004). The WHO Global Salmonella Survey (GSS) showed that between 2000 and 2004 *S.* Heidelberg was the fourth most common serotype among human isolates, and the second most common serotype among non-human isolates in the world. However, because African countries participating in the WHO GSS are predominantly located in West Africa (Galanis *et al.*, 2006), data on NTS from sub-Saharan Africa are limited.

Suppression subtractive hybridization (SSH) is a method for the identification of DNA sequences present within the genome of one strain but absent from the genome of another (Winstanley, 2002), and it has been used previously on several *Salmonella* serovars, including Typhimurium and Enteritidis (Agron *et al.*, 2001; Kang *et al.*, 2006). The

FEMS MICROBIOLOGY LETTERS

method is ideal for identifying genomic regions that vary between strains of the same species (comprising the 'accessory genome'). In this study, we report the use of SSH to screen for sequences in the accessory genome of an African paediatric bacteraemia isolate of *S.* Heidelberg (strain D23734), and report the distribution of such sequences among a collection of *Salmonella* isolates including African NTS, *S.* Heidelberg isolates from Europe, and previously sequenced strains of *Salmonella enterica.*

Materials and methods

Bacterial strains used in this study

African NTS isolates were obtained from existing collections maintained in our University from Uganda, Malawi, Kenya and the Democratic Republic of Congo (Zaire). UK isolates were obtained from human and animal faeces (Table 1). Other isolates were obtained from John Wain (Health Protection Agency, Colindale, London). For UK isolates, serovar designations were confirmed by the *Salmonella* Reference Laboratory, Colindale, London. For the Malawi isolates, serovars other than *S.* Typhimurium and *S.* Enteritidis were confirmed by the National *Salmonella* Reference Laboratory, Galway, Republic of Ireland.

Construction and screening of subtraction libraries

The SSH tester strain was the bacteraemia isolate D23734, from a Malawian child. The serotype of the tester strain was confirmed by both serology and flagellin gene sequencing. The driver strain used in the SSH was the laboratory strain *S.* Typhimurium LT2 for which the complete genome sequence is known (McClelland *et al.*, 2001).

Genomic DNA was isolated from *S. enterica* strains using the Wizard Genomic DNA Purification kit and following the manufacturer's protocol (Promega). SSH was carried out using the Clontech PCR-Select Bacterial Genome Subtraction kit, as recommended by the supplier (Clontech). Tester and driver strain DNAs were digested with RsaI and the PCR amplicons obtained following SSH were cloned into pGEM-T (Invitrogen). The subtraction libraries of RsaI fragments thus constructed were screened by sequencing of plasmid DNA extracted from individual clones using forward (M13-F: 5′-ACGTTGCACAATCCCGGAT-3′) and reverse (M13-R: 5′-CCACCGAAGAAGGAGCAA-3′) vector primers (Cogenics Lark). In order to identify genuinely subtracted sequences, BLASTN searches targeting the genome of *S.* Typhimurium LT2 were conducted. Sequences sharing > 90% identity with the driver genome were omitted from

