Genomic variation of *Salmonella* Typhimurium and dynamics of epidemics

Cristina Fabiola SOTOMAYOR CASTILLO

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Medicine

Sydney Medical School

The University of Sydney

2017

Declaration of Author

I, Cristina Fabiola Sotomayor Castillo, declare that the contents of this thesis consist of original work carried out by the author unless otherwise stated and duly acknowledged. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

Cristina Fabiola SOTOMAYOR CASTILLO

April 2017

“There’s a million things I haven’t done, but just you wait, just you wait”

Lin-Manuel Miranda, “Hamilton, An American Musical”

Acknowledgments

To my supervisor, Dr. Vitali Sintchenko, I will be always grateful for the chance you took by accepting an unknown overseas veterinarian to carry out a project that was so away from my comfort zone and knowledge. You pushed me beyond the fields I had known, and I landed in the genomics world. Never in my dreams would I have thought to learn and be able to understand this world and its concepts. Your kind supervision, support and patience during this period are what I take with me.

My learning curve would have never been the same without the supervision provided by other hospital and university scientists. Dr Qinning Wang, your patience and knowledge, which you have never been selfish with and your permanent support in all the laboratory-based activities were fundamental. I wouldn't have done it without you. You always made me feel good even know at the beginning I had no clue where I was standing in the genomics world of *Salmonella*. I have enjoyed our days at the lab in a way you will never know, and trust me, it is all because of your help and kindness. Elena, you were there to support me when Qinning couldn't due to his daily work. I have learned so much from you and I am lucky to call you my friend. To the Sintchenko's post-docs crew, some of the currently here, some of them moved on: Nadine Holmes, Chayanika Biswas, Ranjeeta Menon, Rebecca Rockett, Verlaine Timms, Rajat Dhakal, each one of you were kind enough to provide support at some stage. Alicia, thank you for all the suggestions and comments at latest stages. Your help made a huge difference. ERL people including Peter Howard, Agnieszka W., and many others through these years, thank you not only for putting up with my presence and questions invading your lab daily.

My PhD friends, Kaitlin, Dave, Andrea B., Vyoma, Ravi, Mai, Alex, Winkie and so many others; Siberia room would have never been the same without you. I am glad I got to know you. We had weddings, babies and so many other adventures together on the way. So many people at ICPMR were by my side on this journey and helped along the way. I will always remember you and be grateful to the time and help you provided. It is an odd feeling arriving to a lab that is as busy as ICPMR, trying to do some work on a field you have never been related to and not knowing where a single pipette is, less to use it. Each one of you took time to help me and I will never forget it.

I also must thank for the opportunity to come to Australia and pursue my PhD. Becas Chile-CONICYT scholarship is the reason why I am here. Thank you for this opportunity.

Last but not least, to my adored husband Claudio and my children. Claudio, you have been with me from the moment you pushed me to apply for the scholarship I never thought I was going to get; you never doubted my capacities, even though I did many times. You sacrificed all that you had accomplished back in Chile to support me in this journey and I could have not done with anyone else by my side. I will always be grateful for your love and support. You are the best partner a woman could have, and I don't only mean because of your awesome cooking skills. The level of patience the partner of a PhD candidate must have is under estimated, and though I know you could not stand me these past months you stayed by my side, loving me, taking over things so I could simply focus on my work. I hope you are not too tired with this last couple of months (I have been a pain, I know, but I will make it up for the rest of our days). My children, Cristobal and Laura; you have had to endure the absence of mum this last part of the journey. I love you so much for your support. Your hugs saying, "you can do it mummy" will last me a lifetime. Thank you, thank you all.

Abstract

Foodborne diseases affect one in six members of the population every year and cause at least 1.5 million deaths annually. Infections acquired through food consumption are common with significant morbidity and mortality costs to the society. Non-typhoidal salmonellosis (NTS) is responsible for the significant proportion of foodborne gastroenteritis worldwide. Molecular identification and characterization of *Salmonella* with assessment of exposure and epidemiological analysis of outbreak-linked cases are main approaches to control the NTS. In this study we focus on *Salmonella* Typhimurium (STM) as the most common causative agent of foodborne NTS in Australia and a frequent cause of community outbreaks.

The specific aim of this thesis is to examine and explain temporal dynamics of STM incidence, using New South Wales (NSW), Australia, as an exemplar region with good access to pathology services as well as an established system of public health surveillance of this notifiable infection. An important part of achieving this goal is the understanding of within- and between-host variations and adaptations in STM genomes and discrimination power of rapidly evolving typing methods. We hypothesized (i) that the application of evolutionary approaches to the analysis of specific clades of STM improves the understanding of salmonellosis epidemic dynamics, and (ii) that polymorphism in STM genomes is a key attribute that, in conjunction with epidemiological evidence, can affect the recognition and investigation of STM transmission events and community outbreaks.

We have examined 11,799 STM isolates recovered from human and environmental samples between 2009 and 2016. Our findings suggest that multi-locus variable sequence typing (MLST) can be successfully applied for molecular serotyping of *Salmonella* isolates circulating in NSW. However, MLST approach lacks discriminatory power to be useful for public health

surveillance. In contrast, multi-locus variable number tandem repeat analysis (MLVA) identified major clades associated with extensive epidemics over the years of this study. Only a small number of MLVA profiles have been associated with particular outbreaks or clusters, masking the diversity of profiles and reducing the potential of epidemiological investigations to elucidate transmission networks. Our retrospective sequencing of STM isolates, including ones associated with a large community outbreak, reconfirmed previously postulated high-resolution of whole genome sequencing (WGS) and the ultimate discriminatory power with the potential to enhance epidemiological investigations and elucidate transmission pathways. Our findings illustrated a relatively constant core genome for STM population over time, translated in stable diversity with predominance of endemic STM MLVA profiles. The adaptive evolution of STM in association within its host, particularly in the experimental model of chronic salmonellosis in mice, was one of the key findings. It involved a limited number of mutations, which did not compromise the ability of STM to maintain the infection and the excretion of STM in stools. The temporal relation between the incidence of STM infections in NSW and the corresponding increase of particular STM clades was unveiled. Our findings indicated that reduction of newly identified MLVA profiles of STM in winter and spring precedes the high activity of human infection in summer.

The comparative genomic analysis performed on carefully selected representatives of sporadic and epidemic STM clades identified genomic polymorphisms within the successful clades when compared to non-successful ones. These observations emphasize the stability of accessory genomes, which was previously underappreciated, but require further *in vivo* validation. Our research findings provided important insights into the dynamics of community epidemics of *Salmonella* Typhimurium and identified types of genomic polymorphisms that contribute to the establishment of epidemic clades. Our results and analyses have also offered important evidence

to guide the interpretation of *Salmonella* Typhimurium public health laboratory surveillance and the translation of whole genome sequencing into more effective control of foodborne diseases.

Table of Contents

Publications and conference presentations associated with the thesis

Phillips A, Sotomayor C, Holmes N, Furlong C, Ward K, Howard P, Octavia S, Lang R, Sintchenko V. Whole genome sequencing of *Salmonella* Typhimurium illuminates distinct outbreaks caused by an endemic multi-locus variable number tandem repeat analysis type in Australia, 2014. *BMC Microbiology* (2016) 16:211.

Wang Q, Sotomayor C, Dhakal R, Menon R, Howard P, Shadbolt C, Hope K, Sintchenko V. The Added Value of Whole Genome Sequencing in the Investigation of *Salmonella* Outbreaks. The Broad Street Pump, Centre for Infectious Disease and Microbiology-Public Health, October 2016, Issue 44.

2015 Cristina Sotomayor, Qinning Wang, Peter Howard, Vitali Sintchenko. Temporal dynamics of human *Salmonella enterica* serovar Typhimurium in New South Wales. Oral presentation at Marie Bashir Institute for Infectious Diseases and Biosecurity Colloquium, Sydney, Australia, November 2015.

2015 Cristina Sotomayor, Qinning Wang, Peter Howard, Vitali Sintchenko. Within-host and between –host genome variation of *Salmonella Typhimurium*: experimental study. Oral presentation at the Centre of Infectious Diseases and Microbiology – Public Health Colloquium, Westmead, November 2015.

2015 Cristina Sotomayor, Qinning Wang, Peter Howard, Vitali Sintchenko. Genomic variation findings presented during acute salmonellosis in balb/c mice Oral presentation at Marie Bashir Institute for Infectious Diseases and Biosecurity Colloquium, Sydney, Australia, November 2015.

2015 Cristina Sotomayor, Qinning Wang, Peter Howard, Vitali Sintchenko. Recent evolution of human *Salmonella enterica* serovar Typhimurium in New South Wales. Communicable Diseases Control Conference, Brisbane, Australia, May 2015.

2014 Cristina Sotomayor, Qinning Wang, Peter Howard, Vitali Sintchenko. Utility of multilocus sequence typing (MLST) for inferring prevalent *Salmonella* enterica serotypes. Oral presentation at the Australian Society for Microbiology Conference ASM 2014. 6^{th} – 9^{th}, Melbourne, Australia.

Scholarships and Awards

2013 – 2016 Becas Chile/CONICYT scholarship

2015 Westmead Charitable Trust Foundation Travel Scholarship

2014 Westmead Charitable Trust Foundation Travel Scholarship

List of Figures

List of Tables

List of Abbreviations

BA	- Blood Agar
BP	- Base pairs
CDC	- Centres for Disease Control and Prevention
CFU	- Colony forming unit
DNA	- Deoxyribonucleic Acid
ERL	- Enteric Reference Laboratory
EU	- European Union
FERG	- Foodborne Disease Burden Epidemiology Reference Group
MLST	- Multilocus sequence typing
MLVA	- Multilocus variable number tandem repeat analysis
NCBI	- National Centre for Biotechnology Information
NGS	- Next generation sequencing
NSW	- New South Wales
PCR	- Polymerase chain reaction
PFGE	- Pulsed field gel electrophoresis
PT	- Phage type
PHE	- Public Health England
SNP	- Single nucleotide polymorphism
SPI	- *Salmonella* pathogenicity island
ST	- Sequence type
STM	- *Salmonella enterica* serovar Typhimurium

µL	- Microlitre
WGS	- Whole genome sequencing

Chapter 1: Introduction

1.1 Foodborne illness: an on-going public health concern

Foodborne diseases affect one in six members of the population every year and cause at least 1.5 million deaths annually (1). This impact is amplified by the globalization of food markets and emerging fresh produce industries. Infections acquired through food consumption can result in clinical complications, which can lead to hospitalization and poor clinical outcomes, usually the very young and the elderly. Changes associated with the spectrum of illness and the food products involved in transmission have taken place. Particularly in terms of current food product demand and supply, dramatic changes have occurred (2). Globalization of the food supply, with large volumes of food being imported from distant countries has facilitated widespread access to a more diverse food selection, resulting in food processing becoming highly industrialized (3).

Animal production for instance, has increased to cover the population's demand which translates into food production animals increasingly raised in close quarters for later slaughtering and processing as part of food production chains. This handling method increases the incidence and transmission of diseases. Developed countries have implemented surveillance systems, however due to limited resources, low-income countries cannot implement the same level of food chain surveillance (4-6). Agriculture has also been impacted by globalization of the food supply, particularly with production of fresh fruits and vegetables which are currently available year round, often transported from warmer countries contributing to the emergence of new pathogens and new trends in foodborne illness presentation (7-8).

Although foodborne diseases are clearly an important cause of morbidity and mortality worldwide, there remains a lack of understanding of their burden, associated risks and pathways of transmission, despite global awareness of foodborne diseases posing a high risk to health and economic development (9).

1.2 The burden of non-typhoidal salmonellosis

Non-typhoidal salmonellosis (NTS) illness is a major cause of foodborne gastroenteritis worldwide. The diverse serovars of NTS are a leading bacterial cause of morbidity and mortality, both in children under 5 years old (10) and in the population in general (11). Infection with NTS has been estimated to cause 93 million enteric infections (90% CI: 61.8–131.6 million) and 155,000 diarrheal deaths each year (12). Despite public health efforts, it remains a growing challenge for both developing and developed countries. The causative agents, members of *Salmonella* spp., colonize a variety of diverse animal reservoirs and are associated with various routes of transmission, making control efforts extremely difficult. In developed countries however, the majority of infections are associated with foodborne exposures (13).

In order to reduce the burden of foodborne salmonellosis, timely recognition of disease outbreaks and accurate identification of the food sources causing disease are crucial. Molecular subtyping of *Salmonella* combined with epidemiological analysis of exposures and outbreak-linked cases are widely used to estimate the relative contribution of different food sources to NTS (14). However, and despite the fact that public health unit and laboratory based surveillance provide useful information, the true NTS burden remains underestimated (15). In order for public health unit and laboratory-based surveillance systems to provide a reliable overview of the population incidence of NTS, an ill patient must first seek medical assistance; only by this initial step a specimen, usually stools, will be collected and submitted to a pathology provider. From this

point, the laboratory in charge of the specimen's analysis will test for enteric pathogens and, if the particular organism is found, a report confirming the positive finding will be generated for the requesting medical practitioner and forwarded to the relevant public health authority (16-17). Jurisdictional Public Health Acts require mandatory reporting of all laboratory confirmed *Salmonella* infections to the relevant health departments and then to the National Notifiable Diseases Surveillance System (NNDSS). Due to the previously mentioned underrepresentation of the total number of cases of salmonellosis, it has been estimated that there are at least 7 salmonellosis cases (95% CI 4–16) occurring in the community for every notification to health departments from pathology providers (18). Hence, notified cases detected by laboratory-based surveillance only represent a fraction of the total community cases of NTS. In terms of global surveillance data, the situation becomes even more difficult since a large number of countries, particularly developing countries, present large variations in their implemented surveillance standards. Several prospective and retrospective studies attempted to establish the notification levels associated with laboratory-based surveillance worldwide, confirming the level of sub-notification involved with NTS cases (15-16, 9).

An additional public health issue is the presence of antibiotic resistance in *Salmonella* serovars. Most infections with non-typhoidal *Salmonella* infections (NTS) are self-limiting, and do not require any antibiotic therapy (17-19). In animals, however, antibiotics are used for several purposes including therapeutics and growth promotion. It is this extensive use, or misuse, that drives development of resistant organisms worldwide, which is emerging as a serious public health threat (20). The need to monitor resistance of *Salmonella* in food products is a necessary safety strategy, as well as the necessary reduction on the use of antibiotics in food production animals (21-22).

More than 2,500 serovars of *Salmonella* enterica can cause NTS. While all these serovars can cause infection in humans, most human gastroenteritis is caused by a limited number of serovars. *S.* Enteritidis and STM, are among the most frequently identified serovars associated with NTS human illness, reaching around 70% of identified cases in the European Union (EU) (6-7,9,23); however, *Salmonella enterica* subsp. enterica serovar Typhimurium (STM) is the predominant serovar in Australia (24-27). The situation with NTS in Australia is similar to the descriptions worldwide; approximately 72% of salmonellosis in Australia is transmitted through contaminated food (28-29). Common foods associated with salmonellosis in outbreak investigations and source attribution studies have included eggs, poultry meat, pork, beef, dairy products, nuts, and fresh produce (28,30-31). *S.* Enteritidis is the most commonly identified serovar after *S.* Typhimurium (28). This former predominates in Western Europe and North America, where laying contaminated hens for egg production have been confirmed as the primary source of human S. Enteritidis infection (6,34), *S.* Enteritidis is not endemic in poultry in Australia. According to the OzFoodNet report (2015), most human infections with *S.* Enteritidis in Australia, have been judged as overseas acquired (35). There are several other serovars present in Australia associated with relatively small numbers of human cases; public health authorities have described them as belonging to particular ecological niches within the country, mostly related to its wide climatic and geographical variation (35-37).

1.3 STM as the main causative agent of NTS identified in New South Wales

Over three thousand *Salmonella* isolates are received and tested every year by the New South Wales Enteric Reference Laboratory (ERL) based at the Pathology West-ICPMR at Westmead Hospital in Sydney, and around half of them have been serotyped as STM. The presence of such large numbers of a single serovar and historically made identification of potential outbreaks caused by STM extremely difficult, justifying the need for more discriminatory subtyping methods (38-39). The multi-locus variable number tandem repeat analysis (MLVA) was introduced to support public health investigations of foodborne salmonellosis outbreaks. MLVA measures the variable length of five STM loci and has been considered faster and more discriminatory than other historical typing methods (40-44).

Representatives from several Australian reference laboratories agreed on the convention by which MLVA typing for STM was implemented. During 2008, the NSW ERL, in collaboration with the Communicable Diseases Branch of the NSW Ministry of Health and the NSW Food Authority, prospectively evaluated MLVA typing by comparing the results available within approximately 2 weeks of the receipt of isolates with epidemiological investigations of suspected clusters (40). Evaluation of results confirmed the higher discriminatory power of MLVA typing when compared to other typing methods such as serotyping and the no longer implemented phage typing. Since 2008, MLVA typing has been implemented in NSW to prospectively subtype STM and identify potential clusters (40-42).

However, endemic MLVA profiles may cause multiple outbreaks along with sporadic cases and a few selected MLVA profiles may represent a large portion of isolates observed in a geographical location, showing that there are cases when the resolution provided by MLVA is no longer sufficient to generate epidemiological links (45).

The population of STM isolates in NSW between 2008 and 2011 remained largely stable in terms of main MLVA profiles dominating the STM spectrum, with almost 50% of isolates classified as belonging to a single phage type (i.e. PT170). Almost a third of these isolates possessed one of three related MLVA profiles (3-9-7-13-523, 3-9-7-12-523 and 3-9-8-12-523), profiles were associated with the largest numbers of STM cases until 2012, typically displaying seasonal peaks related to the warmer months (42). Hence, a large number of STM cases are caused by successful endemic MLVA profiles (34), which may cause multiple outbreaks along with sporadic cases, reducing the potential of epidemiological investigations to elucidate transmission networks (42,45).

1.4 Utility of Whole Genome Sequencing in public health investigations

Whole Genome Sequencing (WGS) offers the ultimate discriminatory power with the potential to enhance epidemiological investigations and elucidate disease transmission pathways. Its advantages over other pathogen characterization methods have been proven for *Salmonella* and other organisms, offering high throughput and high-quality data which are easy to share and interpret between laboratories, nationally and internationally (46-50). Furthermore, it has been shown that WGS can complement existing epidemiological strategies by identifying previously undetected linked cases (51-53).

Recent studies have also demonstrated that sequencing of bacterial genomes can detect chronic carriers, predict the existence of undiagnosed cases within transmission chains, suggest potential routes of transmission as well as identify unrecognized risk factors associated with transmission events (54-56). Evidence suggests that WGS represents a paradigm shift in identification and characterization of foodborne pathogens, including *Salmonella spp.*, with

capacity to replace some of the currently in use diagnostic techniques (57). This will require the development of faster result interpretation systems as well as the development of systems to synthesize share and store the generated genome data at a public health level (58-60).

Nevertheless, availability of clinical and epidemiological information is required to confirm WGS findings regarding genetic relatedness of particular isolates. The synthesis of microbiological, genomic and epidemiological lines of evidence is fundamental for the understanding of dynamics on pathogen's epidemics, including STM, in order to improve the public health surveillance and control of NTS (61-65). Whole genome sequencing has thus been suggested as a complementary tool to enhance traditional epidemiological and surveillance tools, however barriers still persist that prevent widespread global and regional adoption of this technology, mainly related to the acquisition of the technology and the costs involved and the requirement of professionals with the knowledge in bioinformatics and genomics to carry on the analysis (66-68).

1.5 Gaps in the understanding of molecular epidemiology of STM limits our disease control capacity

While the role of gene loss in the shaping of *S.* Typhi as a strict human pathogen that evolved from less virulent zoonotic *Salmonella* has been recognized (61), the genomic variation in bacteria causing NTS has received less attention. For NTS, genomic surveillance at the phylogenetic level has identified the evolution of new common ancestors of *Salmonella,* as well as identifying particular genes, which support gut colonization and faecal shedding (62). It appears that disease-causing strains tend to have a greater number of genes involved in replication, recombination and repair. This description particularly applies to STM ST313, which has been

reported as a multidrug resistant STM as well as a carrier of genomic modifications towards human host-adaptability, potentially becoming completely human adapted, changing the preventive approaches carried out for STM (61).

Recent breakthroughs in genome-wide association studies have analysed the presence of strain-level differences in phenotypes of diverse recombining bacteria, not *Salmonella,* and their effects on bacterial population structure, and have focused on identification of multiple loci within genomes that collectively show a strong association with phenotype (62). However, these methods have not been applied to further understanding of endemic and epidemic clades of STM, which represents a significant gap in the collective knowledge and hence is a major focus of this thesis.

1.6 Aims of the thesis and research hypothesis

The overarching aim of this thesis was to explain temporal dynamics of STM, using New South Wales in Australia as an exemplar region with good access to pathology services as well as an established system of public health surveillance of this notifiable infection. An important part of achieving this aim was to understand within and between-host variation and adaptation in STM genomes by harnessing the discriminatory power of rapidly evolving typing methods. Our *hypotheses* were that:

(H1) the application of evolutionary approaches to the analysis of specific clades of STM improves the understanding of salmonellosis epidemic dynamics

(H2) polymorphism in STM genomes is a key attribute that, in conjunction with epidemiological evidence, can affect the recognition and investigation of STM transmission events and community outbreaks.

1.7 Research objectives

The *specific objectives* of this study were to:

1. Explore the capacity of multi-locus sequence typing (MLST) to identify and differentiate the most common serovars of *Salmonella enterica* co-circulating in NSW (Chapter 5)
2. Investigate the structure and mechanisms of seasonal epidemics of human STM disease in NSW (Chapter 6)
3. Characterize attributes of STM populations that are associated with seasonal epidemics of human STM disease in NSW (Chapter 6)
4. Compare the resolution power of MLST, MLVA and whole genome sequencing for public health laboratory surveillance (Chapters 5 and 6)
5. Examine variations in core and accessory genomes of successful STM clades (Chapter 7)
6. Identify within-and between-host variations and adaptations in STM genomes in acute and chronic infection models (Chapter 8).

1.8 Thesis outline

The thesis is divided into 3 sections. The first section provides a background to the study and contains three chapters (Chapters 1, 2 and 3). The introductory chapter (**Chapter 1**) states the aims, research hypothesis and objectives of the thesis. **Chapter 2** provides an overview of the

burden of foodborne salmonellosis in terms of its global significance, followed by characterization and taxonomy of the *Salmonella* spp., its pathogenic determinants and a description of challenges in laboratory identification as well as the public health surveillance strategies and tools applied to this organism. In **Chapter 3,** the added value of STM genotyping describing the *Salmonella* genus was addressed, as well as describing other contemporary genotyping methodologies such as MLVA-5 typing, to later on fully embrace the description and applicability of WGS for pathogen detection and analysis of either individual non-linked cases or, with further utility, in outbreak investigation.