Table 1. Summary of African and UK *Salmonella* strain sets

Country of origin	Serovar	Number	Details and sources
Uganda	Typhimurium	9	Adult diarrhoea isolates (M. Okong, unpublished data)
	Enteritidis	13	
	Heidelberg	1	
	Stanleyville	1	
Malawi	Typhimurium	17	Children bacteraemia isolates 1998–2004 (Gordon, 2008)
	Enteritidis	4	
	Heidelberg	1	
	Bovismorbificans	4	
	Bukavu	1	
	Sundsvall	1	
Kenya	Typhimurium	4	Adult bacteraemia isolates, 1994–2003 (Kariuki *et al.*, 2005)
	Enteritidis	4	
DRC (Zaire)	Typhimurium	7	Children bacteraemia isolates (Green & Cheesbrough, 1993)
	Enteritidis	5	
UK	Typhimurium	6	Adult diarrhoea isolates (C. Parry, unpublished data)
	Enteritidis	6	
UK Veterinary	Heidelberg	3	Veterinary isolates (P. Wigley, unpublished data)
	Bovismorbificans	4	
Unknown	Heidelberg	3	Gut (Sanger Institute, unpublished data)
Zanzibar	Heidelberg	1	Gut (Sanger Institute, unpublished data)
Kenya	Heidelberg	1	Blood (Sanger Institute, unpublished data)
Tanzania	Heidelberg	1	Faeces (Sanger Institute, unpublished data)
Nigeria	Heidelberg	1	Gut (Sanger Institute, unpublished data)
Peru	Heidelberg	1	Gut (Sanger Institute, (unpublished data)
Thailand	Heidelberg	1	Gut (Sanger Institute, unpublished data)
Malaysia	Heidelberg	1	Gut (Sanger Institute, unpublished data)

further study. Sequences sharing < 90% identity with the genome of the driver strain were further analysed using BLASTN and BLASTX searches of the general database. Similar BLASTN searches were used to determine the presence or the absence of SSH sequences from the genomes of published *Salmonella* genomes, including *S.* Heidelberg SL476 and SL486. All searches were performed using the NCBI website (http://www.ncbi.nlm.nih.gov).

PCR amplification screening of strains

Oligonucleotide primers (Sigma-Genosys) for PCR screening using amplifications are listed in the Supporting Information, Table S1, along with the annealing temperatures used. DNA for PCR amplification was prepared by boiling a suspension of a few colonies in 5% 200 μL Chelex-100 (Bio-Rad) for 5 min. After centrifugation, 150 μL was removed and stored at −20 °C. For PCR amplification, typically 1 μL DNA was used directly in 25-μL volumes containing 1.25 U GoTaq DNA polymerase (Promega), 300 nM each primer, GoTaq buffer, 2.5 mM $MgCl_2$ and 100 mM nucleotides (dATP, dCTP, dGTP, and dTTP). Amplifications were carried out in an Eppendorf MasterCycler thermal cycler for 30 cycles consisting of 95 °C for 1 min, annealing temperature for 1 min and 72 °C for 2 min, with an additional extension time at 72 °C for 10 min following completion of the 30 cycles.

Preparation of cDNA for reverse transcription (RT) PCR amplification

A subset of six *S.* Heidelberg strains was used to determine the expression of SSH sequences. Cells were gown overnight in Luria broth at 37 °C with shaking at 200 r.p.m. Cells were collected by centrifugation of 1 mL of overnight culture at 7000 *g* for 2 min. RNA extraction was performed using the RiboPure™ Bacteria whole RNA isolation kit (Ambion, Applied Biosystems) following the manufacturer's instructions, except that two rounds of the recommended DNAse I treatment of the RNA samples were performed. The RNA was converted to cDNA using the SuperScript II RT kit (Invitrogen). For each RNA preparation, a control reaction lacking the reverse transcriptase was also prepared. We confirmed that expression of the flagellin gene (*fliC*) was detectable in all cDNA samples, but not in any of the controls.

Results

Identification of genetic sequences present in the Malawian *S.* Heidelberg isolate D23734 but absent from the *S.* Typhimurium strain LT2

In order to identify novel sequences that might be associated with the invasive phenotype of *S.* Heidelberg D23734, we carried out SSH using *S.* Typhimurium LT2 as the reference (driver) strain. We chose this strategy with the aim of identifying novel genes common to both invasive and gastrointestinal isolates of *S.* Heidelberg, as well as genes specific to the African invasive isolates. Following SSH, a total of 114 clones were sequenced, resulting in the identification of 52 (46%) genuinely subtracted sequences, 11 of which were repeated more than once, leaving an output of 41 different *S.* Heidelberg D23734 subtracted sequences. A summary of all the SSH sequences is given in Table S2, organized according to their putative function as determined by the BLASTX search. Although many of the SSH sequences matched hypothetical proteins, among a number of matches of potential relevance to virulence were six related to fimbriae, one related to a putative autotransporter, one related to a lipoprotein, two related to transcriptional regulators and one related to a ferrichrome-iron receptor.

Of the 41 SSH sequences, 37 (88%) matched at least partly one or both of the genome-sequenced *S.* Heidelberg strains SL476 or SL486, and the other four matched a *Salmonella* serovar other than *S.* Heidelberg. Using BLASTN, the genome of *Salmonella* Typhi CT18 was screened for the presence of the SSH sequences. Nineteen (46%) of the 41 SSH sequences were found to be at least partly present in *S.* Typhi CT18.

Distribution of SSH sequences among a panel of African and UK NTS isolates

In order to determine the distribution of sequences identified in the genome of *S.* Heidelberg D23734 using SSH, we designed oligonucleotide primers to 14 of the subtracted sequences and screened a panel of African, Tropical and UK human and veterinary NTS isolates using PCR assays (Table 2). The SSH sequences chosen for further study were selected on the basis of (1) potential roles in virulence according to the best BLASTX matches (fimbrial, autotransporter, iron receptor, transcriptional regulator, and lipoprotein) and (2) matches to enzymes or hypothetical proteins with limited distributions among *Salmonella* serovars according to BLASTX searches.

Our collection of invasive NTS isolates from Africa is dominated by *S.* Typhimurium and *S.* Enteritidis. In order to screen for common sequences among NTS invasive isolates in this collection regardless of serotype, we included greater numbers of these more common serovars in the panel, alongside *S.* Heidelberg and other less common serotypes. However, none of the invasive *S.* Typhimurium isolates tested positive for any of the subtracted sequences. SSH sequence D23734-G3 represents a putative fimbrial subunit present in a restricted number of serovars. All *S.* Enteritidis and *Salmonella* Bovismorbificans, and the majority of *S.* Heidelberg isolates, were PCR positive for this

Table 2. Distribution of selected SSH sequences according to PCR assays

	Sub-Saharan African and Tropical Isolates					UK isolates			
SSH sequence putative function based on BLASTX (and sequence identifier)	*S.* Typhimurium (*n* = 38)	*S.* Enteritidis (*n* = 27)	*S.* Heidelberg (*n* = 12)	*S.* Bovismorbificans (*n* = 5)	Other *Salmonella* serovars (*n* = 3)	*S.* Typhimurium (*n* = 6)	*S.* Enteritidis (*n* = 6)	*S.* Heidelberg (*n* = 3)	*S.* Bovismorbificans (*n* = 4)
2-Nitropropane dioxygenase NPD (D23734-4)	–	–	+	–	1 (33)	–	–	+	–
Short-chain dehydrogenase/reductase (D23734-10)	–	–	+	–	–	–	–	+	–
Putative autotransporter/pertactin (D23734-14)	–	–	11 (92)	+	1 (33)	–	–	+	–
Putative fimbrial protein *stkD* (D23734-A1)	–	–	8 (67)	–	–	–	–	2 (66)	–
Putative fimbrial protein *tcfA* (D23734-A5)	–	–	+	–	2 (66)	–	–	+	–
Fimbrial subunit (D23734-G3)	–	+	10 (83)	+	–	–	+	+	+
Probable lipoprotein (D23734-B10)	–	–	7 (58)	+	2 (66)	–	1 (17)	+	+
Ferrichrome-iron receptor (D23734-D2)	–	+	10 (83)	–	1 (33)	–	+	+	–
Transcriptional regulator *tinR* (D23734-D4)	–	4 (15)	+	–	+	–	1 (17)	+	11 (25)
Conserved hypothetical protein (D23734-B9)	–	–	+	–	–	–	–	+	–
Restriction enzyme (D23734-B2)	–	–	+	–	–	–	–	+	–
Hypothetical protein (D23734-5)	–	–	+	–	–	–	–	+	–
Hypothetical protein (D23734-2)	–	–	+	–	–	–	–	+	–
Hypothetical protein (D23734-1)	–	–	11 (92)	–	2 (66)	–	–	+	+