Section two of this thesis focuses on presentation of the strategies performed to fulfil each one of the research objectives. The material and methods chapter (**Chapter 4)** details each of the different approaches, methodologies and techniques applied to address the hypotheses and objectives of this study.

Section three of the thesis corresponds to the presentation of results and its corresponding analysis (Chapters 5, 6, 7, 8, 9). **Chapter 5** focuses mainly on the utility of MLST as a tool for inferring *Salmonella* serovars as well as investigation of the general mechanisms involved in seasonal STM human disease in NSW. **Chapter 6** approaches the main characteristics found within seasonal STM epidemics in NSW as well as comparing the resolution power of three genotype-based identification methods. The analyses presented in **Chapter 7** focus on a set of successful epidemic STM clades, examining variations in core and accessory genomes of these successful STM representatives while comparing them to a set of non-successful STM clades. Finally, **Chapter 8** describes findings generated using two separate animal models for STM infection, one acute and one chronic, identifying variations and adaptations in STM genomes.

This thesis concludes with **Chapter 9,** which summarizes conclusions from the previous chapters, explaining how the findings presented in this thesis can improve characterization of STM infection in NSW by harnessing the emerging field of genomic epidemiology to increase the understanding in terms of STM epidemic dynamics to improve public health laboratory surveillance. By better understanding the nature and molecular markers within successful STM clades, we may be able to streamline the evaluation and implementation of genomic-based surveillance to improve the recognition of transmission pathways and better control of foodborne diseases.

Chapter 2: The burden of foodborne salmonellosis

2.1 Introduction

Infection caused by foodborne pathogens is characterized by gastroenteritis with diarrhoea and related symptoms. There are more than 250 different foodborne diseases described worldwide. *Campylobacter*, Shiga toxin–producing *Escherichia* coli (STEC), nontyphoidal *Salmonella*, and *Listeria* are the main associated causative agents (4,7,9,11,13,69). They are an important cause of morbidity and mortality and a significant impediment to socioeconomic development worldwide. The actual assessment of this foodborne diseases burden is complex: many different pathogens can be transmitted by food, leading to widely different health outcomes (8-9,16,55,70).

Many cases of foodborne illness are self-limiting, self-treated and may not be recognized by the health system. General practitioners and others may treat some, though few, may develop complications that will require extensive treatment and hospitalization beyond the year in which the infection occurred (3). Only a small proportion of those cases are recognized as caused by a hazard associated with food and therefore treated, reported to public health authorities and recorded as part of official statistics for each of those countries (20).

The global community has witnessed changes within foodborne pathogens regarding type of illness, severity and the impact they have in the population worldwide. These changes are even more diverse across regions, countries and communities (2,8-10,7-74). The OzFoodNet Network has described the consequences of diseases potentially transmitted by food in Australia, including salmonellosis; absenteeism from work and the associated costs involved in the actual clinical condition translate into a recognizable burden on the economy (36,37).

This chapter examines the general burden of foodborne disease caused by *Salmonella* and its complexity, including the variety of pathogens member of this group, which can be transmitted through food causing a diverse array of resulting health outcomes. Its analysis would potentially contribute to the policy makers at the individual, population, government and industry level to be aware of the persistence of this situation and the need of more accurate management decisions in terms of control, prevention and surveillance.

2.2 Global significance of bacterial foodborne diseases

In 2007, the World Health Organization (WHO) established the Foodborne Disease Burden Epidemiology Reference Group (FERG), aiming to provide reliable estimates of the global impact of all food-borne diseases. Despite a lack of data from China, Latin America and the Middle East, it was reported that more than 5 billion episodes of diarrhoea were recorded annually among the <5 years old global population, resulting in an estimated 1.15 million deaths a year in South East Asia and Africa (8-9,12). But the problem was not limited to developing countries as might be expected due to the nature of foodborne illness: 419 million episodes of diarrhoea were recorded annually in Europe and 455 million in North America, with under-reporting acknowledged as a limitation of these estimates (8-9,12).

This latest report included data analysis between 2007 and 2015, stating foodborne illness as a public health problem of major significance. Even more, it has been made clear that certain chronic diseases that result from contaminated food consumption can appear long after ingestion; as expected, the causal link is never made for these cases (9,12).

2.3 Global role of foodborne Salmonellosis

Salmonellosis is an important public health problem cause of important morbidity, in addition to also having a significant economic impact worldwide (13,15,75). Although most infections cause mild to moderate self-limited disease, death occurs in those patients with severe infections. In the United States (US) it is estimated that 1.4 million non-typhoidal *Salmonella* (NTS) infections with 400 deaths occurs annually (8-9,12,15,76). The infective dose is usually high, but the bacteria grow well in most food. In food with a high fat content, e.g. chocolate and cheese, the infective dose is very low, and just a few bacteria may be sufficient to cause infection (77).

The patient's susceptibility to infection is variable. For instance in infants, elderly or immune-compromised hosts, the critical infective dose is lower. The onset of disease is often classified as acute, with an incubation period of 1 – 3 (range less than 1 – 10) days, involving general gastro enteric symptoms such as diarrhoea, nausea and vomiting (77). The carrier state is normally 4–6 weeks, but a few cases may be asymptomatic carriers for months or even years. No vaccine is available against non-typhoidal salmonellosis (78). Infection with NTS is generally limited to a specific intestinal event. However, the presence of virulence plasmids has been associated with non-typhoidal *Salmonella* spp. surviving in phagocytes and spreading from the small intestine to the spleen and liver (79). *Salmonella* species are also capable of producing a heat-labile enterotoxin, resulting in the loss of intestinal fluids, causing diarrhoea. This enterotoxin is closely related functionally, immunologically and genetically to the toxin of *Vibrio cholerae* and the heat labile toxin of pathogenic *Escherichia coli* (72,80).

According to the Food Review from the Foodborne and Diarrheal Diseases Branch at the USA Centres for Disease Control and Prevention (CDC), the global human health impact of non-typhoidal *Salmonella* infection (NTS) can be as high as 108 cases and over 155,000 deaths yearly. Many of these cases could be prevented by rapidly identifying sources, which would translate into early public health interventions (12,15,71,74). The actual costs associated to this event are clear through several reports from different countries worldwide. The annual cost of foodborne illnesses in Sweden, according to was estimated to be 1,082 million Swedish krona (US$123 million) (81). Other authors also described that the Danish society saved U.S. $25.5 million by controlling *Salmonella* in 2001 (5,11,76). In Australia, the total cost of foodborne illness according to the Australian Government Department of Health and Ageing (2006) is estimated at $1,249 million per annum and for *Salmonella* infections in particular it costs an estimated $1,387 per notified infection. There is thus a global need to focus efforts on reducing transmission of *Salmonella* infection by food and other routes (24,36-27-28).

Salmonellosis is described in the latest FERG report (WHO, 2015) as one of the most important causes of food-borne disease, resulting in significant economic burden affecting the population and health care systems. Accurate estimates of the burden of diarrheal diseases caused by *Salmonella* species and other foodborne pathogens are needed to effectively set public health goals and allocate resources to reduce disease burden (9,12). In general, but also applicable to the *Salmonella* related cases, laboratory-based surveillance provides useful, however only estimated based on confirmed cases, information. Epidemiological surveillance alone cannot confirm cases, rather providing estimates of probable cases that are subsequently confirmed by laboratory testing (82-83).

For information associated with an illness to be recognized within a laboratory-based surveillance system, patients must seek medical assistance and submit a specimen for laboratory testing (83). The laboratory must test for the pathogen and report a positive finding to the relevant public health authorities. Taking all these steps into consideration, it is not difficult to conclude that cases in laboratory-based surveillance represent only a small fraction of the total community cases (84-86). These cases, which actually get access to testing, are usually associated only to specific subgroups within the general population and thus do not accurately represent the global burden for the condition. Regardless the region, country or population involved, reportable disease data are a key information source for determining the epidemiology of notifiable illnesses (87). Current global connections based upon travelling and migration; food production chain; environmental factors as well as sociocultural elements associated to each community and country are additional factors involved in the presented picture (85-86,88-89).

2.4 Taxonomy of *Salmonella*

Salmonella are motile, non-spore forming, Gram negative, facultative anaerobic bacilli, usually of 0.7-1.5 x 2.5 μm in size, which belong to the family *Enterobacteriaceae* (105). Members of the genus *Salmonella* are described as able to grow well on simple media but almost never ferment lactose or sucrose. Optimum pH for growth between 6.5 – 7.5, however *Salmonella*e can proliferate in the range of pH 4.5 – 9.5 (87). They form acid and sometimes gas from glucose and mannose (106-107). They usually produce hydrogen sulphide (H_2S) on triple sugar iron and decarboxylate lysine and ornithine, and able to hydrolyse indole and urea. Also, they are oxidase negative and catalase positive, generally able to reduce nitrate to nitrite, to grow on citrate as sole carbon source (107).

According to the literature, *Salmonella*e survive freezing in water for long periods of time and are resistant to brilliant green, sodium tetrathionate and sodium deoxycholate, which inhibit other enteric bacteria (108-109).

The role of *Salmonella* in foodborne disease was first documented in the late 1800s when the bacterium was discovered by Eberth and cultured by Gaffky in 1885 (45). However, the association between *Salmonella* and human clinical disease, in the form of typhoid, dates back to the beginning of that century. *Salmonella* was originally designated *Bacillus* cholerae-suis and later isolated by an American veterinary pathologist, D.E. Salmon, from pigs suffering hog cholera. Similar organisms (*Bacillus* cholerae-suis) had previously been isolated from outbreaks of foodborne disease and infected animals. It was Salmon, together with Smith, a co-worker, who first successfully isolated the subsequently named *Salmonella* cholerasuis from pigs (45). In order to classify these organisms and in honour of D.E. Salmon, the genus *Salmonella* was created by Lignières in 1900.

Salmonella can be serologically classified into "serovars", also called serotypes, on the presence and/or absence of O (somatic) and H (flagella) antigens. The "H" antigen (the flagella) derives from the German word "hauch" (breath), first used to describe the swarming of highly motile organisms. "O" derives from the German word 'ohne' (without), first applied to non-swarming (i.e., nonfagellated) bacteria, but now used as a generic term for the LPS somatic antigens of enteric bacteria including *Salmonella*. The Vi antigen was thought to be responsible for virulence (90). The serotyping technique is based on the serologic classification determined using an array of specific antisera. Historically, although "serotype" and "serovar" have both been frequently used, according to the Rules of the *Bacteriological Code* (1990) established by the Judicial Commission of the International Committee on the Systematics of Prokaryotes, the term

"serovar" is preferred to the term "serotype". Kauffmann proposed that each serovar be considered a separate species (91-92).

Salmonella serovars identified after 1966 were designated mainly by their antigenic formula and existence of several species within the genus *Salmonella* was generally accepted. However, some clinically relevant *Salmonella*e identified before 1966 had been given specific names either according to the disease and/or the animal from which the organism was isolated, for example *S.* Typhi and *S.* Typhimurium. Other *Salmonella*e were named for the geographical area where the strain was first isolated, e.g., *S. London* and *S. Panama* (93). These names had been used for a number of years and therefore were adopted without being amended into the new antigenic formula system. Kauffman and Edwards then proposed the creation of *Salmonella* enterica, including all *Salmonellae (*94-95). Subsequently, a similar three-species model was proposed, with "*Salmonella enteritidis*" representing all serovars other than *S. Typhi* and *S. choleraesuis*. Another proposal arose in 1970, which recommended that Kauffmann's "subgenera" be considered a species, i.e., "*S. kauffmannii*" for "subgenus" I, *S. salamae* for "subgenus" II, *S. arizonae* for "subgenus" III, and *S. houtenae* for "subgenus" IV (96).

However, in 1973, on the basis of DNA-DNA hybridization experiments, it was demonstrated that all *Salmonella* strains should belong to a single species (97). In 1982, on the basis of numerical taxonomy and DNA relatedness studies, the name "*Salmonella choleraesuis*" was proposed for the single *Salmonella* species and six subspecies were defined (98-99). In 1989, a single exception was described: one of the subspecies, *Salmonella choleraesuis* subsp. *bongori*, was separated from the other subspecies as a unique *Salmonella* species due to differences demonstrated by DNA relatedness studies (100). This classification proposal using the "cholerasuis" nomenclature as a name for species and serovar caused confusion, so in 1986

"*Salmonella* enterica" was again proposed against the type species of *Salmonella* by the Subcommittee of *Enterobacteriaceae* of the International Committee on Systematic Bacteriology at the XIV International Congress of Microbiology (101).

Le Minor and Popoff of the World Health Organization (WHO) Collaborating Centre formally made the proposal to the Judicial Commission of the International Committee of Systematic Bacteriology in 1987. The word "enterica" was recommended because it had not been used before for a serovar nomenclature. It was also proposed that the seven subgenera of *Salmonella* be referred to as subspecies (subspecies I, II, IIIa, IIIb, IV, V, and VI). Subgenus III was divided into IIIa and IIIb by DNA similarity and phenotypic characteristics. The suggestion was accepted by the Centres for Diseases Control and Prevention (CDC) and other experts and laboratories but there were concerns about new nomenclature and serovar Typhi been overlooked (96). On this same matter, in 1999 Euzéby made an amended request to use "*Salmonella enterica*" as the type species of *Salmonella* and reserve the name "*Salmonella* Typhi" to reflect its clinical importance.

Only until 2002, the real discussion took place, with a final 2005 approval: "*Salmonella* enterica" would replace "*Salmonella choleraesuis*" to become the type species of the genus *Salmonella* (94). Still, the names of some medically relevant serovars, such as Typhi, Typhimurium and Enteritidis were kept due to their frequent use. In summary and based in later publications, the genus *Salmonella* consists of three species namely: *Salmonella* Enterica, *Salmonella* Bongori and *Salmonella* Subterranean (95). A general overview of the current established classification for the genus *Salmonella* is visualised in Figure 2.1.

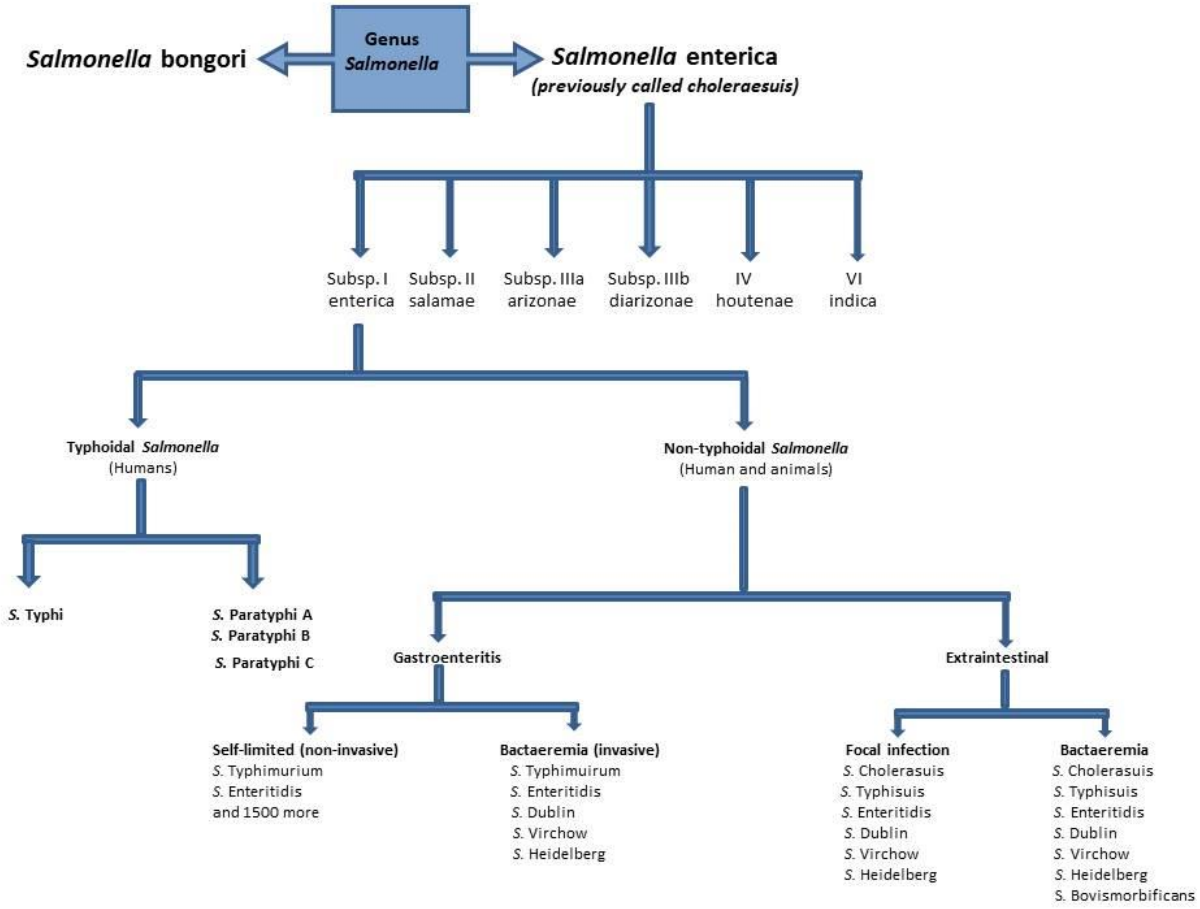


Figure 2.1. Current classification of Genus *Salmonella*

The Kauffman–White Scheme, as first published in 1929 and previously described in this Chapter, currently divides *Salmonella* based on their serological reactions to somatic lipopolysaccharide (O), flagellar (H) and capsular (Vi) antigens (79). These *Salmonella* serovars can also be subdivided into the three following host range types that also influence clinical manifestations/presentations (98):

(i) host generalist (ubiquitous serovars) include *Salmonella* enterica serovar Typhimurium (S. Typhimurium) and S. Enteritidis that cause infections in diverse human and animal hosts; they produce a range of clinical symptoms but mainly acute and self-limiting gastroenteritis (102).

(ii) host-adapted *Salmonella* (confined to a small number of hosts) such as *S.* Dublin in bovine animals and *S.* Choleraesuis in swine results in systemic infections, but these serovars can also infect humans and a limited number of other species (103).

(iii) host-restricted *Salmonella* are associated with severe systemic infections in a single host: *S.* Typhi and *S.* Paratyphi cause enteric fever exclusively in man while *S.* Typhisuis causes paratyphoid in swine, *S.* Gallinarum causes typhoid in fowl and *S.* Abortusovis causes abortions in sheep (104).

Of clinical relevance in human infection is *Salmonella* enterica, which is divided into six subspecies: S. enterica subspecies enterica (subsp. I), S. enterica subspecies salamae (subsp.II), S. enterica subspecies arizonae (subsp. IIIa), S. enterica subspecies arizonae (subsp IIIb), S. enterica subspecies houtenae (subsp. IV), and S. enterica subspecies indica (subsp. VI). Table 2.1 represents a summary of current nomenclature and taxonomic position of *Salmonella* (110).

Table 2.1. Taxonomic classification (writing format) and nomenclature for *Salmonella* spp.

Genus (capitalize, italic)	**Species (italic)**	**Subspecie (italic)**	**Serovar (or serotype)**	**No. of serovars is each species or subspecies**
Salmonella	*Enterica*	e*nterica* (subspecies I)	Cholerasuis, Enteritidis, Paratyphi, Typhi, Typhimurium	1504
		salamae (subspecies II)	9,46:z:z39	502
		arizonae (subspecies IIIa)	43:z29:-	95
		diarizonae (subspecies IIIb)	6,7:l,v:1,5,7	333
		houtenae (subspecies IV)	21:m,t:-	72
		indica (subspecies VI)	59:z36:-	13
	Bongori	subspecies V	13,22:z39:-	22
	Subterranean			

Within these subspecies, the various existing serovars used today to further classify each *Salmonella*, are based on the use of the antigenic classification system. The WHO Collaborating Centre for Reference and Research on *Salmonella* at the Pasteur Institute, Paris, France is responsible for updating the scheme. Every year newly recognized serovars are reported in the journal Research in Microbiology. In the latest report, there were a total of 2,541 serovars in the genus *Salmonella (*111). Serotypes may be said to be 'host adapted' if they are prevalent in one particular host, and also able to colonize and perhaps cause disease in other hosts. Serotypes adapted to man, such as *Salmonella* Typhi and *Salmonella* Paratyphi, usually generate severe septic typhoid syndrome (enteric fever) in humans and these serotypes are not usually pathogenic to animals (112-113). Ubiquitous serotypes, such as *S.* Enteritidis or *S.* Typhimurium, affect both

man and animals. *Salmonella* subspecies enterica include *S*. enterica serovar Typhi that causes typhoid fever, however, non-typhoidal salmonellosis (NTS) is caused by *Salmonella* species other than *S*. Typhi and *S*. Paratyphi (114). For a small number of serotypes, biochemical reactions or reactions against capsule proteins (Vi antigens) are also included. According to the established annotation, the antigenic formula is written with numbers and letters, with antigenic groups separated by colons. Each antigenically distinguishable *Salmonella* possesses a specific O (cell wall) and H (flagellar) antigen; many express alternate phase flagella of two antigenic types (H1 and H2) and a few produce Vi (capsular) antigen (47).

Each *Salmonella* serovar is therefore recognized by its unique combination of antigens (its antigenic formula), and each serotype was accorded species status. For example, the antigenic formula for the serotype Typhimurium is 1,4,5,12:i:1,2 because it carries an LPS protein reacting against the 1,4,5, and 12 antisera and the i phase I H antigen and the 1,2, phase II H antigen (71). Currently there are 67 O-antigens and 117 H-antigens that have been identified (47,95-96). The addition of the antigenic formula by national and international reference laboratories imparts more precise information about the isolates. Table 2.2 summarizes some of the most common *Salmonella* serovars and their antigenic formula.

Table 2.2 Some of the most common *Salmonella* serovars and their antigenic formula

O Group	Serovar	O somatic antigen	Vi antigen (if present)	H flagellar antigen (phase 1)	H flagellar antigen (phase 2, if present)	Antigenic formula*
D	*Salmonella* Typhi	9,12	Vi	D	-	9, 12 (Vi): d: -
A	*Salmonella* ParaTyphi A	1,2,12	-	A	-	1, 2, 12: a: –
C_1	*Salmonella* Cholerasuis	6,7	-	c	1,5	6, 7:c: 1,5
B	*Salmonella* Typhimurium	1,4,5,12	-	I	1,2	1, 4, 5, 12: i: 1, 2
D	*Salmonella* Enteritidis	1,9,12	-	g,m	-	1, 9, 12: g, m: -

*O antigens: boldface numerals

(Vi): Vi antigen, if present

H antigen (Phase 1): lower case letter

H antigen (Phase 2, if present): numeral

2.5 Challenges of laboratory identification of *Salmonella*

When requested by a clinician, the specimen is plated on selective media such as *Salmonella-Shigella* (SS) agar, Hektoen enteric agar, xyloselysine desoxycholate (XLD) agar or desoxycholate-citrate agar, which favour growth of *Salmonella*e and shigellae over other *Enterobacteriaceae*. Chromogenic agars specifically for *Salmonella* recovery are also available (97). The specimens are also often inoculated into the selenite F or tetrathionate broth, both of which inhibit replication of normal intestinal bacteria and permit multiplication of *Salmonella*e. After incubation for 1 – 2 days, the cultured isolate plated on differential and selective media for proper identification and classification. In order to ensure an adequate and accurate identification of each organism, a maximal recovery of *Salmonella* from faecal specimens must be obtained by using an enrichment broth. However, sensitivity of stool culture for recovery of *Salmonella* spp. is estimated to be only 70 % (114).

For the past 25 years, there has been an urgent need to develop faster methods to detect, identify and subtype *Salmonella*, specifically in clinical, food and environmental samples. For clinicians and clinical laboratories, the ideal test should be fast and accurate. Public health associated practitioners on the other hand, value tests that are fast and accurate, generally emphasize on accuracy more than speed, often needing more detailed information regarding isolates than clinicians do in order to epidemiologically link cases (4). For decades, the mainstream of diagnostics for enteric pathogens such as *Salmonella* has been through the use of culture. Most surveillance systems define a case as a culture-confirmed infection. Often isolates are submitted from clinical diagnostic laboratories to public health laboratories where further characterization is performed (115).

Despite traditional serotyping widespread use, it does have a number of drawbacks. Serotyping of *Salmonella* takes around 3 days to complete, is labour intensive, requires the maintenance of over 250 typing sera and 350 different antigens, and is unable to type rough or mucoid strains. Rough strains are incapable of expressing the O antigen, whereas in mucoid strains, it is common to observe the presence of a capsule that prevents the immunologic detection of the O antigen. These strains are known as variants of serotypes that don't express all the general antigenic characteristics, and could be classified as variants among serovars, due to the absence of the recognized antigens which are characteristic for particular serovars.

Furthermore, traditional serotyping is often not sensitive enough to provide the level of discrimination needed for food-borne illness outbreak investigations, and it cannot be used to infer phylogenetic relationships (116). Although traditional culture based methods detect "viable" bacterial cells, they do not detect potentially infectious non-culturable cells, and these techniques are time consuming, labour intensive, and not specific enough to detect and characterize *Salmonella* at the strain level, particularly when a large number of samples are involved (106).

Current testing of food and environmental samples for the presence of *Salmonella* can be divided into three stages: (i) detection of the pathogen; (ii) identification of the isolate as *Salmonella* and its specific serovar designation; and (iii) subtyping of the isolate for association with any clinical cases of salmonellosis (117-118). Confirmation relies on traditional biochemical testing of sugar and nutrient utilization media, which can take days to complete. Even with newer automated technologies that permit simultaneous testing of multiple samples, at least 24 hours are needed for a confirmation of *Salmonella* (117). The largest advance towards faster detection of *Salmonella* has been associated with molecular biology, where PCR and real-time, quantitative PCR (qPCR) are predominantly being applied as the methods of choice for the detection stage of

this process. Many different protocols targeting different genes or gene regions specific to *Salmonella* have been published (118-119).

Most laboratories have adopted culture-independent methods; this decision has been based mostly on the speed factor and also, in some cases, considering that these methods can provide more types of information than were previously available (120-121). However, specimens collected for culture independent testing may, in some cases, be incompatible with culture. In addition, performance characteristics of culture independent tests are variable and different from those of culture (122-123). Culture-independent tests for bacterial enteric pathogens include nucleic acid amplification tests such as PCR and antigen-based methods such as enzyme immunoassays and lateral flow assays. Culture often requires multiple days, whereas some culture-independent methods yield results in approximately one day. Although culture-independent test materials may be more expensive, they often yield cost savings by reducing the need for highly trained microbiologists. It also has the potential for ease of use and lower cost, therefore its adoption may increase the number of tests performed and, hence, the number of cases ascertained. Some tests might also detect pathogens that are not detectable using culture (120). Additionally, many targets have been investigated for the specific and sensitive detection of all *Salmonella*e in food and environmental samples. Though speed is the main advantage for PCR/qPCR assays, there are disadvantages to consider (Table 2.3). Advanced molecular and immunological methods require only a few hours on average to detect the target pathogen from food samples compared to 3 - 4 days using conventional culture-based methods (123).

Table 2.3 Comparison of traditional methods to molecular methods for *Salmonella* detection, identification and subtyping. (From: Bell *et al*, 2016. Recent and emerging innovations in *Salmonella* detection: a food and environmental perspective. Microbial Biotechnology (2016) 9(3), 279–292).

			Technical competency		
Method	**Resolving power**	**Accuracy**	**Performance**	**Analysis**	**Time to resolve**
PFGE	Sub-serotype	100%[d]	Highly trained/must be certified	Highly trained/must be certified	1 – 3 days
Traditional serology	Serotype	Approx. 80%[e]	Highly trained/must be certified	Highly trained/must be certified	Up to 3 days
Phage typing[a]	Sub-serotype	Approx. 80%[f]	Highly trained/must be certified	Highly trained/must be certified	1 – 2 days
PCR/qPCR	Genus to serotype[b]	Varies with protocol and matrix[g]	Moderately trained	Moderately trained	4 – 6 hours
MALDI-TOF	Species	>98% at species level[h]	Easy for clinical workflow	Easy for clinical workflow	< 5 minutes
LC-MS	Serotype to sub-serotype level	98% at serotype level	Moderately trained	Highly trained	< 1 day
WGS	Strain	100%	Easy to perform	Highly trained	3 – 4 days[j]
Metagenomics	Genus to strain[c]	Approx. 11%[i]	Easy to perform	Highly trained	3 – 4 days[j]

[a] Only used for *S.* Typhi, ParaTyphi A, Typhimurium and Enteritidis

[b] Depends on primers and matrix set used; some only detect genus, some serotypes

[c] Depends on sequencing depth, analysis pipeline and available database to query against

[d] If performed with Pulse Net certification standards

[e] Due to rough, mucoid and non-motile strains

[f] Due to ambiguous lysis reaction

[g] Must have a minimum of 10^2 genomes in the reaction in order to get a positive detection

[h] Based on single lab evaluation studies

[I] Based on current pipelines and databases in naturally contaminated cilantro (Jarvis *et al*, 2015)

[j] Depending on analysis time

2.6. Clinical and public health relevance of *Salmonella* serovars

Worldwide *Salmonella* infections, excluding those caused by *S.* Typhi and *S.* Paratyphi, were estimated in 2010 to cause 93.8 cases of gastroenteritis per year million (90% credible interval 61.8–131.6 million), 80.3 million of which are considered foodborne (9,12). According to FERG's 2015 report, NTS infections, when translated into global burden and its numerical location regarding association with foodborne illness, are located in the area associated with relatively low individual impact (DALYs per foodborne case) but having a high impact on a global spectrum (9). In the USA, over 1,000 different *Salmonella* serotypes were reported from 2002 to 2006, however the 100 most common serotypes accounted for about 98 % of the isolates (12). NTS have never been only present on developing countries. In fact, *Salmonella* causes approximately 1.4 million human infections each year in the United States, resulting in 116,000 hospitalizations and 600 deaths (2-3,15). Additionally, to commonly being associated with gastrointestinal illness, *Salmonella* can cause extra-intestinal infections. *S.* Typhi, *S.* Paratyphi, *S.* Choleraesuis and *S.* Dublin are the major serotypes, which cause invasive salmonellosis in humans (15). These invasive *Salmonella* infections may be life threatening. However, secondary bacteraemia and meningitis have been described appearing in less than 10% of cases. Antibiotic therapy would be needed in these severe situations (124). *S.* Typhimurium, *S.* Enteritidis and *S.* Heidelberg, are associated with a relatively low proportion of invasive infections; however, the total number of invasive cases caused by these serotypes appears to be high because they are relatively prevalent among the whole *Salmonella* population (125).

Despite the improvement in sanitation and hygiene, NTS illness continues to impose a significant burden on the population's health in industrialized and underdeveloped countries. It is estimated that 93.8 million cases of gastroenteritis due *Salmonella* spp. occur worldwide leading

to 155,000 deaths each year (9,12). The referred as “extreme age patients”, as well as those with immune suppression or accompanying severe infections such as meningitis, septic arthritis and osteomyelitis, would be at higher risk of developing extra intestinal salmonellosis complications (1). FERG latest report describes the ubiquitous presence of NTS through different regions across the globe. These regions include all different continents, involving cases from countries originally from developing as well as developed regions. It is interesting how the proportion of illness associated with NTS is higher among the regions of Europe, America and the Western Pacific Regions (9,12).

Although children <5 years of age represent only 9% of the global population, 43% of the disease burden from contaminated food occur in this group. Foodborne illnesses from diarrheal and invasive NTS resulted in the largest disease burden, reflecting the ubiquitous nature of *Salmonella*, the severe nature of illness, and the fact that young children are commonly infected (1,2). In these patients, the duration of diarrhoea was estimated to be 4.9 days (minimum of 4.3 – maximum of 8.4 days). In patients over 5 years old, the duration of NTS diarrhoea was 2.8 days (1). Mortality rates vary through regions, as it has been stated earlier on this revision.

WHO classifies countries/regions as medium to high-risk mortality countries and also low mortality countries, which could also reflect the availability and access to adequate health care. Low mortality countries and/or regions include Europe (i.e France, Netherlands), Australia, New Zealand and the United States of America. Medium to high-risk mortality countries/regions include Africa, Asia and Latin America, with no particular gender distribution in any epidemiologically related aspect, though there are differences in terms of dietary habits in some populations according to gender (9,11,12). Case fatality rates have been associated with several and very different risk factors. These are distributed in different levels according to the country

and the population involved in the analysis and include those strictly related to the patient's immune status, the presence of other gastrointestinal pathologies such as diarrhoea and malnutrition as well as cultural habits in terms of food consumption.

Additionally, different *Salmonella* serovars may differ in their invasiveness and risks of mortality. Also, the reports of unsafe food handling, mostly in terms of food preparation and lack of cross-contamination prevention, are some of the other confounding factors which have helped maintaining the presence of *Salmonella* serovars through different regions across the world. Serovars Enteritidis and Typhimurium are the most common in the UK (http://www.hpa.org.uk). Their distribution and presence globally is associated to numerous factors (11). In Africa according to public health statistics, *S.* Enteritidis and *S.* Typhimurium represent 26% and 25% of the isolates respectively. In Asia, Europe and Latin America/Caribbean, *S.* Enteritidis is the most frequent isolate (38%, 87% and 31%, respectively).

In North America, *S.* Typhimurium is the most frequented reported (29%) followed by *S.* Enteritidis (21%) and other *Salmonella* spp. (21%) (6,17). This statistics contrast with information belonging to other regions such as Asia, where NTS invasive disease is quite rare, with the exception of patients suffering from severe immunosuppression (19,22). In Australia, the most common serovar is *S.* Typhimurium, which is also the most commonly identified etiological agent associated with NTS outbreaks (27,32-33-36). Other serovars which are highly incidental in other reasons, such as *S.* Enteritidis, are not endemic in Australian poultry layer flocks and most human infections with *S.* Enteritidis are acquired overseas (25-27,30). Many other serovars occupy particular ecological niches and epidemiological foci in Australia, mostly due to the fact that it is a large country with wide climatic and geo-physical variation across all its regions/states (7).

In general, and according to annual reports, Salmonellosis is highly seasonal in Australia with October being the start of the high season for its presentation. It is also tracking towards a record number of notifications in 2016, with 3,011 notifications over the last quarter which is 1.3 times the quarterly rolling five year mean of 2,352.4. Since 2000, the Australian Government established OzFoodNet, a foodborne disease surveillance system, to improve national surveillance and conduct applied research into the causes of foodborne illness. This entity analyses national-level information on the incidence of diseases caused by pathogens commonly transmitted by food, *Salmonella* included, as well as investigating foodborne disease outbreaks across the country (35-37,126-128). Organizations such as the Public Health Laboratory Network (PHLN), Food Standards Australia New Zealand (FSANZ), the Department of Agriculture, Fisheries and Forestry as well as the National Centre for Epidemiology and Population Health at the Australian National University, are part of the OzFoodNet network. Its role, collaborating on the presentation of surveillance data, allows the characterization and comparison of the foodborne disease problem in Australia (1,9,28).

STM is the most common *Salmonella* serotype in New South Wales (NSW). According to OZFoodNet yearly reports, it regularly accounts for over 50% of all NSW *Salmonella* notifications and also a large number of Australian foodborne outbreaks. The NSW *Salmonella* surveillance system involves a complex network of local and interstate laboratories and notification processes. Local primary laboratories identify *Salmonella* species (spp.), while serovar confirmation as well as subtyping occurs at the closest reference laboratory. STM-induced salmonellosis follows a typical seasonal pattern with an increase in cases during warmer months. There is a decrease on notifications during winter months, which is also applicable to STM (37).

According to the NSW Public Health Bulletin, the majority of cases of human salmonellosis are caused by a small number of serovars (40-41). The 10 most common serovars identified by the NSW Enteric Reference Laboratory between 2010 and 2015 are listed in Table 2.4. These 10 most common serovars across NSW account for a range of 64.2 – 79% of all identified isolates.

Table 2.4 Ten most common *Salmonella* serovars isolated from humans and identified in the NSW between 2010 and 2015

Ranking	**2010 (n = 3,744)**	**2011 (n = 3,473)**	**2012 (n= 2,945)**	**2013 (n= 3,420)**	**2014 (n= 4,300)**	**2015 (n= 4,052)**
1	STM untyped (19.4%)	STM (56.8%)	STM (53.3%)	STM (52.8%)	STM (59.2%)	STM (42%)
2	STM – PT 170 (14.6%)	*S.* Enteritidis (5%)	*S.*Enteritidis (5.2%)	*S.*Enteritidis (4.4%)	*S.*Enteritidis (3.6%)	*S.*Enteritidis (4%)
3	*S.* enterica – unknown serovar (8%)	*S.* Virchow (4.6%)	Monophasic (ser. 4,5,12:i-)# (3.5%)	*S.* Virchow (3.3%)	*S.* Virchow (2.8%)	*S.* Saint Paul (3.3%)
4	STM (untypable) (6.4%)	*S.* Wangata (2.6%)	*S.* Virchow (2.9%)	*S.* ParaTyphi B bv Java (2.3%)	*S.* ParaTyphi B bv Java (2.3%)	*S.* Java (3%)
5	STM – PT 9 (5%)	*S.* Infantis (2.1%)	*S.* ParaTyphi B bv Java (2.8%)	*S.* Infantis (2.2%)	*S.* Wangata (2%)	*S.* Virchow (2.5%)
6	*S.* Infantis (3.9%)	*S.* ParaTyphi B bv Java (2.1%)	*S.* Birkenhead (2%)	*S.* Wangata (2.2%)	*S.* Infantis (1.9%)	*S.* Birkenhead (2.3%)
7	*S.* Birkenhead (2.8%)	*S.* Birkenhead (2%)	*S.* Singapore (1.9%)	*S.* Birkenhead (2.1%)	*S.* Birkenhead (1.8%)	*S.* Wangata (2%)
8	STM – PT 135 (2.1%)	*S.* Saint Paul (1.4%)	*S.* Wangata (1.8%)	*S.* Saint Paul (1.4%)	*S.* Saint Paul (1.1%)	*S.* Infantis (2%)
9	*S.* Singapore (1.5%)	*S.* Bovimorbificans (1.3%)	*S.* Infantis (1.4%)	*S.* Bovimorbificans (1.2%)	*S.* Stanley (0.9%)	*S.* Bovimorbificans (1.8%)
10	*S.* Enteritidis-untyped (1.4%)	*S.* Newport (1.1%)	*S.* Saint Paul (1.2%)	*S.* Stanley (1.2%)	*S.* Chester (0.8%)	*S.* Stanley (1.3%)

(n = number of records for that year. The percentage of each *Salmonella* serovar for the year is given in brackets. This information was obtained from OzFoodNet-Enhancing Foodborne Disease Surveillance across Australia. NSW OzFoodNet Annual Report 2010, 2011, 2012, 2013, 2014 and 2015.)

Since it is a heterogeneous serovar, additional molecular subtyping methods are needed in order to identify and relate specific strains to specific outbreaks. Within the different subtyping techniques available, MLVA (Multilocus variable-number tandem-repeat) typing appeared as a rapid alternative to successfully differentiate strains. In fact, in 2006 and 2007, routine STM MLVA subtyping was implemented in Queensland and NSW respectively and though is about to be superseded by newest tests and methods, it is still the performed typing method for STM (26).

2.7 Current characterization of *Salmonella* for public health surveillance

Emerging technologies such as Next Generation Sequencing (NGS) offer big opportunities providing even more rapid and accurate detection and genetic characterization of various *Salmonella* (118). Costs will remain relatively higher than currently available rapid methods until further technological and commercialization development occurs and may still require skilled technical and computer support to conduct analysis of the massive data sets (129). Currently, WGS requires a pure culture for the generation of a complete genome sequence, making the pathogen detection directly from food or another samples a challenge (130).

Several approaches for molecular detection and characterisation of *Salmonella* have emerged, including PCR-based, from the current genomics era. All of these methods propose to generate more efficient automation and data management, with an eventual final goal of a fully automated laboratory diagnostics workflow, with outputs from one approach strengthening other approaches, being directly or indirectly related (131-132). In Australia, all laboratory-confirmed *Salmonella* infections are reported to state and territory health departments, and subsequently to the National Notifiable Diseases Surveillance System (NNDSS).

In Australia, the Public Health Laboratory Network (PHLN) has developed a standard case definition for the diagnosis of diseases, which are notifiable in Australia. In this particular case, the "confirmed laboratory case definition" for *Salmonella* distinct two distinct syndromes: enteric fevers (typhoid and paratyphoid fevers) and gastroenteritis/salmonellosis, where the other over 2,500 serovars are classified. It requires of laboratory confirmation testing to be performed before notifying the public health authorities (40). This confirmation includes preliminary serology at public or private clinical laboratories, followed by confirm/biochemical identification, complete serological identification, and any other specific subtyping method to narrow down the identification step. Particularly in New South Wales (NSW), primary laboratory diagnosis of salmonellosis is usually made through the culture of stools and protocols stay close to the general previously made description (29,40-42).

The most common subtyping methods currently in use are the pulse-field gel electrophoresis (PFGE) and multiple loci sequencing typing (MLST) (52,133). PFGE is based on the separation of chromosomal DNA fragments after restriction enzymatic digestion. MLST is based on sequencing analysis of particular house-keeping genes and it has higher discriminatory power, excellent data analysis capability and, in contrast to PFGE, it has higher reproducibility between laboratories (134-143). Public as well as private diagnostic laboratories submit *Salmonella* cultures for serotyping to the NSW Enteric Reference Laboratory (ERL) at the Centre for Infectious Diseases and Microbiology Laboratory Service (CIDMLS), NSW Health Pathology. All isolates are tested to confirm their identity. The traditional tube methods of biochemical testing are the gold standard for reference laboratories (40-41). Though serovar identification within *Salmonella* is not relevant for patient's management, it is an important piece of information for epidemiological and outbreak investigation. For those common and successful serovars further

subtyping including the use of molecular based techniques such as multilocus variable-number tandem-repeats analysis (MLVA) are needed and have been implemented, becoming the method of choice in several laboratories worldwide till date (37,40-41, 127).

In Australia, the Public Health Laboratory Network (PHLN) has developed a standard case definition for the diagnosis of diseases, which are notifiable in Australia. In this particular case, the "confirmed laboratory case definition" for *Salmonella* distinct two distinct syndromes: enteric fevers (typhoid and paratyphoid fevers) and gastroenteritis/salmonellosis, where the other over 2,500 serovars are classified. It requires of laboratory confirmation testing to be performed before notifying the public health authorities (41). Laboratory confirmation includes preliminary serology at public or private clinical laboratories, followed by confirm/biochemical identification, complete serological identification, and any other specific subtyping method to narrow down the identification step. Particularly in New South Wales (NSW), primary laboratory diagnosis of salmonellosis is usually made through the culture of stools and protocols stay close to the general previously made description (37,41).

The traditional tube methods of biochemical testing are the gold standard for reference laboratories (42,45). Though serovar identification within *Salmonella* is not relevant for patient's management, it is an important piece of information for epidemiological and outbreak investigation. For those common and successful serovars further subtyping including the use of molecular based techniques such as multilocus variable-number tandem-repeats analysis (MLVA) are needed and have been implemented, becoming the method of choice in several laboratories worldwide (43,44,144-146). Public Health England, the USA CDC, MDU in Victoria and NSW ERL all routinely perform *Salmonella* WGS. In NSW, this is a complementary analysis to the still up to date implemented MLVA typing (147).

2.8 Public health relevance of reservoirs of salmonellosis

The vast majority of *Salmonella*e are primarily pathogenic in animals that constitute the reservoir for human infection; these include poultry, pigs, rodents, cattle and domestic pets (148-150). The organism almost always enters via the oral route through contaminated food or drink. Its main niche is the gastrointestinal tract of humans and animals (151-152). Many animals including cattle, rodents and fowl are naturally infected with a variety of *Salmonella*e, having the bacteria in their tissues (meat), excretions or even eggs (150).

Salmonella can be transmitted through the entire food chain from animal feed, primary production, and all the way to households or food-service establishments and institutions. In humans, salmonellosis is generally contracted through the consumption of contaminated food of animal origin (mainly eggs, meat, poultry and milk), although other foods including green vegetables contaminated by manure, have been implicated in its transmission under "cross contamination" (153-155). *Salmonella* spp. have a variety of animal reservoirs and routes of transmission that can result in human infection, including domestic or wild animal contact, human to human contact, water and soil (156).

In Australia, approximately 72% of salmonellosis cases are estimated to be transmitted through contaminated food (157-158). Common foods associated with salmonellosis in outbreak investigations and source attribution studies include eggs, poultry meat, pork, beef, dairy products, nuts, and fresh produce (157,159). *Salmonella*e are difficult to eradicate from the environment. However, because the major reservoir for human infection is poultry and livestock, reducing the number of *Salmonella*e harboured in these animals would significantly reduce human exposure (160). For example, in Denmark, all animal feeds are treated to kill *Salmonella*e before

distribution, resulting in a marked reduction in salmonellosis. Other helpful measures include changing animal slaughtering practices to reduce cross-contamination of animal carcasses as well as vaccination of food production animals such as chickens (121). To control salmonellosis, education and communication must target not only food producers, but those who prepare and provide food products. Protecting processed foods from contamination, providing training in hygienic practices for all food-handling personnel in slaughterhouses, food processing plants, and restaurants, cooking and refrigerating foods adequately in food processing plants, restaurants and homes are some of the areas where action must be taken in order to achieve control over the risk of infection (83,84,121).