The table shows the number of strains that tested positive for a given sequence by PCR assay; percentages are given in parentheses. The group of sub-Saharan African and Tropical *S.* Heidelberg isolates includes the tester strain D23734. +, All of the strains (100%) were PCR positive; –, all of the strains tested were PCR negative.

sequence (Table 2). Sequence D23734-G3 shares 94% identity with a putative gene identified in the *S.* Enteritidis phage type 4 strain P125109 (SEN2799; putative fimbrial subunit protein) (Thomson *et al.*, 2008).

According to the PCR assays, two other fimbriae-related sequences (representing *tcf* and *stk* fimbriae, respectively) were either present only in *S.* Heidelberg isolates (*stk*) or found in all *S.* Heidelberg isolates, but also in some other serovars (*tcf*). Either *S.* Enteritidis or *S.* Bovismorbificans was PCR positive for three other SSH sequences, but all *S.* Enteritidis and *S.* Bovismorbificans isolates were PCR negative for 10 of the SSH sequences (Table 2). For eight of the 14 SSH sequences, all *S.* Heidelberg isolates were PCR positive. We found no evidence among the African isolates for distribution according to geographical source (data not shown).

Distribution of SSH sequences among genome-sequenced *Salmonella* serovars

We searched *Salmonella* genomes in the database for the presence of the 14 SSH sequences. Thirteen of the SSH sequences were found to be present in the genomes of both *S.* Heidelberg strains SL476 and SL486. However, D23734-B10, part of a gene encoding a putative lipoprotein, was absent from both genomes. After *S.* Heidelberg, D23734 SSH sequences were most commonly found in *Salmonella* Paratyphi A (Table 3).

Expression of genes represented by SSH sequences

A subset of six *S.* Heidelberg strains was chosen for RT-PCR screening to detect the expression profile of seven of the SSH sequences. According to PCR assay results, each of the six strains contained each of the seven SSH sequences tested. The results are shown in Table 4. The majority of SSH sequences lie within genes expressed in all of the African and Tropical *S.* Heidelberg isolates. The exception was SSH sequence D23734-A1 (putative *stk* fimbrial gene), for which expression could not be detected in three of the isolates. For three of the SSH sequences, expression could not be detected in the UK veterinary isolate KMS1977 (Table 4).

Discussion

We used SSH to identify 41 sequences that were present in the accessory genome of African bacteraemia isolate D23734, but absent from *S.* Typhimurium LT2. Our results suggest that there is little variation between the genomes of African invasive and faecal isolates of *S.* Heidelberg. We did identify one SSH sequence (D23734-B10) lacking from the two genome-sequenced *S.* Heidelberg strains but present in the majority of African isolates. The presence of this sequence in some other invasive serovars, such as Typhi,