2.9 Pathogenic determinants of *Salmonella*

Most of the serovars with exception of *Salmonella* enterica serovars Gallinarum and Pullorum, present a peritrichous flagella (the 'H' antigen). This multiprotein complex grants the bacteria the necessary structures for swimming and swarming motility (159). *Salmonella* strains will typically express either of two sets of genes encoding the flagellar antigens, and two distinct H-antigen 'phases' are therefore said to exist, and both need to be expressed in order for the serotyping to be performed (89). The mean infective dose to produce clinical or subclinical infection in human is $10^5 – 10^8$ *Salmonella*e. Host factors such as gastric acidity, normal intestinal microbiota and local intestinal immunity certainly contribute to resistance to *Salmonella* infections (115-116). Once ingested, *Salmonella* spp. must survive the low pH of the stomach, adhere to the small intestine epithelial cells and overcome host defence mechanisms to enable infection. *Salmonella* spp. possesses a number of structural and physiological virulence factors, enabling them to cause acute and chronic disease in humans (160).

The virulence of *Salmonella* spp. varies with the length and structure of the O side chains of lipopolysaccharide molecules located on the bacterial cell surface. Resistance of *Salmonella* spp. to the lytic action of complement (part of the immune response) is directly related to the length of the O side chain (161). Other important virulence factors include the presence and type of fimbriae, which determines the ability of *Salmonella* spp. to attach to host epithelium cells (148), as well as the expression of genes responsible for invasion of host cells (162-163). As shown in Figure 2.2, *Salmonella* will successfully invade non-phagocytic cells by inducing their own uptake.

Virulence genes involved in invasion and those required for intracellular survival are clustered in large chromosomal DNA regions designated *Salmonella* pathogenicity islands (SPIs), which are defined as large gene cassettes within the *Salmonella* chromosome that encode determinants responsible for establishing specific interactions with the host, particularly associated with bacterial virulence (164-168). SPIs contribute to host cell invasion and intracellular pathogenesis. Currently, 12 SPI have been described, and though *Salmonella* subspecies and serovars may present differences in terms of SPIs size, structure, function and distribution, there are also several SPIs conserved throughout the genus (168). For example, SPI-1 is distributed among *Salmonella spp.* and required for bacterial invasion into the intestinal epithelial cells, while systemic infections and intracellular accumulation of *Salmonella* spp. are dependent on the function of SPI-2 (164-165).

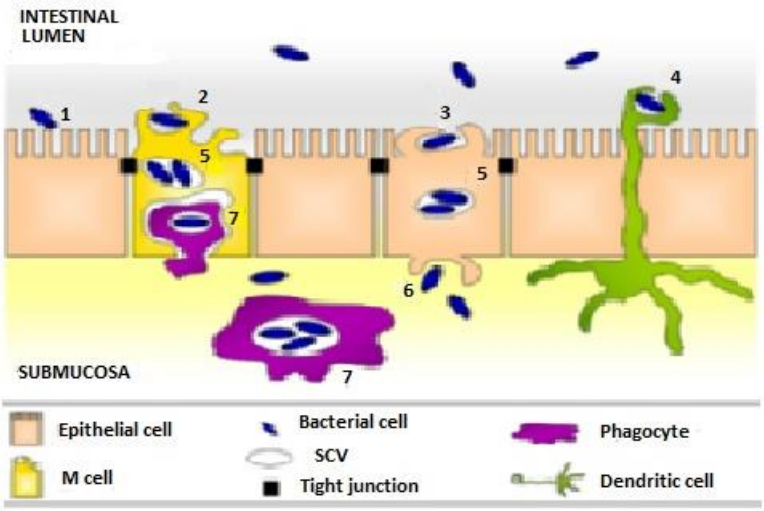


Figure 2.2 Pathogenesis model of *Salmonella enterica se*rovar Typhimurium. **1**: *Salmonella*e attach to the intestinal epithelium using adhesins encoded within SPI-3 and SPI-4. **2 and 3**: Invasion and engulfment of bacteria, mediated by virulence factors encoded within SPI-1 and SPI-5. **4:** alternatively, dendritic cells can directly take up bacteria from the sub mucosa layer. **5**: Once inside the cytoplasm, *Salmonella* is located within the *Salmonella containing vacuoles* SCVs where it replicates. **6:** The SCV releases the internal cells to the sub mucosa. **7:** Bacteria are internalized within phagocytes and located again within a SCV. Finally, these phagocytes can disseminate through the lymph and blood stream. Figure from *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. (From: Fabrega A and Vila J. *Salmonella enterica* serovar Typhimurium skills to succeed in the host: virulence and regulation. Clin. Microbiol. Rev. April 2013; 26(2):308-341).

SPI-1 and SPI-2 encode type III secretion systems, consisting of multiprotein complexes in charge of building channels across the bacterial and epithelial cell membranes. These structures result in the efficient translocation of bacterial effectors directly into the epithelial cell cytoplasm, which after interaction with the cell's own structures, are capable of altering the host cellular functions, resulting in intracellular bacterial survival and posterior colonization (168). Following invasion, *Salmonella* stays within a membrane compartment known as the *Salmonella*-containing vacuole (SCV; Figure 2.2), where the bacteria actively remodels this compartment and establishes itself, becoming capable of survival and replication (164-166).

In other aspects, multiple antibiotic resistant strains of *Salmonella* have also emerged; one example is *S.* Typhimurium definitive phage type 104 (DT104). Multi-resistant *S.* Typhimurium ST 104 infects both humans and animals, such as cattle and sheep (169). To date, this organism is not endemic in Australia, although it is a significant health problem in European countries, North America, the Middle East, South Africa and South-East Asia. Similar situation applies to ST 34 and its multiple antibiotic resistant profile (44). *Salmonella* spp. is one of the most commonly identified pathogens worldwide, able to cause a variety of infections in humans and animals (3).

Due to the over 2,500 different serovars within the species *S. enterica*, typing methods for discriminating bacterial isolates from the same species, which used to be essential epidemiological tools, became insufficient due to their limited ability to examine the relatedness of isolates. Methods exploring at the molecular level became revolutionary, adding new tools for enhanced surveillance and outbreak detection. Depending on the setting, one or more genotyping methods might need to be considered. The following chapter describes the implementation and use of particular genotyping technologies for STM.

Chapter 3: Added value of genotyping of STM in disease control

3.1 Introduction

In Chapter 2, the taxonomy of *Salmonella* was discussed in detail, showing the complex classification and diversity involved within these organisms. The focus of this Chapter is *Salmonella* genotyping and the importance of it for public health control of salmonellosis. Traditional serotyping remains widely used, and the identification of *Salmonella* serovars remains an important public health diagnostic need (147). However, this technique presents several obstacles discussed in detail in Chapter 2 (subchapters 2.4 and 2.5) that include the time frame involved to generate accurate identification results, labour, training and the maintenance of a large number of antisera to actually serotype each *Salmonella* isolate (79).

Hence, there is a proclivity to test, implement and even some day replace traditional typing techniques with genomic based methods completely, which could potentially save time and mostly labour involved in the characterisation of pathogens such as *Salmonella enterica* serovar Typhimurium (STM). Nevertheless, genomic methods also require bioinformatics skills in addition to the time needed to analyse this data (148).

3.2 Genome of *Salmonella*

The first analysis of a *Salmonella* genome was through the construction of linkage maps, where the genes were located along a map. Individual genes were points representing particularly defined mutations, without any additional information regarding size or even nucleotide sequence of the gene. *Salmonella enterica* serovar Typhimurium strain LT2 (STM-LT2), the principal strain for cellular and molecular biology, was isolated in the 1940s, going through several changes in

term of the descriptive approach. The circular map that was originally created had no physical length of the chromosome. This was later on redefined by predicting the length of the chromosome in kilobases (kbs) of DNA, based on the analysis of joint transduction of genes (149). The publication of the first completely sequenced *Salmonella* genome happened in 2001, particularly for *Salmonella* Typhimurium LT2 (150). They sequenced the 4,857-kb chromosome and 94-kb virulence plasmid of S. Typhimurium strain LT2 and compared it against 8 other enterobacteria (Figure 3.1). STM-LT2 genome presents a size of 4,857,432 base pairs (bp), with a G+C content of 53%. The general characteristics of the STM-LT2 genome are summarized in Table 3.1.

Table 3.1 STM LT2 genome main characteristics

Parameter	**Chromosome**	**Plasmid pSLT**
Size (bp)	4,857,432	93,939
G + C content	53%	53%
rRNA clusters	7	0
tRNAs	85	0
tRNA pseudo gene	1	0
Structural RNAs	11	1
CDS (including pseudo genes)	4,489	108
CDS pseudo genes	39	6

Genomic comparisons between sequenced enterobacterial genomes including other *Salmonella* serovars and other enteric pathogens such as *Escherichia coli* K12 and O157:H7 have been performed (150). These two species can be considered related phylogenetically in the distant past since they share this large amount of genetic material, which typically varies between 2,500 and 3,100 orthologous genes (between 50% and 70% of the whole genome. As expected, there is also large homology between *Salmonella* spp. (>98%), for DNA and also amino acids.

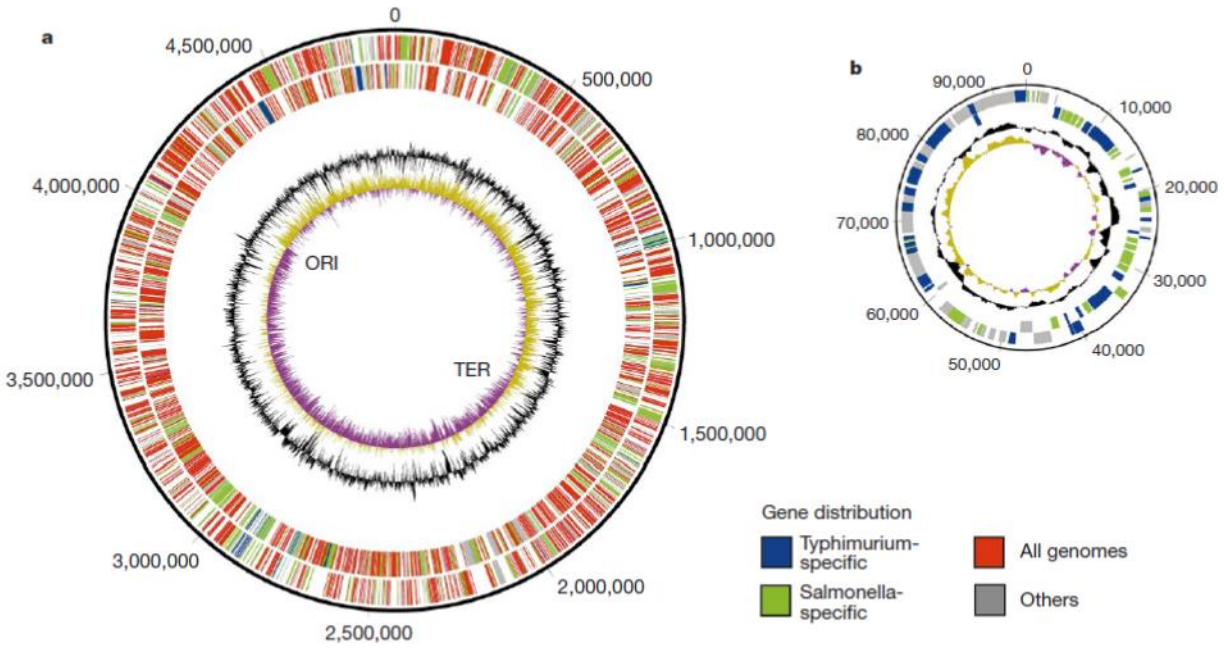


Figure 3.1 The *Salmonella enterica* serovar Typhimurium LT2 genome. **a:** STM LT2 chromosome compared to 8 related enterobacteria. Base pair numbering is indicated outside the outer circle. The outer two circles represent the coding orientation, with the forward strand on the outside and the reverse strand on the inside. Red indicates close homologues in all eight genomes. Green indicates genes with a close homologue in at least one other *Salmonella* (S. Typhi, S. paratyphi A, S. paratyphi B, S. arizonae or S. bongori) but not in E. coli K12, E. coli O157:H7 and K. pneumoniae. Blue indicates genes present only in STM LT2. Grey indicates other combinations. The black inner circle is the G+C content; the purple/yellow innermost circle is the GC bias. The positions of the origin of replication (ORI) and terminus (TER) are shown. **b**: The plasmid pSLT. Base pair numbering is indicated outside the outer circle. The plasmid is not to scale. The colour scheme is the same as for A. (*From McClelland et al, 2001. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. Letters to Nature, Vol 413, 25 October 2001*).

Callister *et al* (2008) described that, at the sequence level, the genes that belong to the core genome are highly similar, but it is uncertain whether they have maintained similar function in both species since they inhabit different ecological niches and are not functionally independent of the specie-specific genes. Kaushik et al (2017) reviewed the diversity of gene cassettes which gene expression is known to be optimized throughout evolution toward the changes in an organism's lifestyle and the niche that it occupies, which would explain the prevalence of these organisms worldwide and through time.

For other enterobacterias, homology is from 80% minimum for DNA and >88% for amino acid sequences. Distinct phenotypic characteristics have been observed amongst the 2,500 *Salmonella* known lineages, the genetic basis of which has been determined using genomic sequence analysis. These characteristics have been used as a way to identify distinct differences between organisms from similar species (151). As an example, a comparison of *Salmonella* serovar subspecies I genomes which infect both animals and humans, demonstrated that these pathogens share around 90% of their genes. This large portion within the genome has been named "core genome".

The core genome includes mainly genes that are essential for the bacterial cell in terms of growth and survival, and are often related to code for enzymes involved on the biosynthetic pathway. Core genes, are not only highly conserved in terms of sequence identity but also in terms of their position within the genome (152). The remaining 10% of genes has been described as unique to each of the different lineages and has been labelled as "accessory genome". This accessory genome, contains mostly genes of prophages and genes of unknown function which have contributed to the genetic diversity of *Salmonella spp.,* for example, the *Salmonella*

Typhimurium serovar (153). The accessory genes are those related to pathogenesis. Chromosomes of enteric bacteria consist in collinear regions intermixed with so-called islands, which could be unique to certain species. These islands, also known as *Salmonella* pathogenicity islands (SPIs), are main components of the accessory genome together with prophages, which sometimes encode pathogenic functions. The acquisition of SPIs can occur horizontally, enabling bacteria to rapidly gain complex virulence functions from other species (139). There are 23 different SPIs, which have been described so far within the accessory genome (154). Five SPIs (SPI-1 to SPI-5) are common to all serovars of *S.* enterica while the rest is distributed among different serovars and/or strains. Laing *et al* (2017) showed a concordant relationship within *S. enterica,* between the core and accessory genomes. Their results indicate that the accessory genome is a specific selection within particular niches, establishing a complement of genes and regulatory elements that enable as well as increase the survival of the *S. enterica* strains.

The definition of core and accessory genome is not absolute. It has been stated that genes could be defined to be part of the core genome taking into consideration the actual strains being compared. For example, when performing comparison between 2 strains, core genome would be defined as those genes that are common to the 2 members of the comparative pair (153). However, there will be a high proportion (approximately 90%) of genes shared between different members of the genus, and the majority of differences that exist between *Salmonella* serovars under analysis would be primarily due to differences localized in the accessory genome. This was described by comparing genomes from *Salmonella* paraTyphi C strain RKS4594 and *Salmonella* cholerasuis strain SC-B67. They shared 4,346 genes, equivalent to 96.66% of the S. paratyphi C strain RKS4594 and 98.2% of the cholerasuis strain SC-B67 (155).

In terms of the clinical response generated by *Salmonella* pathogens, they may not have as many genes in common. This was demonstrated when comparing the human typhoid agents S. Paratyphi A and S. Typhi. These only have 89.8% and 90.2% common genes at core genome level, which was equal to 4,008 genes only (156). These mentioned genomic differences might imply the influence that independent adaptation processes associated with the host are taking place. Host-pathogen interactions would be then responsible of the gene modifications taking place in the described comparison, despite the expected closeness of these two infectious agents.

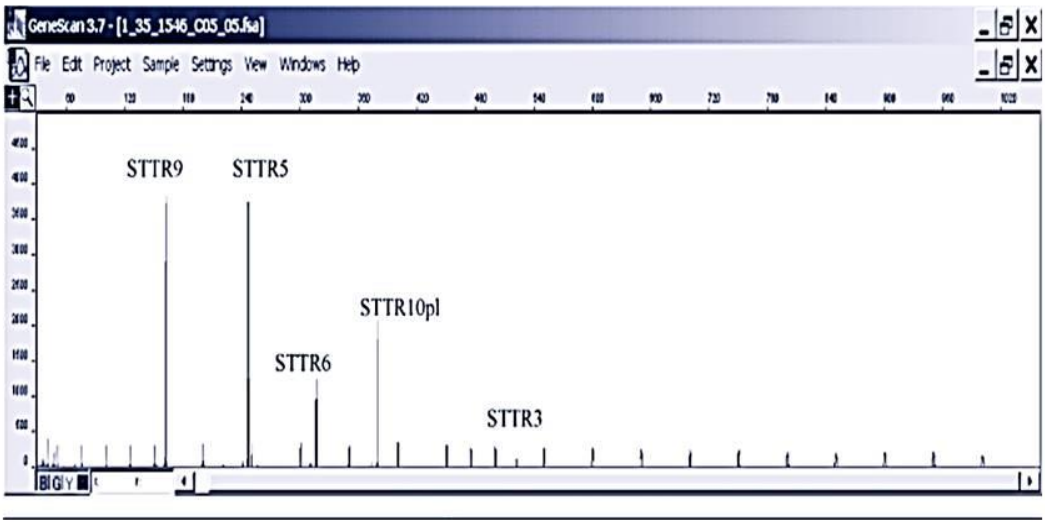
3.3 STM MLVA for laboratory surveillance

Pulse-field gel electrophoresis (PFGE) has been used worldwide for many years to subtype *Salmonella enterica* serovar Typhimurium within phage types (157). Unfortunately, phage typing often makes differentiation of epidemiologically closely-related isolates difficult, which can interfere with epidemiological investigations of outbreaks. Furthermore, PFGE can be relatively slow in terms of performance and results as well as subjective in terms of interpretation (158). Combined, these factors can lead to delays in identifying outbreaks and their sources in order to prevent further spread. The development of MLVA typing has offered a solution to this situation: a shorter turnaround time combined with high discriminatory power enables the characterization of subspecies during epidemics from a single serovar or a phage type (159).

Several MLVA schemes have been developed for *Salmonella*, including a general scheme for S. enterica subspecies enterica, one for *S.* Typhi and for *S.* Typhimurium (119). This particular typing method became widespread among many public health laboratories for the investigation of *Salmonella* and other foodborne organisms. Although there are several other methods still available, MLVA generates reproducible and comparable results, by using the same standardized

techniques, can be easily shared between different laboratories (160). The technique of MLVA is based on the polymerase chain reaction (PCR) and involves amplification of targeted DNA fragments using fluorescently-labelled (161). The generated products (amplified DNA segments or amplicons) contain variable copies of tandem repeats which are then measured in terms of length, either by gel or capillary electrophoresis. This method provides reproducible and accurate sizing of fragments (within 1 base + or -) and can be performed more rapidly than other molecular-based methods. It has several additional advantages over other molecular methods including reduced cost, minimum equipment required, speed, reliability as well as high reproducibility.

The chromatogram representation (Figure 3.2) provides the approximate lengths for each locus. The sizes of the genotyped fragments as determined by the analysis program are relative due to slight differences in DNA migration through the capillary matrix. These differences depend on the DNA sequence, which affects the secondary structure of the single strand, DNA as well as the type of fluorescent marker associated to the primer.



Locus	Flanking length*	Amplicon length	Repeat length	Repeat No./ code
STTR9	144	162	9	2/ 3
STTR5	175	247	6	12/ 13
STTR6	264	318	6	9/ 10
STTR10pl	311	377	6	11/ 12
STTR3	106	523	27/33	2+11/ 523

*All lengths are in base pairs.

Figure 3.2 An example of a S. Typhimurium MLVA chromatogram. Generated by ABI capillary genotyping of fluorescently-labelled multiplex PCR products. Each peak corresponds to a specific MLVA locus. Source: New South Wales Enteric Reference Laboratory at the Centre for Infectious Diseases and Microbiology, ICPMR, NSW Health Pathology, Westmead.

The raw genotyping data is then compared to a standard curve prepared by sequencing a number of different sized fragments from each locus and comparing the actual fragment length to the sizes provided by genotyping. This may also be confirmed by comparing the nucleotide sequence of the fragment with the genomic sequence of type strain STM- LT2 in the NCBI Genbank. The number of repeats for each locus is inferred by subtracting the known length of the flanking sequence from the total amplicon length, and dividing the result by the known length of each repeat sequence. In recent years MLVA has almost completely replaced PFGE as the method of choice for genetically separating closely related *Salmonella* strains. STM was one of the first serovars of *S.* enterica to be analysed using the MLVA system (161-162). This MLVA typing scheme was first proposed for the analysis of *S.* Typhimurium in 2004 (162), with later on development of the assay targeting five more variable loci. The five involved loci are – STTR9, STTR5, STTR6, STTR10pl ('pl' referring to the fact that this locus – STTR10pl – is on a plasmid while the other loci are on the chromosome) and STTR3. These five loci are summarized in Table 3.2.

The MLVA profile is usually expressed as a series of numbers of particular length, each of which represents the number of copies of repeated sequences at each one of the loci under analysis in a standard and particular order (163). This method does not make any assumptions on the nature of the sequence. For STM, loci are designated as STTR – (*Salmonella* Typhimurium tandem repeat), plus an arbitrary number.

Table 3.2 Designated loci, size and sequences for *S*. Typhimurium MLVA. (i) as described in http://www.pasteur.fr

Locus	Length (bp)	Tandem repeat sequence
STTR-9	9	Gtctgcgat
STTR-5	6	Accacg
STTR-6	6	Gcaagg
STTR-10	6	Cctgtt
STRR-3	27, 33*	Cgatgtgaccccgcccgatgatagcggcgatga

***** The length of repeat sequences at locus STTR-3 corresponds to a combination of 27 and 33 base pair repeats.

After all five loci of an isolate are assigned a particular fragment length, a profile (or string) is prepared. One way of expressing the profile is providing the actual fragment lengths in base pairs. Another subtyping method, proposed by Lindstedt *et al* (2004), which is the method used throughout Australia, is assigning numbers that reflect the number of tandem repeats contained within a particular fragment. Table 3.3 relates particular fragment lengths to the Australian MLVA coding system.

Table 3.3 Fragments relationship to the Australian MLVA coding system

Locus	**STTR-9**	**STTR-5**	**STTR-6**	**STTR-10**	**STTR-3**
Fragment length (bp)	162	313	324	371	523
Australian code	03	24	11	11	523

As shown in Table 3.3, the profile 3-24-11-11-523 indicates 3 repeat copies at locus 1, 24 repeat copies at locus 2, and so on. Where repeat copies differ in length, allele codes can be assigned to distinguish specific types of repeat combinations (132). Since repeat lengths in STTR3 can differ, this particular locus is sometimes expressed as the total size in base pairs (163). This

one in particular has been described as a complicated locus to analyse. The 5-locus STM multiplex scheme has been widely adopted in Europe, Australia and other countries (34,164-165). Some countries, such as Norway and Denmark, have been able to successfully harmonize the performed analysis (166). Whilst there are various possible formats in which the MLVA profile could be expressed, so far, none has been globally adopted (167).

In 2008, representatives from Australian reference laboratories agreed that for all loci except STTR3, the result will be expressed as 0 if there is no amplicon (i.e. the locus is absent); 1 if the size of the amplicon corresponds with that of the flanking region (i.e. the locus is present, but no repeat sequences are present); 2 if the amplicon length corresponds with the sum of the flanking region and one repeat, and so on. For the STTR3 locus, which is complicated by the potential presence of variable numbers of repeats of two different lengths, it was agreed that the actual amplicon length would be given. The agreement on displaying information is represented in Figure 3.3; this provided a method by which Australian laboratories can compare results (32). Genetic relatedness of isolates can be established by analysis of their MLVA profiles. In NSW, MLVA fragment lengths data from each STM screened isolate are entered in the NSW STM-MLVA Database together with any relevant epidemiological details such as date, source and location. Using this, data dendrograms can be generated providing a summary of the epidemiological scenario behind the appearance of particular MLVA profiles.

In general, loci STTR-9 and STTR-3 are considered stable loci as there is slight or no allelic variation between isolates that are closely related based on phage type and epidemiological data (119). Consequently, any differences observed at either of these two loci for any given isolates are, regardless of the data for the remaining three loci, considered sufficient to call the isolates *unrelated* (distinct), conclusion which should be supported providing necessary epidemiological

data (165). Allelic variation in loci STTR-5, STTR-6 and STTR-10, has been found to be frequent since any of these three loci may undergo one or more tandem repeat changes with relative ease. For these particular loci, confirmation of relationships between isolates may need supplementary information based on additional pieces of information.

An example of genetic relatedness based on MLVA profiles is shown in Table 3.4.

Table 3.4 Examples of genetic relatedness between STM isolates based on MLVA profile

	MLVA-5 locus: tandem repeat numbers					
Profile	**STRR-9**	**STRR-5**	**STRR-6**	**STRR-10**	**STRR-6**	**Genetically related to profile 1***
1	2	23	10	10	523	N/A
2	2	23	10	9	523	Probable
3	2	24	12	10	523	Possible
4	2	24	12	11	523	Distinct

* The presented categories for genetically related to profile 1 corresponds to a proposal by the Rural Industries Research and Development Corporation, Australian Government, 2013.

The MLVA method has shown a high discriminatory power between isolates belonging to the same serovar, whether obtained from human, food or animal sources. It is also a method that has proven to be suitable for identifying links between potential outbreak cases and sources (168). However, common or endemic MLVA profiles may cause multiple outbreaks along with sporadic cases and a few selected MLVA profiles may represent a large portion of isolates observed in a geographical location (34). In terms of MLVA profiles analysis, it typically relies on distance-

based methods to infer relationships between isolates and their different profiles based on dendograms or phylogenetic tree constructs, as proposed by PulseNet MLVA protocols (http://www.pulsenetinternational.org).

Dendograms are usually based on the number of differed loci, which is a rather basic approach that could implicate some problems including the linking of profiles which share loci instead of actually inferring relationships beyond the loci based profile (168). So, high genetic diversity outside the five typing loci within particular MLVA profiles may need additional analysis beyond MLVA typing in order to accurately identify links between cases and to resolve outbreaks of salmonellosis. In such instances, Whole Genome Sequencing (WGS) has presented itself as the ultimate discriminatory power tool, with the potential to improve epidemiological investigations and determine transmission pathways (Section 3.4) (169).

As concluding remarks for this subchapter, MLVA typing has proven to have enough discriminatory power to, in general, effectively differentiate STM isolates. Its discriminatory power is even superior to other typing methods. It is also technically simple, inexpensive to perform and has been used as a helpful method for epidemiological investigation. Hence, common MLVA profiles may represent a large portion of isolates observed in a geographical location, without possibility of accurately differentiating them. WGS has offered highest discriminatory power when compared with MLVA typing, improving epidemiological investigations and transmission pathways analysis.

3.4 WGS for pathogen detection and analysis

Advances in genomics are transforming public health laboratory surveillance (148). The term "genomics" came into sight around 1986 by Dr. Thomas Roderick, a US geneticist, who was mostly focussed on the study and comparison of genomes of diverse species, including concepts of evolution and relationships (170). It was actually the Human Genome Project the instance that encouraged the actual revolution in terms of sequencing technologies. Through its initiatives, high-throughput WGS was established as an important tool involved in the study of human and microbial organisms at a deeper, modified and improved level based on the methods previously published (171). The first aim was to search for more efficient and richer in information sequencing methods, including sequencing of longer segments of DNA and/ or whole chromosomes. Advances in software and computer science helped producing large amounts of sequence data across the genome that was then assembled using newer computation technology (172).

As WGS continually becomes quicker and more cost effective, it is likely to be used as a routine epidemiological typing tool due to its ability to discriminate between sporadic and outbreak related cases, which may be indistinguishable using traditional molecular subtyping methods including MLVA (173). The development of high-throughput next generation sequencing technology has focused on the establishment of increasingly simple bench top technology (148). These have improved the capacity to perform efficient and lower-cost WGS, enabling its implementation as a tool to be used in clinical diagnostics and public health microbiology almost in real-time (174). Generally, WGS technology follows a mostly standardized workflow, beginning with extraction of homogenous microbial DNA, which in the case of STM, is usually from a single colony selected from a pure culture. WGS requires high-quality, intact, non-degraded

DNA (175). The extracted DNA is then subjected to library preparation (176), which involves preparation of a double-stranded DNA library consisting of cleaved DNA fragments ligated to ‘adapters’ which are known sequence fragments, specific to the particular sequencing platform.

Preparation of a DNA library is a crucial step, required to generate material suitable for sequencing (177), and there are several library preparation methods specific for the various sequencing platforms. However, they all have in common that the DNA fragments must always be linked with particular adapters. These adapters will enable multiple samples to be pooled in a single run and subsequently de-identified and assembled for analysis. Prior to adapter attachment, normalization and clean-up of the samples must be performed; that way a standardized amount of each sample is loaded into the sequencer. Once the DNA fragments have been loaded into the sequencer and sequencing data have been generated, the output is referred to as reads, which are short or long base pair sequences which have been inferred from the DNA template by the sequencing process. Figure 3.3 summarizes the steps involved in one of the library preparation procedures available.

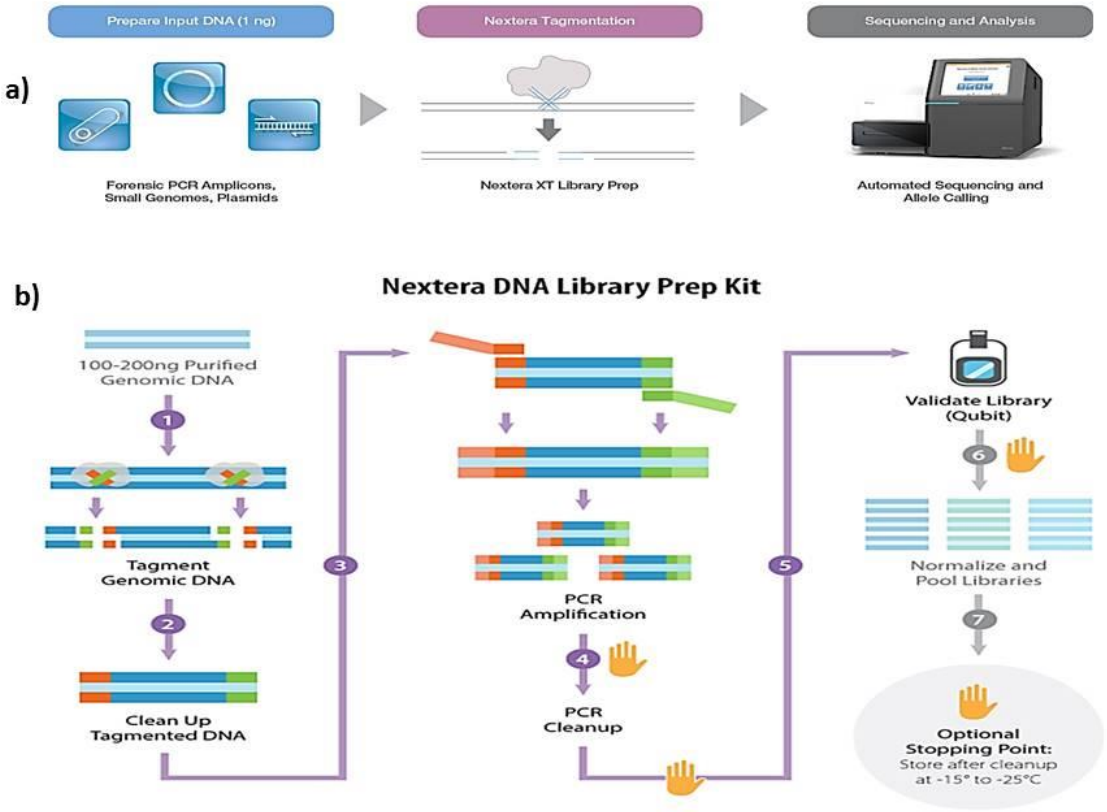


Figure 3.3 General library workflow for WGS. Simple overview from input DNA, DNA tagmentation and sequencer (Source: Illumina Nextera DNA reference guide, 2016).

Read length has bioinformatics implications in terms of analysis; authors have referred to short sequencing reads at lengths below 500 bp and long reads above this length (178). Once the output is revealed from sequencer, the options are either read mapping to a reference genome or reads can also be assembled by using *de novo* assembly into longer contiguous sequences.

Mapping is a term used to describe the alignment of short sequence reads to a longer specific reference sequence; d*e novo* assembly refers to the reconstruction of contiguous sequences without making use of any reference sequence (179). There are several quality metrics such as coverage (or sequencing depth), Q30 scores and cluster density which enable the determination of data quality post sequencing; among these, Q scores are defined as a property that is logarithmically related to error among the base calling process performed by the sequencer. A Phred quality score (Q30) is equivalent to the probability of an incorrect base call in 1,000 times, providing a base call accuracy of 99.9%. Q30 is considered a benchmark for quality in WGS (180). Sequence coverage refers to the average number of reads per locus that the sequencer has been able to analyse in order to achieve confident base calling. It is expressed as an average or median of all the coverage values per base. Sequencing at high levels of coverage provides the generation of high quality and accurate sequences, reliable in terms of accurate detection of variants within particular genomes. Human WGS studies are usually performed at least at 40x coverage (181). Cluster density is another term to consider when analysing sequencing results. It is an important metric that influences the total data output.

"Clonal clusters" are the amplification product from the library preparation and should be around 170-200 raw density (K/mm^2). The density of these clusters impact in terms of data quality and total data output. Under clustering may result in high data quality but lower data output whereas over clustering leads to lower data quality in general (182).

Different factors may contribute to under and over clustering; insufficient library clean-up, inaccurate library quantification and nucleotide diversity within the library are the main general ones. There are other several elements, particularly involved on the quality of the sequencing data, which must be taken into consideration: overall GC content, duplicate reads and necessary amount of reads to be able to interpret results are among them. Different tools provide summarized statistics in relation with quality control (176). Trimming low quality data and reads can be performed with the help of several software and scripts. Failure to remove sequences that are not within expectations can disrupt the final conclusions based upon the sequencer resulting output information (178). Figure 3.4 summarizes the steps involved in data analysis of WGS sequencing.

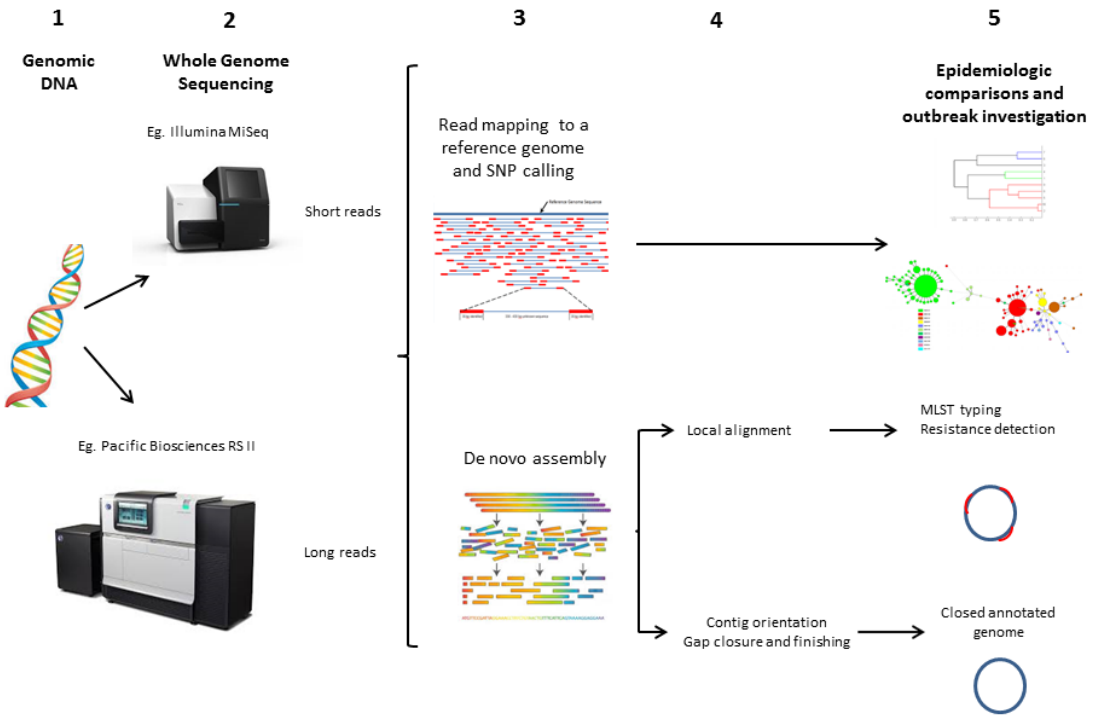


Figure 3.4 General overview of WGS data analysis workflow**. 1)** Sequencing starts with good quality DNA extract. **2)** This is loaded into any of the available next-generation sequencers, which will be able to produce either short genome reads or long genome reads . **3 & 4)** Analysis may be performed by either mapping reads against a specific reference genome or by assembling these reads into longer contiguous sequences (*de novo* assembly). **5)** If mapping to a reference genome is used, this can lead to epidemiological comparisons and outbreak investigation. The resulting *de novo* assemblies can be used for further analysis including antibiotic resistance detection. (Figure from Kwong J, Mccallum N, Sintchenko V, Howden B. Whole genome sequencing in clinical and public health microbiology. Pathology. 2015; 47(3): 199–210.

Worldwide, there are numerous sequencing platforms available and hence several considerations must be taken into account when determining which platform should be used for a particular clinical microbiology application (148). Currently the market offers several next-generation sequencing methods, all of which have advantages and disadvantages (178). Table 3.5 summarizes some of the most popular sequencing technologies and their main characteristics.

Table 3.5 Summary of the currently most used sequencing platforms

Traditional sequencing	
Sanger sequencing	Still widely used of short DNA segments sequencing.
Shotgun sequencing	Involved fragmentation of long strands of DNA into numerous smaller segments for Sanger sequencing, therefore higher price.
Next-generation sequencing technologies (NGS)	
Pyrosequencing (Roche 454)	One of the earlier NGS technologies. Currently not much in use.
SOLiD sequencing (Life Technologies)	Less popular than Ion Torrent, likely to be over shadowed by newer technologies.
Ion semiconductor sequencing (Life Technologies Ion Torrent)	Popular due to lower cost and speed of sequencing for short read sequences. Higher error rates and poor coverage for high A-T or G-C regions.
Illumina sequencing	Current market leader. High sequencing throughput; low error rate and low cost. Limitation of short read sequences.
Single molecule real-time sequencing (Pacific Biosciences)	Novel method. High raw errors. Overlapping reads can produce accurate consensus sequences.
Emerging technologies	
Nanopore sequencing (Oxford Nanopore)	Benchtop sequencing technologies leader. Portability. Capable of generating long-sequence reads. Still being refined.

Following, is a list of the main aspects that should be taking into consideration when considering the use of sequencing technology:

i) Cost: 10 years ago the sequencing costs of a million base pairs sequenced were approximately US$1000; the costs are now below US$0.10 (183). Despite the decrease in price, there remain large expenses involved on the implementation of this kind of equipment, including set up, sequencing reagents cost and also bioinformatics processing costs.

ii) In-house sequencing or service provider purchasing: The main consideration on this particular mater is the investment associated with the purchase of sequencing technology. If this can be handled, turnaround times for analysis and data generation can improve dramatically. External service providers may involve longer times related to analysis and data.

iii) Sequencing contents: There are differences regarding the available sequencing technologies. Some of them can sequence less than 10 bacterial genomes within hours; others can sequence up to 100 in a single run, which may take around 3 days. This will certainly impact on the costs implicated on the sequencing run.

iv) Versatility: Considering that sequencing technology is explosively evolving, the platform's capability to upgrade and modify its sequencing practices is a relevant factor.

v) Data quality: This factor, in practical terms, might be one of the most relevant ones. Without good quality of sequencing results, there would be not much information to report. Each sequencing platform is equipped with a tool that will be able to provide information about the quality of the sequencing reads.

There are some limitations that should be taken into account when evaluating NGS technologies. Primarily, most of the analyses are carried out through the use of single nucleotide variants (SNPs), which are identified from comparative analysis against a reference genome (175). This implicates that the analysis is dependent on the quality and the appropriate selection of the reference genome, which could jeopardize the resolution and accuracy needed in order to carry out adequate public health and/or clinical decisions.

3.5 Added value of whole genome sequencing in public health microbiology

The development of microbial genomics has provided access to large amounts of information; genome analysis is currently an important part of the clinical microbiology field, particularly thanks to the availability of large number of available human/animal bacterial pathogens supporting the fields of diagnostics, epidemiology, pathophysiology and treatments (184). Microbial genomics enable identification and analysis of particular markers such as antibiotic resistance genes, virulence genes and genes transference that may be important to consider when making decisions regarding to treatment and prognosis of the current infection under study. The four major applications of WGS in the context of pathogen characterization are: identification, typing, resistance detection and virulence gene detection (185). Initially, genome sequences were mainly generated to answer research questions, mostly related to evolution concepts (144). Currently, WGS plays a key role particularly for those organisms that are unable to be identified based on the routine established screening methods and where culture preparation for further identification is not a possibility (148).

Metagenomics, the culture-independent analysis of all nucleic acids from a sample, has the potential to improve the detection of both known and novel microorganisms. Laboratories can rely in a single sequencing based test, which would allow identification of most microorganisms in a sample without the need of a traditional culture (186-187). This relatively new approach to infectious disease investigations has become increasingly common in public health laboratories worldwide, being performed directly from clinical samples belonging to outbreak investigation or complex diseases (188).

The technology of WGS is rapidly becoming routine and shared between diagnostic fields. The potential of accurate pathogen detection tests by metagenomic sequencing has been demonstrated in numerous studies and clinical contexts, holding great hopes as an improvement tool for infectious disease diagnostics (189). On the typing field, there's hardly any question regarding the explicit association/application between typing bacterial pathogens using WGS in terms of infection control, surveillance and/or outbreak investigation. There are a number of traditional typing methods, described in detail in Chapter 3/Subchapter 3.3, which are often performed in centralized reference laboratories and not in routine diagnostic ones. This case is well represented with the surveillance of *Listeria monocytogenes*, for which established typing methods include serotyping, binary typing, ribotyping, MLVA typing, PFGE and MLST typing. With this diversity of diagnostic methods, as mentioned earlier, comparison of results becomes difficult and challenging. Sequencing technology provides an opportunity to perform analysis at a widely comparative as well as highly discriminatory level (190).

The resistance detection related application for WGS is based on the fact that previous research has showed the capability of this science to assist with antimicrobial resistance detection. Analysis using WGS generated data can detect particularly acquired resistance by detecting

particular gene expression profiles (191). In terms of its relationship with the treatment field, WGS could reduce diagnosis time and therefore reducing the potential risk of exposing a patient to ineffective drugs increasing the risk of generating additional or new resistance profiles for particular cases (192). The ability to detect virulence genes is an important use of the WGS generated data. Research related to the ability of WGS to detect presence of particular virulence genetic markers that increase its pathogenicity has been published in organisms such as *Staphylococcus* aureus and *Escherichia* coli and its Shiga toxin (193).

Evidence suggests that WGS is most likely to be used in reference laboratories, tertiary hospital laboratories and research-associated laboratories (194-195). Costs are certainly one of the main significances preventing the general use of NGS in clinical and public health microbiology worldwide. It also highlights the perception of the utility associated to this technology in clinical and public health microbiology, facilitating national and international collaborations (185).

3.6 Genomics and its use in STM-related outbreak investigation

Data generated using WGS complements existing epidemiological tools by enabling reassembly of transmission networks and identification of possibly undetected epidemiological links. WGS has been used to identify outbreak isolates from non-outbreak isolates of some *Salmonella* serovars, including STM (196). From a public health and epidemiologic perspective, robust and higher resolution genomic analysis provided by WGS has turned out to provide details into transmission pathways for several significant pathogens. Although the majority of these published investigations have been conducted retrospectively, the findings in all of them agree to highlight the potential of WGS as a real-time infection control tool (43).

Salmonella associated outbreaks have been involved in diverse scenarios worldwide, with identified sources including agricultural produce such as tomatoes (197); poultry and poultry related products (198); ready –to-eat food suppliers such as dine-in or take away restaurants (199); kitchen utensils and appliances (200); livestock production (201) and also related to cases involving wildlife (202). Considering that an outbreak will always involve at least 2 cases, with a high risk of increasing these numbers in a short period of time, prompt subtyping of *Salmonella* and synthesis of laboratory findings into public health actions are critical in reducing delays in outbreak investigation (203). WGS has proven to have the potential to discriminate between sporadic and outbreak isolates that would not be able to be differentiated by currently implemented subtyping methods (24). Even more relevant, the technology has taken outbreak investigations to near real-time (196).