Table 3. BLASTX screening results for D23734 SSH sequences

SSH sequence putative function (and sequence identifier)	S. Heidelberg (SL476)	S. Heidelberg (SL486)	S. Choleraesuis (SC-B67)	S. Typhi (Ty2)	S. Typhi (CT18)	S. Agona (SL483)	S. Javiana (GAMM 040433)	S. Kentucky (CDC191)	S. Kentucky (CVM29188)	S. Saintpaul (SARA23)	S. Saintpaul (SARA29)	S. Schwarzengrund (CVM19633)	S. Schwarzengrund (SL480)	S. Newport (SL254)	S. Newport (SL317)	S. Paratyphi A (ATCC9150)	S. Dublin (CT.020211853)
2-Nitropropane dioxygenase NPD (D23734-4)	+	+	–	–	–	+	–	+	+	–	–	+	+	–	–	–	–
Short-chain dehydrogenase/reductase (D23734-10)	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Putative autotransporter/pertactin (D23734-14)	+	+	–	–	–	–	–	–	–	–	+	–	–	–	–	+	+
Putative fimbrial protein *stkD* (D23734-A1)	+	+	–	–	–	–	–	+	+	–	–	–	–	–	–	+	–
Putative fimbrial protein *tcfA* (D23734-A5)	+	+	+	+	+	–	–	–	–	–	–	+	+	–	–	+	–
Fimbrial subunit (D23734-G3)	+	+	–	–	–	–	–	–	–	–	–	–	–	+	–	–	+
Probable lipoprotein (D23734-B10)	–	–	–	+	+	–	–	–	–	–	+	–	–	+	+	+	–
Ferrichrome-iron receptor (D23734-D2)	+	+	–	–	–	+	–	+	+	–	–	–	–	–	–	+	–
Transcriptional regulator *tinR* (D23734-D4)	+	+	+	+	+	–	+	–	–	–	–	+	+	–	–	+	–
Conserved hypothetical protein (D23734-B9)	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Restriction enzyme (D23734-B2)	+	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Hypothetical protein (D23734-5)	+	+	–	+	+	–	–	–	–	–	+	–	–	–	–	–	–
Hypothetical protein (D23734-2)	+	+	–	–	–	+	–	+	+	–	–	–	–	–	–	–	–
Hypothetical protein (D23734-1)	+	+	–	+	+	+	+	+	+	–	–	+	+	–	–	+	–

Table 4. Expression profile of seven SSH sequences among a panel of six *Salmonella* Heidelberg strains

SSH sequence putative function (and sequence identifier)	D23734 (Malawi)	845 (Malawi)	20031619 (Zanzibar)	20040049 (Kenya)	20041283 (Peru)	KMS1977 (UK Veterinary)
Putative autotransporter/pertactin (D23734-14)	+	+	+	+	+	+
Putative fimbrial protein *stkD* (D23734-A1)	+	+	–	–	–	+
Putative fimbrial protein *tcfA* (D23734-A5)	+	+	+	+	+	+
Fimbrial subunit (D23734-G3)	+	+	+	+	+	–
Probable lipoprotein (D23734-B10)	+	+	+	+	+	+
Ferrichrome-iron receptor (D23734-D2)	+	+	+	+	+	–
Transcriptional regulator *tinR* (D23734-D4)	+	+	+	+	+	–

suggests that the genomic region represented by this sequence may merit further study. However, we found the SSH sequence to be present also in UK veterinary isolates of *S.* Heidelberg, and demonstrated the expression of the putative lipoprotein-encoding gene in all six isolates tested, including one of the veterinary isolates. Hence, there was no clear association of either the presence of the gene or the expression of the gene with the invasive isolates. Likewise, the potentially virulence-related putative autotransporter/pertactin-related SSH sequence D23734-14 was found in 92% of the African/Tropical *S.* Heidelberg isolates and all of the UK veterinary isolates, and was expressed in all the *S.* Heidelberg isolates tested. Interestingly, based on PCR assays, this sequence was present in the five African isolates of *S.* Bovismorbificans, but absent from UK gastrointestinal isolates of this serovar. The sequence was also present in some other invasive serovars. However, overall, there was no clear correlation between SSH sequence distributions and those serovars generally regarded as more virulent/invasive, including serovars Cholaerasuis, Dublin, Schwarzengrund and Newport (Threlfall *et al.*, 1992; Chiu *et al.*, 2006).

Salmonella Heidelberg is in the same serogroup as *S.* Typhimurium (serogroup B, antigen profile 1,4,5,12 : *r* : 1,2). In a study of the clonal diversity of eight NTS serovars, using enzyme electrophoresis to detect allelic polymorphisms, it was reported that *S.* Heidelberg, when compared with serovars Choleraesuis, Dublin, Derby, Enteritidis, Typhimurium, Infantis, and Newport, showed the least diversity among strains from Europe and the Americas (Beltran *et al.*, 1988). However, we detected variations between African/Tropical isolates of *S.* Heidelberg with respect to six of the SSH sequences used for PCR assays.