It has been proven that, while carrying out analysis based on a common STM-MLVA profile known to have been associated with two food-borne outbreaks along with sporadic cases in NSW, by sequencing these STM isolates associated with acute gastroenteritis can highlight distinct foodborne community outbreaks within groups of isolates sharing the same MLVA profile. Genomic analysis itself can also further differentiate sporadic cases from outbreak cases and cluster sporadic isolates within endemic MLVA profiles of STM, which can significantly improve the resolution of public health laboratory surveillance (35). The current technology involving genomic sciences and its applicability to the public health and microbiology field are numerous and are currently being approached worldwide. NSW Health is taking advantage of it by combining the information provided by traditional typing methods and the added value that data generated by WGS technology can provide to STM surveillance. Its use is becoming more massive

every year across Australia due to accessibility and understanding of the technology and its required analysis.

Aims 2 and 3 from this thesis intend to confirm the real utility this technology presents in different scenarios such as the genomic variability within already identified successful STM MLVA clades and the genomic variability present when acute and chronic STM infection in animals takes place, in the hope that some findings will contribute to increase the knowledge surrounding this infection and therefore also contributing in decision making processes against this organism.

Chapter 4: Material and Methods

4.1 Introduction

In order to carry out a full description and analysis of the mentioned aims of this research work and therefore to test the hypothesis considered throughout the investigation process, different approaches have been considered. There are 3 different aims of the project requiring different methods that are to be addressed in this thesis.

The examination of the temporal dynamics of STM infections, which were characterized by specific MLVA profiles tested in the STM isolates and observed through the years, has been presented as a quantitative analysis by summarizing the isolate MLVA profile dynamics obtained from the cases taking place across the mentioned time period within NSW population (objective 1). Also, a small section is presented aiming to test the ability of MLST to infer serotypes using 100 randomly selected *Salmonella enterica* isolates that were previously serotyped by the conventional method. The top 10 most common *Salmonella* serovars in NSW were included and evaluated.

Whole Genome Sequencing (WGS) was applied to STM isolates belonging to objective 2. These isolates were classified as successful and non-successful clades based on the yearly presentation numbers associated to non-typhoidal *Salmonella* (NTS)-STM human cases with some particular MLVA profiles. NSW Health has identified these particular MLVA types as associated with large numbers of STM cases each year. By using WGS technology, we aimed to explore the genomic differences in a set of STM isolates with certain specific MLVA profiles. This would help to explain the mechanism of diversity by the presentation of successfulness and non-successfulness population clades that have been persistent t in NSW over a five years' time period.

In the animal experiment section (objective 3), WGS was performed on a serial of STM isolates obtained through a mouse passage experiment. This section was aiming to explore the within- and between host genomic changes of STM developed through oral inoculations of STM for five generations in mice. This was aimed to exam the transmission mechanism of the pathogen pathogenicity by an animal model and to potentially emulate transmission pathways present in human infections.

4.2 Clinical and environmental isolates included in the study

All *Salmonella* isolates from clinical and environmental sources included in this study were obtained from the NSW Enteric Reference Laboratory (ERL), Institute of Clinical Pathology and Medical Research (ICPMR)-Pathology West. The conventional serotyping based on Kauffman–White Scheme (available on line at: http://www.pasteur.fr/sante/clre/cadrecnr/salmoms/WKLM_En.pdf) were performed for all of the isolates and MLVA detection and profile assigning for STM isolates were conducted previously by this laboratory. The detailed information of the isolates and the particular analyses performed on the isolates within the aims of research will be presented throughout the materials and methods chapter.

The study was approved by the Western Sydney Local Health District Human Research Ethics Committee (WSLHD HREC Reference Number LNR/16/WMEAD/405).

4.3 Multi-locus sequencing typing (MLST)

4.3.1 Bacterial isolate preparation

A set of 126 randomly selected and previously serotyped *Salmonella* e*nterica* isolates collected between 2010-2012 were subjected to MLST typing following the scheme published previously (205). Frozen cultures stored at -80^{o}C were revived on the selective media xylose lysine desoxycholate (XLD) agar plate overnight at 37^{o}C. A single colony from the overnight XLD plate was sub cultured on the Blood Agar (BA) media overnight at 37^{o} C. A single colony was picked up from the BA plate and re-suspended in 200 µL water.

4.3.2 DNA extraction and PCR amplification

DNA was extracted from the 200 µL suspension by boiling for 10 minutes. The suspension was then centrifuged at 14,000xg for one minute and the supernatant was used for PCR. The PCR amplification for *Salmonella* MLST was performed by using the HotStarTaq Master Mix Kit ® (Qiagen, Germany). The sequences of seven primer sets according to the published MLST scheme for *Salmonella* (Kidgell *et al*, 2002) and the seven housekeeping genes involved in the MLST typing were given in Table 4.3.2.

Table 4.3.1 The primer sequences and genes for *Salmonella* MLST typing (Kidgell *et al*, 2002):

Primer	Sequence	Housekeeping gene	PCR product size (bp)
*thrA*F	5'-GTCACGGTGATCGATCCGGT-3'	*thr*A (aspartokinase+homoserine dehydrogenase)	852 bp
*thrA*R	5'-CACGATATTGATATTAGCCCG-3'		852 bp
*purE*F	5'-GACACCTCAAAAGCAGCGT'-3'	*pur*E (phhohoribosylaminoimidazole carboxylase)	510 bp
*purE*R	5'-AGACGGCGATACCCAGCGG-3'		510 bp
*sucA*F	5'-CGCGCTCAAACAGACCTAC-3'	*suc*A (alpha ketoglutarate dehydrogenase ehydrogenase)	643 bp
*sucA*R	5'-GACGTGGAAAATCGGCGCC-3'		643 bp
*hisD*F	5'-GAAACGTTCCATTCCGCGC-3'	*his*D (histidinol dehydrogenase)	894 bp
*hisD*R	5-GCGGATTCCGGCGACCAG-3'		894 bp
*aroC*F	5'-CCTGGCACCTCGCGCTATAC-3'	*aro*C (chorismate synthase)	826 bp
*aroC*R	5'-CCACACACGGATCGTGGCG-3'		826 bp
*hemD*F	5'-GAAGCGTTAGTGAGCCGTCTGCG-3'	*hem*D (uroporphyrinogen III cosynthase)	666 bp
*hemD*R	5'-ATCAGCGACCTTAATATCTTGCCA-		666 bp
*dnaN*F	5'-ATGAAATTTACCGTTGAACGTGA-3'	*dna*N (DNA polymerase III beta subunit)	833 bp
*dnaN*R	5'-AATTTCTCATTCGAGAGGATTGC-3'		833 bp

The PCR reaction was prepared in a total volume of 25 µL reaction mix including 12. 5 uL of 2X Hot Star master mix, 0.12 µL of each of the forward and reverse primers at 0.2 µM concentration, 10.26 uL of the molecular grade and 2 lug DNA extract. The thermocycler profile used was as follows: 95° C for 15 minutes (1 cycle), 95° C for 30 seconds; 55° C for 60 seconds; 72° C for 60 seconds (30 cycles), 72° C for 10 minutes; hold 22° C (1 cycle). All PCR products were visualized on a 2% Agarose gel. Once successful PCR amplification was confirmed, a PCR clean-up was performed by adding 2 µL of ExoproStar ® enzyme (GE Healthcare, USA) in 9ul of PCR product and incubated at 37° C for 30 minutes, followed by a 80°C inactivation step for 30 minutes. 1 µL of the primer, either forward or reverse, at a concentration of 9 µM was then added to the mix and ready to be sequenced by an external service provider.

4.3.3 Sequence result analysis

The sequence data was visually checked and analysed using software Chromas® (Technelysium, Australia) by trimming of low quality nucleotides at the beginning and the end of the sequence. The trimmed forward and reverse sequencing reads were then assembled into a contig. Once the contigs were obtained from all of the seven genes for each isolate, the contig sequences were then submitted to a website based MLST Databases at The University of Warwick (http://mlst.warwick.ac.uk/mlst/mlst/dbs/Senterica) for the allele number and the sequence type assignments. The sequence type (ST) was then compared to the corresponding serotyping result previously identified by the NSW ERL.

4.4 Multilocus variable-number tandem repeat typing (MLVA)

MLVA typing has been performed on all STM isolates at the NSW ERL as part of the NSW state Public Health *Salmonella* enterica serovar Typhimurium Surveillance program, using a national harmonized method based on a previously published method (41). The existing MLVA results were analysed for Aim 1 intending to explore the temporal dynamics and diversity of STM isolates over a time period based on MLVA profiles.

4.4.1 PCR amplification for MLVA typing

MLVA typing was performed through a multiplex PCR reaction by amplifications of five loci including STTR9, STTR5, STTR6, STTR10pl and STTR3 on the STM genome using five sets of primers labelled with fluorescent dyes NAD, HEX and FAM, respectively. The reaction was achieved in a 30µl reaction mix prepared using Qiagen HotStar Taq PCR Mix (Qiagen, Germany) with the mix components as described in Wang *et al*, (2008). The PCR amplification conditions were as follows: an initial denaturation at 95°C for 15 minutes, followed by 25 cycles of 94°C for 30 seconds, 60° for 1.5 minutes and 72° for 1.5 minutes. A final 72°C extension for 10 minutes was performed. The standard gel electrophoresis was performed to visualize the amplification products before the detailed fragment length analysis by capillary gel electrophoresis (CGE).

4.4.2 Fragment length analysis by CGE

The lengths of PCR products amplified on the five loci were measured by CGE using an ABI 3130xl Genetic Analyser (Applied Biosystems) by a service provider. The CGE result was presented as chromatogram data, which was analysed using the Peak Scanner software (Applied BioSystems) to determine the size length of each amplicon. The size length of each sequence was converted to a repeat number determined based on the *S.* Typhimurium LT2 complete genome

sequence (Genbank Accession Number NC_003197). The MLVA result was reported as a string of five numbers in an order of STTR9-STTR5-STTR6-STTR10pl-STTR3. The first four digits represent the number of repeats and the fifth is the actual length of the sequence. In the cases of no amplicon was generated, "0" was assigned.

4.4.3 Study of bacterial population dynamics and cluster definition

All STM isolates serotyped and MLVA typed by NSW ERL during 2010 and 2015 were included in this study (N=11,799). All duplicate isolates and environmental/food samples were removed from the dataset. Epidemiological approaches applied to this investigation include the definition of cluster, which included the following criteria: the identification of five or more isolates with the same MLVA profile over a period of 4 weeks within a particular geographical location (e.g., neighbouring residential postcodes). These clusters have been related to what the NSW Health Department investigated; and the descriptive epidemiological analyses of cluster demographics, including distributions of age groups and gender. MLVA profiles were used in the classification of clusters as endemic/new patterns, successful clades. The divergence observed through time was the main focus of the analysis of the MLVA profiles. Simpson's diversity (D), Richness of species (St) and Population velocity (Pvi) indexes/time scale were also calculated, based on ecological diversity concepts (Magurran, 1988).

Simpson's diversity (D) formula corresponds to:

$$D = 1 - \frac{\sum n\,(n-1)}{N\,(N-1)}$$

where n = number of individuals of each species in time (number of MLVA profiles), N = total number of individuals of all species in time (total number of STM).

All temporal changes in STM population were represented by quarter and yearly figures. Minimum spanning trees were generated using BioNumerics v.6.5 (Applied Maths, Sint-Martens-Latem, Belgium) to measure the genetic distance between the successful MLVA clades from all typed isolates in NSW using the categorical coefficient and no priority rules for the algorithm.

4.5 Whole Genome Sequencing (WGS)

For objective 2, sixteen isolates from three different MLVA profiles that were regarded as successful clades by their presentations of high incidence rates across NSW throughout the years, were selected for WGS. These isolates had a known association with human STM clinical infections with three MLVA profiles identified as STM-clade 3-9-7-12-523, STM-clade 3-10-13-12-496 and STM-clade 3-17-9-11-523. For the first two profiles, six isolates from each were selected. These made up of a single isolate collected and identified each year between 2010 and 2015. For STM-clade 3-17-9-11-523, due to its appearance since mid-2012, only four representatives from each year between 2012 and 2016 were included.

In addition, three extra isolates of different MLVA profiles were also selected for comparison by WGS. These MLVA types were regarded as sporadic or non-successful MLVA profiles based on their presentations at low rates of incidence and association to STM human infections in NSW. Two of these, STM-clade 3-9-7-15-523 and STM-clade 3-16-11-11-523, not only presented low rates of STM human infections compared to the successful clades, but also, these MLVA profiles were very similar to that of the successful clades by presentations of less

than two tandem-repeat differences at one or maximum two loci (Table 4.5), which may suggest a major genetic variation manifested by the repeat differences.

Table 4.5 Details of selected STM isolates from successful and sporadic MLVA profiles for WGS

Isolate	**Year**	**MLVA profile**
STM-01_S46 (STM1)	2010	3-9-7-12-523
STM-2_S27 (STM2)	2010	3-10-13-12-496
STM-3_S58 (STM3)	2011	3-9-7-12-523
STM-4_S48 (STM4)	2011	3-10-13-12-496
STM-5_S13 (STM5)	2012	3-9-7-12-523
STM-6_S16 (STM6)	2012	3-17-9-11-523
STM-7_S35 (STM7)	2012	3-10-13-12-496
STM-8_S31 (STM8)	2013	3-9-7-12-523
STM-9_S8 (STM9)	2013	3-17-9-11-523
STM-10_26 (STM10)	2013	3-10-13-12-496
STM-11_50 (STM11)	2014	3-9-7-12-523
STM-12_S23 (STM12)	2014	3-17-9-11-523
STM-13_S52 (STM13)	2014	3-10-13-12-496
STM-14_S19 (STM14)	2015	3-9-7-12-523
STM-15_S30 (STM15)	2015	3-17-9-11-523
STM-16_S7 (STM16)	2015	3-10-13-12-496
SRR2538292 (STM17)	2010	5-10-12-9-490
SRR2538293 (STM18)	2012	3-16-11-11-523
SRR2538304 (STM19)	2012	3-9-7-15-523

Comparative analysis of these set of isolates was performed focusing on allele difference between successful and non-successful isolates as well as differences for the successful MLVA profiles through time.

For objective 3, transmission pathways of STM in animal model, analysis was focused on the identification of MLVA allele differences and the single nucleotide polymorphism (SNP) by WGS. SNP analysis was performed by using CLC Genomic Workbench (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/). Detailed bioinformatics analyses are described in subchapter 4.6.

4.5.1 DNA extraction for WGS

The DNA from isolates was extracted using an automated DNA extraction instrument of Chemagic Prepito-D ® (PerkinElmer, USA). This equipment allowed us to extract DNA from 12 samples within 80 minutes utilizing the Prepito NA Body Fluid Kit ® (PerkinElmer, USA) and the Plasma protocol following the manufacturer instructions. The procedure is based on magnetic particle separation by usage of beads, which are magnetized by an external magnet, and enables isolation of high quality DNA. This involves: 1) Preparation of a input sample of 200 lug fresh culture colony suspension in Phosphate Buffered Saline (PBS) buffer at ~4 McFarland (~x10^8/mL CFU); 2) Addition of the 200 lug culture suspension with 450 µL Lysis buffer containing 10 µL Proteinase K in a deep-well plate (DWP) and uploading to the Chemagic Prepito-D instrument; 3) Preparation of 150 lug of the magnetic beads and 100 lug DNA elution buffer in the chemagic Tip & Tube Rack and uploading to the instrument; 4) Proceed to the extraction operation following the onscreen protocol. Once the extraction protocol finished, the extracted DNA was treated by adding 1 µL RNase A® (Qiagen, Germany) to degrade single-stranded RNA in the extracts. Additional DNA purification was performed according to the following procedures if required:

45 µL of each DNA extract was transferred to a 96-deep-well plate. Then, after vortexing AMPure XP® (Beckman Coulter, USA) magnetic beads for 1 minute, 81 µL were added to the 45 µL extract. This is a highly efficient purification system to increase DNA quality with no salt carryover.

i) The plate was placed on an OrbiShaker MP® micro plate shaker/vortexer (Benchmark Scientific, USA) for 2 minutes at 1500 rpm followed by 10-minute incubation at room temperature. The plate was then pulse centrifuged for 5 seconds using the PlateFuge Microplate Centrifuge® (Benchmark Scientific, USA).

ii) The plate was placed on Invitrogen™ Ambion™ Magnetic Stand-96 magnetic holder (Invitrogen, ThermoFisher Scientific, USA) for 3 minutes in order to remove supernatant by pipetting without disturbing the beads.

iii) Leaving the plate on the magnetic holder, 200 µL of freshly prepared 80% EtOH was added to each well and incubated for 30 seconds before aspiration using a multi-channel pipette. This wash step was performed twice.

iv) The plate was then span down for 5 seconds and again placed on the magnetic holder in order to remove any EtOH leftover. The remaining beads after EtOH removal were air dried at room temperature for 10 minutes while still on the magnet.

v) Once dry, 30 µL of 10mM Tris-HCl pH 8.0 was added as elution buffer and pipetted up and down 30 times until beads were completely re suspended. Then plate was taken into the shaker for 2 minutes at 1500 rpm, followed by a 5 second spin.

vi) Plate was placed on the magnetic holder again, this time for 3 minutes, allowing transference of the 30 µL of supernatant without the beads into new Lo bind collection tubes.

The DNA extracts were then analysed using a NanoDrop® spectrophotometer (NanoDrop™ ThermoFisher Scientific, USA) to determine the DNA quality by measuring absorbance ratios at A260:A230 and A280:A260. The ratios for the acceptable DNA quality were between 1.8 to 2.0 for both absorbances. A Qubit® assay (Qubit® ThermoFisher Scientific, USA) was also used to determine the final DNA concentration of the extract. The Qubit® working solution was prepared by diluting the Qubit® dsDNA BR Reagent 200-fold in Qubit® dsDNA BR buffer.

Subsequently, 190µL of the Qubit® working solution was placed into two 0.5mL tubes for standards 1 and 2, and 198µL of the working solution in 0.5mL tubes for the samples. Afterwards, 10µL of each Qubit® standard was added to the appropriate tubes and mixed by vortexing for 2-3 seconds. After that, 2µL of DNA extract was added to 198µL of the working solution. The tubes were then allowed to incubate at room temperature for 2 minutes. The selection of DNA and dsDNA Broad Range mode on the Qubit® 2.0 fluorometer (Qubit® ThermoFisher Scientific, USA), was made, based on the assay type to be used for measurement as well as expected DNA concentration.

4.5.2 Library preparation for (WGS)

The Illumina Nextera XT Library preparation kit was used for the genome library preparation. The procedures are outlined below.

Quantitation of DNA by Picogreen measurement: the input DNA for library preparation was 1 ng as required by the Illumina specification. The quantitation of DNA was performed by using Quant-iT™ PicoGreen® dsDNA Assay Kit (ThermoFisher Scientific, USA). A 20X TE buffer (10mM Tris-HCl and 1mM EDTA) was diluted with sterile, distilled, DNase-free water to generate a 1 X Tris EDTA (TE) buffer with a pH of 7.5. The DNA samples were diluted 50-fold in 1X TE buffer

in a 96-well plate to a final volume of 100µL. A 2µg/mL working solution of dsDNA was made by diluting a 500µg/mL Lambda DNA standard stock 250-fold to make a 2µg/mL working solution. A second dilution of the 2µg/mL working solution was made to achieve a concentration of 1ng/mL for the high range standard curve. This was the reference standard for the assay. The standards were added to the 96-well plate with the samples. The Quant-iT™ PicoGreen® reagent (ThermoFisher Scientific, USA) was prepared by a 200-fold dilution of the concentrated reagent with 1XTE buffer. To each sample, 100µL of the Quant-iT™ Picogreen working solution was added, and incubated for 2-5 minutes at room temperature in the dark. Fluorescence was measured using the Victor X Plate Reader (Perkin Elmer Inc. USA) with wavelengths at 485 nm for excitation and 528 nm for emission. The DNA concentration was calculated based on standards of known concentrations referred to the standard fluorescence curve. Samples were then diluted to 0.2 ng/µL with 10mM Tris buffer pH 8.0 using the JANUS® NGS Express Automated Liquid Handling Workstation (PerkinElmer Inc., USA).

Tagmentation of the sample DNA: 5 µL of the sample DNA at 0.2 ng/µL was added to each well of a 96-well hard shell TCY Nextera XT Tagment Amplicon (NTA) plate. Following that, 10µL of Tagment DNA (TD) buffer (Illumina, USA) was added to each well, and mixed with sample DNA by pipetting up and down. 5µL of Amplicon Tagment Mix (ATM) (Illumina, USA) was added to the wells and mixed by pipetting up and down. The NTA plate was then sealed with a MicroSeal® 'B' adhesive seal (Bio-Rad, USA), and centrifuged at 1500 rpm at 20°C for 1 minute. The NTA plate was then placed in a thermocycler at 55°C for 5 minutes to instigate tagmentation. The tagmentation reaction is neutralized through the addition of 5µL of Neutralise Tagment (NT) buffer (Illumina, USA). The NTA plate is sealed again with the MicroSeal® 'B' adhesive seal

(Bio-Rad, USA) and centrifuged at 1500 rpm for one minute and incubated at room temperature for 5 minutes.

Index PCR amplification: The NTA plate was placed on a TruSeq Index Plate Fixture (Illumina, USA) and 15 µL of the Nextera PCR Master mix (NPM) (Illumina, USA) was added to each well. With a multichannel pipette, 5µL of index 2 primers was added to each column of the NTA plate, and 5µL of index 1 primers was added to each row of the NTA plate. The NTA plate was then sealed with a MicroSeal® 'A' adhesive seal and centrifuged at 1500 rpm at 20°C for one minute. The plate was placed in a thermocycler at 72°C for 3 minutes, 95°C for 30 seconds, followed by 12 cycles of 95°C for 10 seconds, 55°C for 30 seconds, and 72°C for 30 seconds, then 72°C for 5 minutes and held at 10°C.

Post-PCR clean up: The NTA plate was centrifuged at 1500 rpm for 1 minute. A multi-channel pipette was used to transfer 50 µL of each PCR product and 30 µL of AMPure XP magnetic beads (Beckman Coulters, USA) into a 96-well MIDI labelled as CAA (Clean Amplified Plate). The CAA plate was then shaken on a microplate shaker at 1800 rpm for 2 minutes, and incubated without shaking for 5 minutes. The supernatant was carefully removed and discarded while the CAA plate was on a magnetic stand. The magnetic beads were then washed with 200 µL fresh 80% ethanol and the plate was incubated on the stand for 30 seconds. The supernatant was once again removed and discarded. The ethanol wash step was repeated once, removing all remaining ethanol. The plate was kept on the magnetic stand and allowed to air-dry for 15 minutes. The CAA plate was then removed from the magnetic stand, and 52.5 µL of resuspension buffer (RSB) (Illumina, USA) was added to each well. The CAA plate was then shaken on a microplate shaker at 1800rpm for 2 minutes, and further incubated at room temperature for 2 minutes without shaking. The CAA plate was once again placed on the magnetic stand for 2 minutes. 50 µL of the

supernatant in the CAA plate was transferred to a new clean 96-well TCY plate (CAN) ready for normalization.