Two of the SSH sequences (D23734-B2 and D23734-B9) were present in all *S.* Heidelberg isolates and in none of the other *Salmonella* serovars tested in this study, including all of those in the database. Thus, these sequences may represent *S.* Heidelberg-specific markers. Genome sequences are available in the database for two American isolates, SL476 (GenBank accession number CP001120), a multidrug-resistant strain, and SL486 (ABEL01000001–ABEL1000048) a drug-susceptible strain. With respect to the distribution of SSH sequences, we found no difference between these strains. Indeed, with the exception of *Salmonella* Saintpaul and *Salmonella* Newport, all of the same-serovar genomes showed no difference in terms of SSH sequence distributions. *Salmonella* Saintpaul strain SARA23 was the only strain negative for all 14 SSH sequences, and is described as falling within the main clade of the Saintpaul serovar, whereas its partner strain SARA29, which we found to be positive for three SSH sequences (D23734-5, -14, and -B10), has been described as an outlier (http://www.jcvi.org/salmonella/index.shtml). *Salmonella* Newport SL254 and SL317 are strains from the two distinct lineages that exist within the *S.* Newport serovar (http://www.jcvi.org/salmonella). We found that both strains carry the lipoprotein-associated SSH sequence D23734-B10, but strain SL254 also carries the fimbrial-associated SSH sequence D23734-G3.

The genomes of *S. enterica* possess numerous fimbrial gene clusters implicated in host colonization and adaptation. Indeed, the genome of *S.* Typhimurium LT2 alone carries 11 fimbrial operons, some of which have been implicated directly in virulence (van der Velden *et al.*, 1998; Humphries *et al.*, 2001). The repertoire of fimbrial operons varies between serovars, with some widely distributed, but the others restricted to a limited number of serovars (Townsend *et al.*, 2001; Porwollik & McClelland, 2003). Our observations indicate that fimbrial gene clusters make a major contribution to the accessory genome of *S.* Heidelberg. The *stk* gene cluster has been reported to be specific for *S.* Paratyphi A (Edwards *et al.*, 2002). Based on analysis of the distribution of SSH sequence D23734-A1, we can add *S.* Heidelberg and *Salmonella* Kentucky to this, although our PCR assay data suggest that the operon may not be carried by all *S.* Heidelberg. We also demonstrate that the expression of the *stk* operon gene could be detected from some, but not all, of the *S.* Heidelberg isolates tested. Given the proven role of fimbriae in pathogenicity, the variable carriage of this fimbrial operon gene (*stkD*), and the variable expression even among those strains carrying the gene, leads us to the conclusion that the role of this operon merits further investigation.

Salmonella Typhi carries the fimbriae designated *tcf*, for Typhi colonizing factor (Folkesson *et al.*, 1999). It has been

reported previously that the *tcf* operon is present in *S.* Heidelberg and other invasive serovars such as Paratyphi A, Sendai and Choleraesuis (Townsend *et al.*, 2001). We identified two SSH sequences matching genes within this operon, namely *tcfA* (D23734-A5) and *tcfD* (D23734-E10). In addition, we identified an SSH sequence matching *tinR* (D23734-D4), which lies downstream of the *tcf* genes in *S.* Typhi, and encodes a transcriptional regulator (Folkesson *et al.*, 1999). However, based on the S. Heidelberg SL476 genome, it appears that this proximity does not occur in *S.* Heidelberg, and our observations suggest that the distributions of *tinR* and *tcfA* differ among both our strain collections and the serovars represented in the database. Our PCR assays indicated that the SSH sequence D23734-A5 (representing *tcfA*) was present in 100% of African, Tropical and UK isolates of *S.* Heidelberg, as well as in the two genome-sequenced isolates SL476 and SL486. Expression of *tcfA* was detected in all six isolates of *S.* Heidelberg tested.