Library normalization: 20 µL of the supernatant from the CAN plate was transferred to a MIDI plate labelled as Library Normalization Plate (LNP). A mixture of 4.4 ml library normalizations additives (LNA1) and 800µl library normalization Beads (LNB1) was prepared for the normalization of about 96 samples. 45 µL of the combined LNA1/LNB1 was added to each well of the LNP plate containing libraries. The LNP plate was sealed with a MicroSeal® 'B' adhesive seal and shaken on a microplate shaker at 1800rpm for 30 minutes. The LNP plate was then placed on a magnetic stand for 2 minutes and 80 µL of the supernatant was removed and discarded. The LNP plate was removed from the magnetic stand and the beads are washed twice with 45 µL Library Normalization Wash (LNW1) in each well. The supernatant was then removed and discard. 30µL of 0.1N NaOH was added to each well, the plate was re-sealed and shaken at 1800rpm for 5 minutes. During those 5 minutes, a new 96-well plate was prepared by adding 30µL of library storage buffer (LNS1) to each well (Storage plate SGP). The LNP plate was placed on the magnetic stand for 2 minutes then 30µL of the supernatant from the LNP plate was transferred to the SGP plate. The SGP plate was sealed with MicroSeal 'B' adhesive seal and then centrifuged at 1000 g for 1 minute. The plate was store at 4oC ready for library quantitation and sequencing.

Kapa library quantification qPCR: The Kapa Library Quantification Complete Kit (KAPA Biosystems) was used to determine if the library preparation worked and the quantity of the library. Firstly, the libraries were diluted by 1 in 8000 in a DNA dilution buffer (10 mM Tris-HCl, pH 8.0). The Kapa q PCR master mix was prepared from the kit and then 16 µL of the mix was dispensed into each well of a LightCycler® 480 96-well plate. 4 lug of diluted DNA samples and Kapa

standards provided with the kit were dispensed into the wells to make up a 20 µL reaction mix. Molecular-grade PCR water was added to negative control wells instead of DNA. The PCR was performed in a LightCycler® 480 machine and the thermal profile for amplification consisted of a denaturation step of 95°C for 5 minutes, followed by 35 cycles of amplification, involving 95°C for 30 seconds and 60°C for 45 seconds. All samples and standards were tested in duplicates. The sample quantitation was referred from the Ct values generated by standards. A Ct value between 13 and 18 was to be considered sufficient concentration for sequencing. Samples that did not generate a Ct value may indicate the failure of indexing the sample needs to be re-indexed. Samples with a Ct value >18 may be of very low concentration.

Library pooling: The SGP plate was centrifuged at 1000g for 1 minute before the library pooling. 5 µL of each library from SGP plate was pooled together into a LoBind tube (Eppendorf, Germany) labelled as pooled amplicon library (PAL). A 30 µL of the PAL aliquot was mixed with 30 µL of freshly prepared 0.2 N NaOH in a separate LoBind tube to denature double-stranded DNA libraries into single strands. The denatured library was then combined with 30 µL of 200mM Tris-HCl at pH 7.0 to neutralize denatured library. The library was then diluted with a pre-chilled Hybridization buffer (HT1) by taking a volume between 30 to 70 µL of the library solution in 870 to 830 µL of HT1 buffer depending on the cluster density. A PhiX genome provided by Illumina was used as a 1% spike-in sequencing control as recommended by the Illumina. The final library to be ready for sequencing was prepared by mixture of 195 lug of the denatured and diluted library, 1.3 lug of denatured and diluted PhiX control at 20pM, and 1103.7 lug of the pre-chilled HT1 (1.3 ml in total).

NextSeq 500 sequencing setup and loading: The WGS was performed using Illumina NextSeq 500 platform. Following the Ilumina's instruction, the Flow Cell, NextSeq reagent and buffer

cartridges were removed from the freezer and thawed at room temperature for 30 minutes before loading the library and upload to the machine. The entire 1.3 ml of the pooled denatured and diluted library was transferred into the designated reservoir on the reagent cartridge. Following the onscreen step by step instructions on the sequencing machine, the flow cell, the reagent cartridge and buffer cartridge were upload to the machine and lot numbers of all item were checked and proceed to sequencing once the run was set up on the BaseSpace Cloud Server-Sequence Hub ® (Illumina Inc, USA) following the procedure instructions.

4.6 Bioinformatics applied for STM genome analysis

Various bioinformatic approaches were employed through this research for aims related to genomic variation in the animal model as well as the comparative genomics for successful and non-successful STM clades. They have been separated into different method description sections in order to make it clearer to the reader.

4.6.1 Identification and analysis of Single Nucleotide Polymorphisms (SNPs)

SNPs were determined using, first, raw reads mapping against a reference genome and secondly by de novo assembled sequences mapped against the reference. FastQ files were imported into CLC Genomics Workbench v 7.0 (CLC bio, Aarhus, Denmark) and reads were trimmed to remove Nextera transposase adapter sequences then mapped to the reference genome of S. Typhimurium LT2 (NCBI GenBank Accession No. NC_003197). Quality-based variant detection was performed using settings of a minimum neighbourhood quality of 15 and minimum central quality of 20. Variant detection thresholds were set for a minimum coverage of 10 reads and minimum variant frequency of 75 %. De novo assembling was performed using the CLC Genomic Workbench by default setting except the contig length was set at great than 500 bp.

All predicted SNPs were visually/manually checked by locating the SNP positions in the read-mapping files and comparing the consensus calls to the reference genome. Gene and amino acid changes associated with each SNP were determined from the annotated reference genome. All identified SNPs were concatenate and made the alignment against the reference genome LT2 using Snippy (https://github.com/tseemann/snippy). The resulting alignment was then performed phylogeny analysis using MEGA 7, which generate a SNP tree using maximum likelihood method.

4.6.2 Analysis of core and accessory genomes

Sequencing data was de novo assembly using SPAdes assembler using accurate parameter. The genome assemblies were than annotated using Prokka ®. The output GFF3 file from Prokka was used as input file for the pan genome analysis using Roary ® (220). The output files from Roary were generated either as a .csv file for a list of annotated genome with the presence and absence of core and accessory genes, or a core gene alignment phylogeny tree and a pangenome matrix.

4.7 Animal model for STM transmission pathways

In order to analyse the presence of these potential variations across the genome and the description of genomically related changes, potentially associated with transmission events, is that an animal model causing STM infection was used, to resemble potential changes occurring on the human to human transmission events.

4.7.1 Selection of a STM isolate for the animal model

A human isolate of STM with a MLVA profile of 3-9-7-12-523 was selected as the initial inoculum in the mouse passage experiment. The isolate was collected from a human stool sample in 2010 and was revived from the culture collection at NSW ERL. The MLVA profile of this

isolate was selected since it had been identified as extremely successful, with its presence associated with numerous salmonellosis cases in NSW and also linked in particularly large outbreaks in terms of number of cases involved, not only in NSW but also in other states across Australia (34). The isolate was sub-cultured on the XLD culture media overnight at 37^{o}C. A single colony from the XLD plate was subculture on a BA plate at 37^{o} C overnight, from which the inoculum sample was prepared for mice experiment.

The inoculum for mice experiment was prepared in PBS buffer as a suspension containing x10^{5}/mL colony-forming units (CFU) of bacterial cells. A volume of 200 µL was inoculated to a randomly selected mouse by oral gavage. To check the actual live cells inoculated into the mouse, a serial dilution of the original inoculum was prepared in PBS by 1 in 10^{4}, 10^{3} and 10^{2} dilutions and a triplicate sample of 10 µL from each dilution was inoculated on a BA plate and incubated at 37^{o}C overnight. The colonies from each dilution were visually counted and the average colony numbers from the triplicate samples were used to calculate the live cells inoculated in to the mouse.

.

4.7.2 Selection of experimental animals

For the acute infection model, an albino, laboratory bred BALB/c mouse strain was selected. In total, 29 female BALB/c strains at age of 6 weeks old, weighing 16 – 18 grams were selected for the acute infection experiment. In the subsequent chronical infection experiment, the mouse strain 129X1/SvJ was selected. This strain has been reported that it presents clinical symptoms of non-typhoidal salmonellosis in a more similar way to that present in human STM cases (Sondberg and Jelsbak, 2016). A total of 10 mice at age of 6 weeks old, weighing 16 – 18 grams, were selected. All experimental mice were kept at the Animal House facilities at Westmead Hospital to establish *Salmonella* infection. Upon arrival, mice were randomly distributed into

pairs. A week of adjustment and settling in was given and they were not deprived of either water or food during settling time. Cages were supplied with bedding for environmental enrichment.

All animal experiments described were approved by the Western Sydney Local Health District Animal Research Ethics Committee (Protocol Number 5126.06.14).

4.7.3 Animal Data Collection

For individual identification purposes, all mice were marked using permanent ink and distinctive colours. During settlement week, and also later after infection, daily clinical examinations took place including daily recording of the following parameters: weight, activity level (categorized as normal, reduced activity and no activity), feeding habits (categorized as normal, reduced feeding, no feeding); appearance of fur (categorized as normal, ruffled, over ruffled); hunched position (categorized as normal, partial round and complete round), ataxia/tremor (categorized as normal, mild and over). After the first week of settling and if clinical examinations wouldn't provide any signs of potential immune system alterations (weight loss or any other abnormal clinical finding), inoculation took place.

4.7.4 Animal inoculation

Infections were performed using the oral gavage technique by administrating 200 µL of inoculum at an infective dose of 10^5 cfu/mL. Appropriate manual restraint methods as well as technique skills were needed in order to minimize the risk of oesophageal trauma as well as avoiding dosage into the lungs, which usually results in the animal showing immediate signs of respiratory distress. If this was to be observed, then the animal should have been humanely killed. Negative controls consisted of healthy, 200 µL saline inoculated mice.

4.7.5 Stool sample collection

Mice were manually and visually inspected after inoculation took place. Initial stool collection took place 72 – 96 hours post inoculation; later on, twice a week stool collections were performed for every inoculated animal. These were collected using a vial containing nutrient broth, incubated at 36^{o} C overnight, vortexed vigorously and then plated on XLD agar plates for another overnight incubation. Enumerated and labelled positive plates were stored in STGG broth if later needed. Positive stool samples were collected for at least 4 days post-inoculation. If the clinical condition of the animal allowed, stools were collected up to 4 weeks post- inoculation. The last one of these stool collection series was used to prepare the subsequent passage inoculum. This one was prepared and treated following the same previously described method of inoculum preparation in subsection 5.7.2. In cases where no positive stool collection was detected, re inoculation of the mice took place under the exact same inoculum preparation.

4.7.6 Tissue collection

Tissues were collected after a minimum of 10 days after the initial positive stool cultures were obtained. Infected animals were culled based on clinical symptoms following the animal ethics guidelines for researches affiliated with The University of Sydney. Use of a carbon dioxide (CO_2) gas chamber was the preferred culling method. For the second stage of the animal experiment, where mice strain modifications occurred, and no CO_2 gas chamber was available. In these circumstances, dislocation of cervical vertebrae was performed. Once the appropriate euthanasia was confirmed, the tissue collection was performed by placing the mouse on adequate surface against the animal's back. Anterior and posterior limbs were held by needles, so the abdomen and skin was not folded and ready to be disinfected with sterile 70% isopropyl alcohol pads.

A longitudinal incision from the sternum's xiphoid process was performed. This allowed abdomen cavity completely exposed for tissue collection. The collected tissues included the liver, spleen and intestine. Both liver and intestine tissue was collected by dissection of the organ and a collecting sample of approximately 1 cm in length. The spleen was collected as total organ.

All tissues were enriched in nutrient media broth and plated on selective XLD agar plates and incubated overnight at 36° C. Once the *Salmonella* positive tissue cultures were obtained on the XLD plate, a single colony was sub-cultured on BA plates and incubated overnight at 37° C. The DNA was extracted from the BA culture following the DNA extraction procedure described in subchapter 4.5 for WGS. In the chronical animal experiment, stool isolate-based inoculation was also performed. Negative controls were established by an inoculation of 200 µL saline in the mice.

4.8 Statistical analysis

Statistical methods for the data belonging to aim number 1, including Spearman's rank correlation coefficient and uni-variable linear regression were performed to investigate the correlation between observed and expected cases of STM; 95% confidence intervals and *p*-values were calculated. A value of $p<0.05$ was considered statistically significant when Chi square test was performed. For objectives 2 and 3, level of significance for comparative results were also calculated with Chi square test. All statistical analysis was performed by using IBM SPSS software® (International Business Machines Corp., US).

Chapter 5: Discriminatory power of MLST and MLVA subtyping methods for STM characterization

5.1 Introduction

The notion that circulating clades of *Salmonella* that cause human disease can significantly evolve over time is explored in this chapter. The results presented directly address the following research objectives: (I) To explore the capacity of MOST to identify and differentiate between the most common *Salmonella enteric* servers co-circulating in NSW, and (ii) to capture and characterize STEM population diversity in New South Wales by use of MOVE typing results. The results presented herein describe the molecular epidemiology of STEM in NSW between 2010 and 2015, and propose a testable framework for monitoring circulating STEM clades. The MLST typing results were presented and defended as oral presentation by the candidate at the Australian Society for Microbiology Annual Scientific Meeting in Melbourne in 2014.

5.2 Utility of MLST for inferring prevalent *Salmonella* enterica serotypes

In order to confirm the ability of MLST to infer serovars as well as including serovars, which were prevalent for NSW and had not been previously reported by others, we randomly selected 126 *Salmonella enterica* isolates serotyped by the NSW Enteric Reference Laboratory between 2010 and 2012. These isolates represented the nine most common serovars associated with human disease in NSW in 2013. The proportion of isolates representing each serovar selected for inclusion in this study correlated with the prevalence of each serovar within the NSW population in 2013 (Table 5.1). In 2013, the majority of human NTS cases in NSW were caused by STM (71.7%; n= 91) followed by *S.* Enteritidis (7.9%; n=10) and *S.* SaintPaul (4.7%; n= 6).

Table 5.1 The S. *enterica* isolates included in this study. The selected serovars followed the respective proportions displayed by *Salmonella* in NSW during year 2013.

Serovars	Proportion within sample set (%)	No of isolates selected
S. Birkenhead	0.8	1
S. Bovimorbificans	3.4	4
S. Enteritidis	7.9	10
S. Infantis	2.5	3
S. Montevideo	2.5	3
S. SaintPaul	4.7	6
S. Stanley	1.7	2
S. Typhimurium	71.7	91
S. Virchow	4.7	6
Total	**100**	**126**

Once MLST typing was performed and trimmed sequences submitted to the NSW ERL *Salmonella* database as described in Chapter 4 (Subchapter 4.3.3) a specific Sequence type (ST) was identified for each isolate. All STM isolates belonged to ST19 (n=86). Similarly, all isolates of SaintPaul serovar were assigned to ST50 (n= 6), S. Virchow strains were classified as ST16 (n= 6), S. Infantis was associated with ST32 (n= 3) and serovars Birkenhead and Stanley were classified as ST424 (n= 1) and ST29 (n= 2), respectively. Thus, serovars could be distinguished based on MLST STs, differences between the profiles within serovars were observed. The sequence type and MLST profiles for each serovar are detailed in Table 5.2. For those isolates belonging to particular serovars where different ST types were inferred, the MLST allelic profile could differ by up to 5 alleles. This is the case of the isolates within the serovar Montevideo.

Two distinct ST types were assigned, ST138 (n= 2) and ST316 (n=1). When comparing these two allelic profiles, only the genes *dna*N and *his*D generated the same profile. All the remaining loci, when compared between each other, have a different assigned allele number.

Differences within STM monophasic isolates were also observed; 3 different ST types were assigned including ST85, ST19 and ST34. Comparison within their allelic profiles was performed

finding differences within all 7 genes for ST85. Differences between ST19 and ST34 were present only at *dnaN*. Allelic profiles from STM and one of the STM monophasic isolates (ST34) were close, only differing in profile belonging to gene *dna*N. Serovars can still be distinguished based on the last position of the profile.

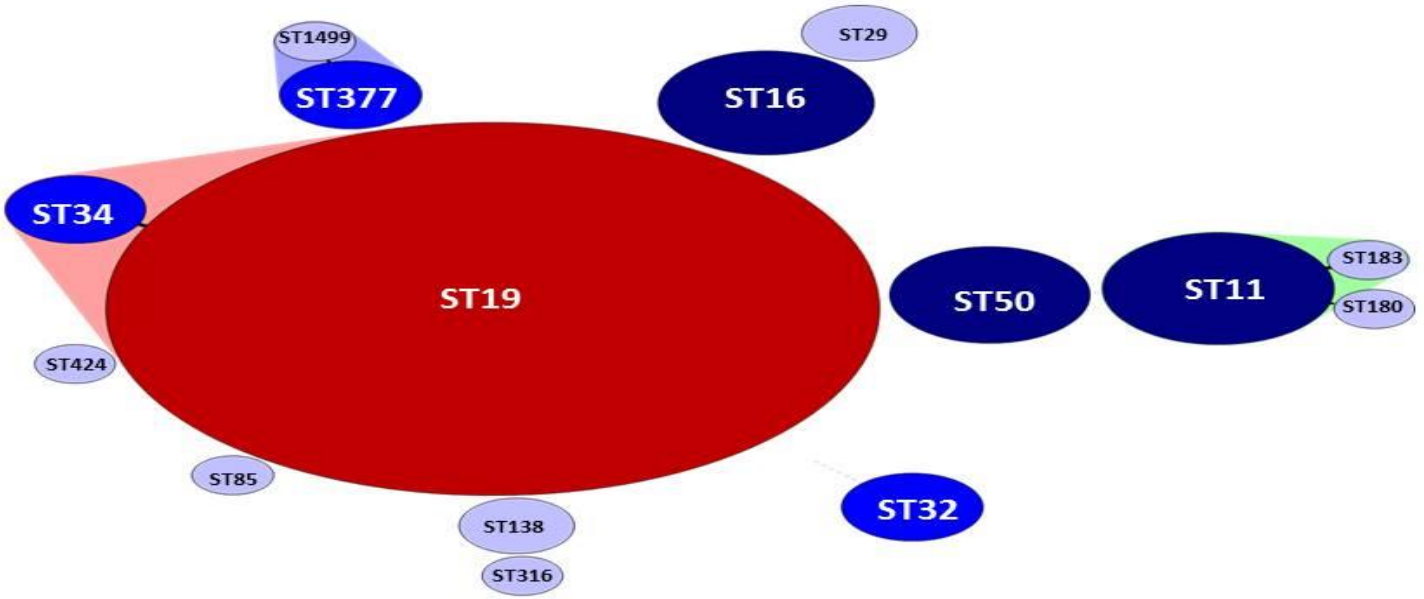
Table 5.2 ST designation for serovars of S. enterica and individual allelic profiles

Serovar	ST	Number of isolates	Allelic Profile							MLST pattern
			aroC	dnaN	hemD	hisD	purE	sucA	thrA	
S. Birkenhead	424	1	143	31	18	143	5	142	22	143-31-18-143-5-142
S. Bovimorbificans	1499	1	2	59	23	64	38	19	12	2-59-23-64-38-19-12
	377	3	2	59	23	64	38	61	122	2-59-23-64-38-61-122
S. Enteritidis	11	8	5	2	3	7	6	6	11	5-2-3-7-6-6-11
	183	1	5	2	3	7	60	6	11	5-2-3-7-60-6-11
	180	1	41	2	3	7	5	6	10	41-2-3-7-5-6-10
S. Infantis	32	3	17	18	22	17	5	21	19	17-18-22-17-5-21-19
S. Montevideo	138	2	11	41	55	42	34	58	4	11-41-55-42-34-58-4
	316	1	43	41	16	42	35	13	111	43-41-16-42-35-13-111
S. Saint paul	50	6	5	21	18	9	6	12	17	5-21-18-9-6-12-17
S. Stanley	29	2	16	16	20	18	8	12	18	16-16-20-18-8-12-18
S. Typhimurium	19	86	10	7	12	9	5	9	19	10-7-12-9-5-9-19
S. Typhimurium monophasic	85	1	45	4	8	44	27	9	8	45-4-8-44-27-9-8
	19	1	10	7	12	9	5	9	2	10-7-12-9-5-9-2
	34	3	10	19	12	9	5	9	2	10-19-12-9-5-9-2
S. Virchow	16	6	6	7	10	10	8	10	14	6-7-10-10-8-10-14

The distribution of the ST types and the relationships between them was visualized using minimum spanning trees. The size of the nodes represented the number of isolates included within that particular node, therefore the larger the node, the larger number the isolates within it. Also, linkage between nodes, expressed by the coloured shading between the ST19 and ST34 nodes, the

ST377 and ST499 nodes and the ST11, ST183 and ST180 nodes, represents a close relationship and genetic distance between sequence types within the same *Salmonella* serovar (Figure 5.1). The results suggest 100% concordance between serovars obtained by serotyping of somatic and H antigens and those inferred from MLST results. Our findings confirmed the ability of MLST-7 scheme to predict *S. enterica* serovars that dominate ecological niches in the NSW.

Figure 5.1 Minimum spanning tree containing *Salmonella enterica* ST types (n=126). The red, green and blue shading refers to ST type relatedness based on allele profile within same serovars. The nodes are assigned different colours based on the proportion of isolates contained in each one of them. Red = > 80 isolates, dark blue = between 5 – 10 isolates, light blue = between 3 – 4 isolates, purple = < 2 isolates.



5.3 Frequency distribution of STM infections based on MLVA profiles

We then reviewed MLVA profiles associated with individual sequence types in order to compare the discriminatory power of two typing methods. The previous subchapter aimed to the capacity of MLST to identify and differentiate between the most common *Salmonella enterica* serovars co-circulating in NSW, which was accomplished, however, there were several distinct MLVA profiles within individual STM STs. Within the 86 isolates serotyped as STM and further classified as ST 19, there were 23 different MLVA profiles identified within these isolates (Table 5.3). These findings reflect the need of a superior resolution tool to accurately investigate the distribution of STM infection in NSW. In order to accomplish this, a total of 11,799 STM isolates were genotyped using MLVA-5 between the 1st of January 2010 and 31st of December 2015. All duplicate isolates and environmental/food samples were removed from the dataset. The number of MLVA typed isolates ranged from 1571 in 2012 to 2547 in 2014; year 2014 represents the highest number of typed isolates during the study period.