A third fimbrial operon, represented by the SSH sequence D23734-G3, was present in serovars Enteritidis, Bovismorbificans, Saintpaul, Newport and Dublin as well as all but two *S.* Heidelberg isolates, and was expressed in all *S.* Heidelberg isolates tested, except for the UK veterinary isolate KMS1977. Expression of genes associated with a further two of the seven SSH sequences tested, encoding a putative ferrichrome-iron receptor and transcriptional regulator, respectively, were also expressed in all *S.* Heidelberg isolates except KMS1977. These observations suggest that although we found little evidence for genome content variations between invasive and gastrointestinal isolates of *S.* Heidelberg, we did find variations in gene expression. It is possible that differences in gene expression play a role in the different pathogenic abilities exhibited by isolates of this serovar.

Acknowledgements

We would like to acknowledge the guidance and support of Professor Tony Hart, who sadly died in September 2007. We would also like to thank Martin Okong, Chris Parry, John Cheesbrough, Melita Gordon, John Wain, Paul Wigley and Sam Kariuki for providing strains. This work was funded by the Medical Research Council in the form of a capacity building studentship for C.B.

References

Agron PG, Walker RL, Kinde H, Sawyer SJ, Hayes DC, Wollard J & Andersen GL (2001) Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar enteritidis. *Appl Environ Microb* **67**: 4984–4991.

Bahwere P, Levy J, Hennart P, Donnen P, Lomoyo W, Dramaix-Wilmet M, Butzler JP & De Mol P (2001) Community-acquired bacteremia among hospitalized children in rural central Africa. *Int J Infect Dis* **5**: 180–188.

Beltran P, Musser JM, Helmuth R, Farmer JJ III, Frerichs WM, Wachsmuth IK, Ferris K, McWhorter AC, Wells JG & Cravioto A (1988) Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S.* choleraesuis, *S.* derby, *S.* dublin, *S.* enteritidis, *S.* heidelberg, *S.* infantis, *S.* newport, and *S.* typhimurium. *P Natl Acad Sci USA* **85**: 7753–7757.

Berkley JA, Lowe BS, Mwangi I *et al.* (2005) Bacteremia among children admitted to a rural hospital in Kenya. *N Engl J Med* **352**: 39–47.

Burt CR, Proudfoot JC, Roberts M & Horowitz RH (1990) Fatal myocarditis secondary to *Salmonella* septicemia in a young adult. *J Emerg Med* **8**: 295–297.

Chiu CH, Chuang CH, Chiu S, Su LH & Lin TY (2006) *Salmonella enterica* serotype Choleraesuis infections in pediatric patients. *Pediatrics* **117**: e1193–e1196.

Demczuk W, Soule G, Clark C, Ackermann HW, Easy R, Khakhria R, Rodgers F & Ahmed R (2003) Phage-based typing scheme for *Salmonella enterica* serovar Heidelberg, a causative agent of food poisonings in Canada. *J Clin Microbiol* **41**: 4279–4284.

Edwards RA, Olsen GJ & Maloy SR (2002) Comparative genomics of closely related salmonellae. *Trends Microbiol* **10**: 94–99.

Folkesson A, Advani A, Sukupolvi S, Pfeifer JD, Normark S & Lofdahl S (1999) Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. *Mol Microbiol* **33**: 612–622.

Galanis E, Lo Fo Wong DM, Patrick ME, Binsztein N, Cieslik A, Chalermchikit T, Aidara-Kane A, Ellis A, Angulo FJ & Wegener HC (2006) Web-based surveillance and global *Salmonella* distribution, 2000–2002. *Emerg Infect Dis* **12**: 381–388.

Gordon MA (2008) *Salmonella* infections in immunocompromised adults. *J Infect* **56**: 413–422.

Gordon MA, Walsh AL, Chaponda M, Soko D, Mbvwinji M, Molyneux ME & Gordon SB (2001) Bacteraemia and mortality among adult medical admissions in Malawi – predominance of non-typhi salmonellae and *Streptococcus pneumoniae*. *J Infect* **42**: 44–49.

Graham SM (2002) Salmonellosis in children in developing and developed countries and populations. *Curr Opin Infect Dis* **15**: 507–512.

Green SD & Cheesbrough JS (1993) *Salmonella* bacteraemia among young children at a rural hospital in western Zaire. *Ann Trop Paediatr* **13**: 45–53.

Humphries AD, Townsend SM, Kingsley RA, Nicholson TL, Tsolis RM & Baumler AJ (2001) Role of fimbriae as antigens and intestinal colonization factors of *Salmonella* serovars. *FEMS Microbiol Lett* **201**: 121–125.

Kang MS, Besser TE, Hancock DD, Porwollik S, McClelland M & Call DR (2006) Identification of specific gene sequences conserved in contemporary epidemic strains of *Salmonella enterica*. *Appl Environ Microb* **72**: 6938–6947.

Kariuki S, Revathi G, Gakuya F, Yamo V, Muyodi J & Hart CA (2002) Lack of clonal relationship between non-typhi *Salmonella* strain types from humans and those isolated from animals living in close contact. *FEMS Immunol Med Mic* **33**: 165–171.

Kariuki S, Revathi G, Kariuki N, Muyodi J, Mwituria J, Munyalo A, Kagendo D, Murungi L & Hart CA (2005) Increasing prevalence of multidrug-resistant non-typhoidal salmonellae, Kenya, 1994–2003. *Int J Antimicrob Ag* **25**: 38–43.

Kariuki S, Revathi G, Kariuki N, Kiiru J, Mwituria J, Muyodi J, Githinji JW, Kagendo D, Munyalo A & Hart CA (2006) Invasive multidrug-resistant non-typhoidal *Salmonella* infections in Afric: zoonotic or anthroponotic transmission? *J Med Microbiol* **55**: 585–591.

Lavy CB, Thyoka M & Pitani AD (2005) Clinical features and microbiology in 204 cases of septic arthritis in Malawian children. *J Bone Joint Surg Br* **87**: 1545–1548.

McClelland M, Sanderson KE, Spieth J *et al.* (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852–856.

Porwollik S & McClelland M (2003) Lateral gene transfer in *Salmonella*. *Microbes Infect* **5**: 977–989.

Rice PA, Craven C & Wells JG (1976) *Salmonella* Heidelberg enteritis and bacteremia. An epidemic on two pediatric wards. *Am J Med* **60**: 509–516.

Thomson NR, Clayton DJ, Windhorst D *et al.* (2008) Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res* **18**: 1624–1637.

Threlfall EJ, Hall ML & Rowe B (1992) *Salmonella* bacteraemia in England and Wales, 1981–1990. *J Clin Pathol* **45**: 34–36.

Threlfall EJ, Ward LR, Frost JA & Willshaw GA (2000) The emergence and spread of antibiotic resistance in food-borne bacteria. *Int J Food Microbiol* **62**: 1–5.

Townsend SM, Kramer NE, Edwards R *et al.* (2001) *Salmonella enterica* serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect Immun* **69**: 2894–2901.

van der Velden AW, Baumler AJ, Tsolis RM & Heffron F (1998) Multiple fimbrial adhesins are required for full virulence of *Salmonella typhimurium* in mice. *Infect Immun* **66**: 2803–2808.

Vugia DJ, Samuel M, Farley MM, Marcus R, Shiferaw B, Shallow S, Smith K & Angulo FJ (2004) Invasive *Salmonella* infections in the United States, FoodNet, 1996–1999: incidence, serotype distribution, and outcome. *Clin Infect Dis* **38**: S149–S156.

Wilmshurst P & Sutcliffe H (1995) Splenic abscess due to *Salmonella* Heidelberg. *Clin Infect Dis* **21**: 1065.

Winstanley C (2002) Spot the difference: applications of subtractive hybridisation to the study of bacterial pathogens. *J Med Microbiol* **51**: 459–467.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Oligonucleotide PCR primers used in this study.
Table S2. Summary of SSH.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.