The majority of isolates (96.5%) were recovered from stool samples; the remaining isolates were obtained from blood cultures, urine or other tissue samples and aspirates. The number of MLVA profiles present every year varied, with an average of 372 different types (range: 295 – 398). Specifically, 436 MLVA profiles were documented in 2010, 398 in 2011, 295 in 2012, 360 in 2013, 353 in 2014 and 390 in 2015, respectively. The number of yearly STM typed isolates according to particular MLVA profiles is detailed in Table 5.4.

Table 5.3 MLVA profile distribution within STM serovar, STM monophasic and ST types.

Serovar	ST	Number of isolates	MLST Pattern	MLVA Patterns
S. Typhimurium	19	86	10-7-12-9-5-9-19	3-10-13-12-496
				3-10-14-12-496
				3-10-15-12-496
				3-10-7-13-523
				3-10-8-9-523
				3-12-15-13-523
				3-12-15-14-523
				3-12-16-13-523
				3-13-10-12-523
				3-13-14-10-523
				3-14-10-13-523
				3-14-11-12-523
				3-14-8-13-523
				3-16-8-15-523
				3-17-15-14-523
				3-17-9-12-523
				3-9-6-13-523
				3-9-7-12-523
				3-9-7-13-523
				3-9-7-14-523
				3-9-7-15-523
				3-9-8-12-523
				3-9-8-13-523
S. Typhimurium monophasic	85	1	45-4-8-44-27-9-8	NA
	19	1	10-7-12-9-5-9-2	3-12-15-13-523
	34	3	10-19-12-9-5-9-2	4-15-12-0-490
				4-15-11-0-490
				NA

NA= (no applicable). Corresponds to 2 MLVA profiles which at the time of isolates analysis and selection were not available as part of the NSW ERL *Salmonella* database.

Table 5.4 Total STM isolates and number of different MLVA profile identified each year, NSW 2010 – 2015

Year	No. isolates (% of total)	Number of MLVA profiles
2010	2,054 (17.4)	436
2011	1,972 (16.7)	398
2012	1,571 (13.3)	295
2013	1,807 (15.3)	360
2014	2,547 (21.6)	353
2015	1,848 (15.7)	390
Total	11,799 (100)	Average MLVA types: 372

Mean age of patients regardless of gender was 30.09 ±23.81 years, with 41.1% of cases aged less than 14 years. In terms of the presence of the new cases of *Salmonella* evaluated through time and a potential association to demographic factors, no statistically significant differences among either gender or age group were found, the proportion of female (48%) and male (52%) cases was similar ($p>0.05$). Younger age groups, such as children between 0 – 4 years old, were over-represented in terms of new cases, but again, no statistical significance was found for this association. A gradual increase in the number of STM cases was observed during the study period, with some successful endemic STM activity identified. Among the MLVA patterns, six closely related clades, based on the 1 tandem repeat difference present within locus STRR10 and STRR6, were predominant: 3-9-7-13-523; 3-9-8-13-523; 3-9-7-14-523; 3-9-7-12-523; 3-9-7-15-523 and 3-9-8-12-523. These clades represented almost 25% of all STM isolates analysed during the study period and could be clustered as part of one identifiable complex (STM Complex 1).

STM Complex 1 has been present constantly in NSW, being involved in a large number of outbreaks and sporadic cases since the implementation of MLVA typing in 2007. MLVA profile 3-9-7-13-523 was regularly identified and was established as one of the most successful endemic clades through the state. This finding was confirmed by reviewing the whole MLVA dataset for years 2010 – 2015. However, from 2012 the incidence of infections with MLVA profile 3-9-7-13-523 isolates decreased, from 546 during 2010 to 30 in 2015.

One repeat difference clade, MLVA profile 3-9-8-13-523, which was also regularly identified since earlier years also decreased, from 121 isolates in 2011 to only 4 during 2015. A similar decrease was evident for almost all members of STM complex 1 from mid-2012 until late 2016. Profile 3-9-7-14-523 was successful between 2010 and 2011, before decreasing to undetectable levels during 2015. Another significant complex (STM Complex 2) present in NSW

included MLVA profiles 3-16-9-11-523, 3-16-9-12-523, 3-17-10-11-523, 3-17-9-11-523 and 3-17-9-12-523. Though some of them were not detected in NSW earlier than mid 2012 - early 2013, there has been a slight increase in the number of outbreaks caused by STM Complex 2 from 2015. The number of isolates involved in STM infection and clustered within complex 2 remains less than the ones presented within STM complex 1. STM complex 2 exhibit a larger variety of related clades with one tandem repeat difference either in loci 3 or 4. This complex emerged strongly in 2012, being identified in over 150 cases of disease. Even though the incidence of complex 2 infections decreased during 2016 and early 2017, it's still observed yearly, with numbers of new cases of approximately 45.

When the incidence of a particular complex 2 MLVA profile decreased presentation during a particular year i.e. profile 3-17-9-12-523 between years 2013 and 2014, profiles with a one tandem repeat difference, i.e. 3-17-9-11-523, became associated with a large number of STM cases (refer to Table 5.5 to review this relationship). Distribution of the most common MLVA profiles identified during the study period was also represented using minimum spanning trees (MST) (Figure 5.2). Even though MLVA profile members of STM complex 1 have been present in NSW for longer period of time, when reviewing the total number of clades per complex between 2012 and 2015, it becomes evident that this complex has not been the most successful one in years, hence STM Complex 2 has become highly successful in terms of STM cases presentation as presented in Table 5.5.

During 2010, a large number of isolates were part of STM Complex 1. Following years, 2011-2012, exhibited the same complex distribution: majority of STM cases still attributed to endemic profiles including 3-9-7-13-523 and 3-9-8-13-523; since 2012 though, the appearance of new successful patterns, displayed as large nodes challenging those linked to complex 1 became

evident. Several large nodes were evident in year including patterns belonging to both described complexes, 1 and 2, as well as additional new MLVA profiles (3-10-14-12-496 and 3-10-13-12-496), which were linked to particularly large outbreaks (Table 5.6). The location of theses nodes within the whole spanning tree create a separate branch that was not present between 2010 and 2012 (Figure 5.2). Year 2014 presented the largest number of STM isolates through the 6 years studied. Presentation of complex 1 and 2 remained, but complex 1 is represented by smaller nodes, being replaced by complex 2 isolates as well as novel MLVA profiles causing specific outbreaks (Table 5.6). The characterization presented during year 2014 is similar for year 2015 in terms of the predominant presence complex 1, in addition to previously define as sporadic MLVA profiles.

Table 5.5 The most prevalent MLVA clades over the period of this study (total counts and percentages shown).

MLVA profile	Complex	2010	2011	2012	2013	2014	2015	Total
3-9-7-13-523	1	546 (51.8%)	264 (25%)	106 (10%)	90 (8.5%)	19 (1.9%)	30 (2.8%)	1055
3-9-7-12-523	1	61 (23.5%)	24 (9.2%)	14 (5.4%)	19 (7.3%)	102 (39.2%)	40 (15.4%)	260
3-9-7-14-523	1	130 (42.2%)	106 (34.4%)	10 (3.3%)	49 (15.9%)	13 (4.2%)	0	308
3-9-8-13-523	1	82 (18.7%)	121 (27.6%)	130 (29.7%)	89 (20.4%)	12 (2.7%)	4 (0.9%)	438
3-16-9-11-523	2	ND	ND	ND	26 (16.1%)	91 (56.1%)	45 (27.8%)	162
3-16-9-12-523	2	ND	ND	63 (41.7%)	37 (24.5%)	49 (32.5%)	2 (1.3%)	151
3-17-10-11-523	2	ND	ND	ND	5 (4.5%)	93 (84.5%)	12 (11%)	110
3-17-9-11-523	2	ND	ND	5 (1.5%)	40 (11.6%)	213 (61.7%)	87 (25.2%)	345
3-17-9-12-523	2	ND	ND	151 (46.6%)	163 (50.3%)	8 (2.5%)	2 (0.6%)	324
Total clades per complex	Complex 1 Complex 2	819	515	260 219	247 271	146 454	74 148	2061 1092

ND = non-detected

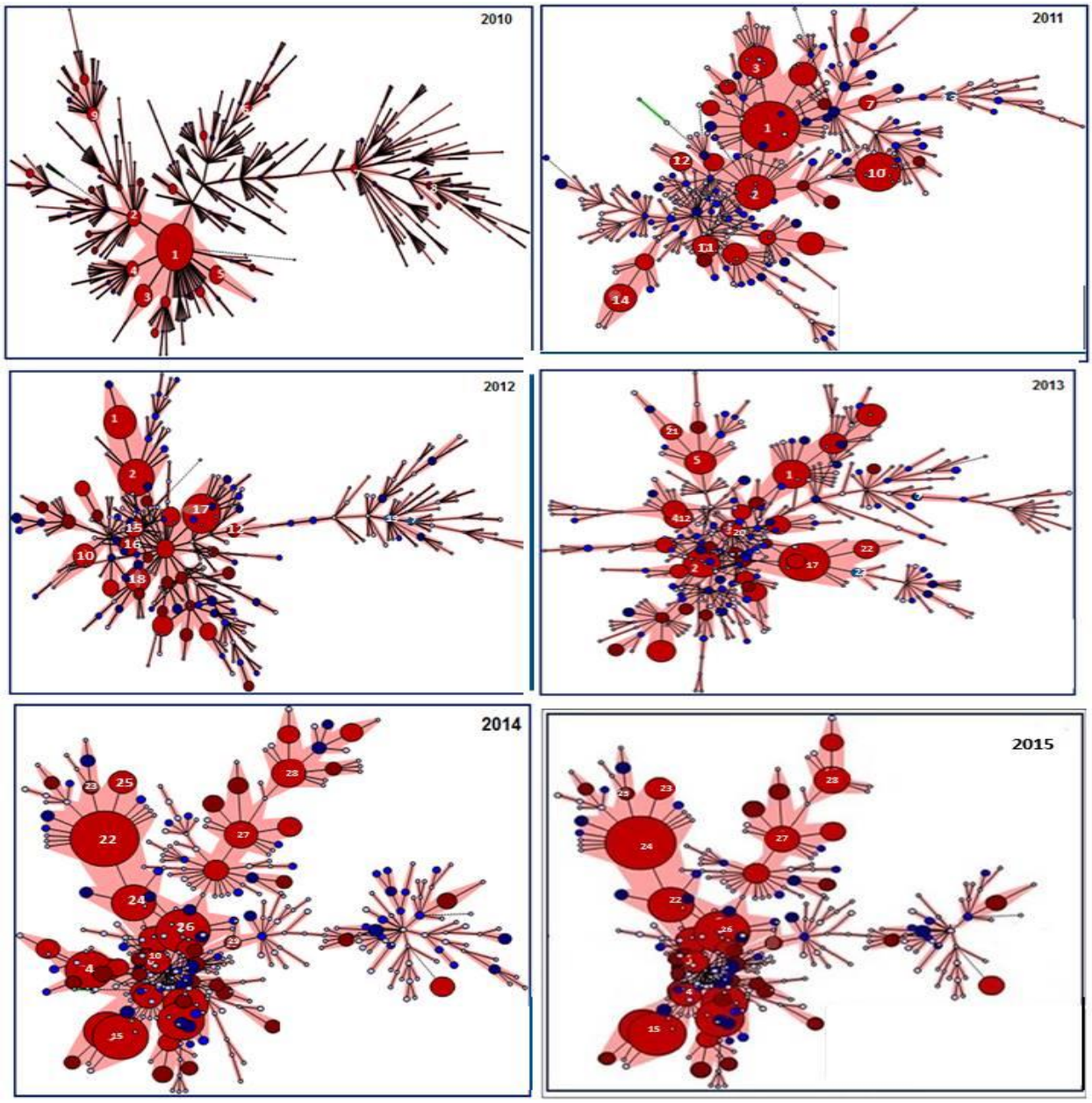


Figure 5.2 Minimum spanning tree for main MLVA profiles through 2010-2015. Numbers 1 – 5, 15 -16 within nodes correspond to profiles belonging to STM complex 1. Numbers 17 – 18, 22-25 within nodes correspond to profiles belonging to STM complex 2.

In order to further explore the distribution of STM, we then analysed the seasonal trends of the most common MLVA profiles associated with STM cases followed the expected STM seasonal pattern as displayed in Figure 5.3, with an increased number of cases during the warmer months. Two seasonal peaks, concentrating the larger number of cases based on "month of sample for culture collection", have been usually observed: during Quarter 1 (January – March) and Quarter 4 (late September - December). The peaks in terms of increased number of cases during warmer months seem to be expanding to include cooler months such as May. For this month, current presentation of STM cases has shifted from an average of 6.4% of cases between 2010–2012 to an almost 12% between 2013 – 2016. Regarding MLVA profile distribution across the whole period, Figure 5.3 displayed the dominance of particular profiles belonging to clonal complexes through identifiable years.

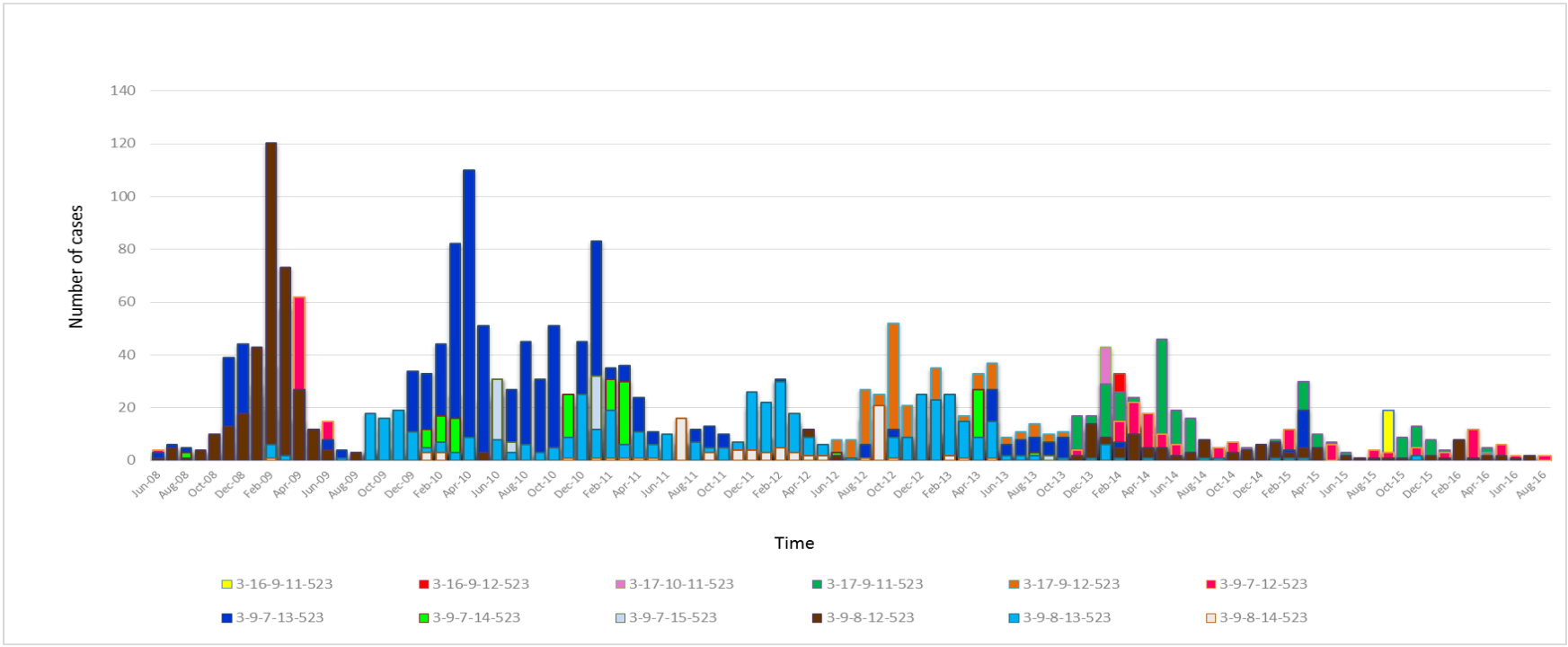


Figure 5.3 Temporal trends of successful STM MLVA profiles, NSW 2008 – 2016. The bars show the individual count for each of the considered MLVA profiles. Bars are coloured to represent the individual MLVA patterns specified. The bar height represents number of cases of each profile/month, etc.

5.4 MLVA profiles associated with community outbreaks

Over the 5-year study period, there was no significant change regarding clustering of typed STM isolates as shown in Figure 5.2. However, these clusters present a particular trend in NSW: majority of them are small in size (less than 25 confirmed cases) and often within a limited geographical area (based on patient's residential postcode), therefore usually demonstrate close geo-spatial distribution. In occasions, there are not always environmental samples clearly identified along the epidemiological investigation which test positive during outbreak investigation; however, based on epidemiological surveillance, a majority of these clusters were linked to a common source. Between 2010 and 2015, 113 confirmed outbreaks associated with STM were identified. The majority of the outbreaks occurred within Q4-Q1. When analysed over time, annual differences in the predominant PT of outbreak-associated isolates were observed.

Table 5.6 describes the presentation of outbreaks yearly. In 2010, the MLVA profiles commonly identified as the source of 18 outbreaks belonged to PT170 corresponded to 3-9-7-13-523, 3-9-7-12-523 and 3-9-7-15-523, all of them presenting 1 repeat difference within STTR10. From year 2011, PT170 organisms were still present associated to outbreaks but also un-related novel profiles emerged, such as 3-10-7-13-523. Rare MLVA types such as 3-11-11-9-523 and 3-14-8-12-523 belonging to PT135 and PT6 were associated with a limited number of outbreaks. 2011 presented less number of outbreaks than the previous year (n=13); 84.6% of them were linked to a particular food vehicle, predominantly raw eggs.

Profiles 3-17-9-12-523 and 3-10-8-9-523, belonging to PT153 were highlights among the endemic PT170 organisms identified during 2012. This year was characterized by the largest number of outbreaks of salmonellosis (n=27), 85% of them were associated with STM. In 2013

the number of STM outbreaks decreased (n=9), and these were mostly associated with PT135 profiles 3-17-9-12-523, 3-13-11-9-523, as well as the endemic PT170. 2014 displayed an increase in number of outbreaks (n=26), however trend continued as described earlier in terms of a decrease in association of them to endemic PT170 clades; PT170 was still represented but mostly related to other MLVA profiles. An increase on PT135 representatives, including uncommon MLVA profiles was also observed. Regarding PT 9 related patterns such as 3-10-7-12-523, 3-10-13-11-496, and their respective clades, confirmed association to several outbreaks during 2015 was observed, manifesting a diversification of MLVA patterns associated with STM outbreaks through time (Table 5.6).

Table 5.6 Significant STM associated foodborne outbreaks in NSW during 2010 – 2015

Year	PT	MLVA profile	No. of outbreaks	Cases within outbreak
2010	170	3-9-7-13-523	9	82
	170	3-9-7-15-523	3	19
	170	3-9-7-12-523	2	25
	170	3-10-7-15-523	2	5
	170	3-9-7-14-523	1	16
	170a	3-14-8-14-523	1	7
	9	3-10-13-12-496	1	4
	9	3-21-12-13-523	1	9
	9	3-27-16-12-526	1	168
	135a	3-11-11-9-523	1	7
	204	3-11-10-9-523	1	4
	6	3-14-8-12-523	1	10
2011	170	3-9-7-13-523	1	6
	170	3-9-7-14-523	1	6
	170	3-9-8-13-523	1	17
	170	3-9-8-14-523	1	13
	44	3-10-8-9-523	2	93
	135	3-12-9-10-550	1	9
	135	3-13-11-9-523	1	4
2012	170	3-9-7-13-523	3	27
	170	3-9-9-12-523	5	65
	170	3-10-7-15-523	1	15
	170	3-9-8-13-523	4	19
	170	3-10-7-13-523	1	14
	135a	3-10-8-9-523	1	11
	135a	3-15/16-10/11-523	1	8
	135	3-13-9-11-550	1	4
	44	3-9-8-14-523	1	14
		3-27-8-21-496	1	9
		3-14-9-14-523	1	10
		3-17-9-12-523	3	23
2013	170	3-9-7-14-523	1	7
	170	3-9-8-9-523	1	5
	135a	3-10-7-14-523	1	49
		3-23-23-11-523	1	17
		3-27-8-21-496	1	8

Year	PT	MLVA profile	No. of outbreaks	Cases within outbreak
2014	135	3-12-12-9-523	3	26
	135a	3-10-7-12-523	3	45
		3-12-11-14-523	3	93
	170	3-9-7-12-523	3	21
		3-17-10-11-523	2	57
		3-24-12-10-523	2	17
		3-17-9-11-523	1	19
		3-16-9-12-523	1	26
		3-25-13-10-523	1	6
		3-26-13-8-523	1	13
		3-26-7-20-496	1	11
	170	3-9-8-11-523	1	4
		3-9-8-12-523	1	16
		3-13-10-11-523	1	4
		3-26-13-8-523	1	7
2015	170	3-9-7-12-523	1	5
		3-26-17-10-523	1	5
		3-16-9-11-523	1	8
		3-12-11-14-523	2	16
		3-12-12-9-523	1	11
		3-14-9-13-523	1	5
		3-10-9-9-523	1	5
		3-26-13-8-523	1	4

5.5. Using WGS to distinguish outbreaks caused by a particular MLVA profile

Analysis based on the use of WGS was performed in order to examine its utility in discriminating sporadic and outbreak linked STM infections within the same endemic MLVA profile. To accomplish this, two epidemiologically independent outbreaks (outbreak A and outbreak M) of a single MLVA profile belonging to STM complex 2, MLVA profile 3-17-10-11-523, in NSW between January and May 2014, along with interspersed sporadic cases were analysed. Outbreak A was linked by NSW Health Authorities to chicken liver pate produced weekly from fresh ingredients at a Sydney café (Food Authority NSW, personal communication, 14 August 2014). Outbreak M appeared to be caused by contamination of a number of foods and environmental surfaces at a Sydney hot bread shop (Food Authority NSW, personal communication, 14 August 2014). All sporadic cases occurred within two months of one or both outbreaks as shown by Figure 5.4.

Of the 85 isolates of MLVA 3-17-10-11-523 recovered from human cases diagnosed between January and May 2014, 56 isolates were epidemiologically classified as belonging to either outbreak from the available data and therefore sequenced for this analysis. A breakdown of the 56 isolates regarding outbreak and type of isolate can be reviewed in detail in Appendix A. Following WGS, two case isolates from Outbreak A and two from Outbreak M were excluded from the analysis due to the poor quality of sequences. The total number of case isolates included in the study was 52.

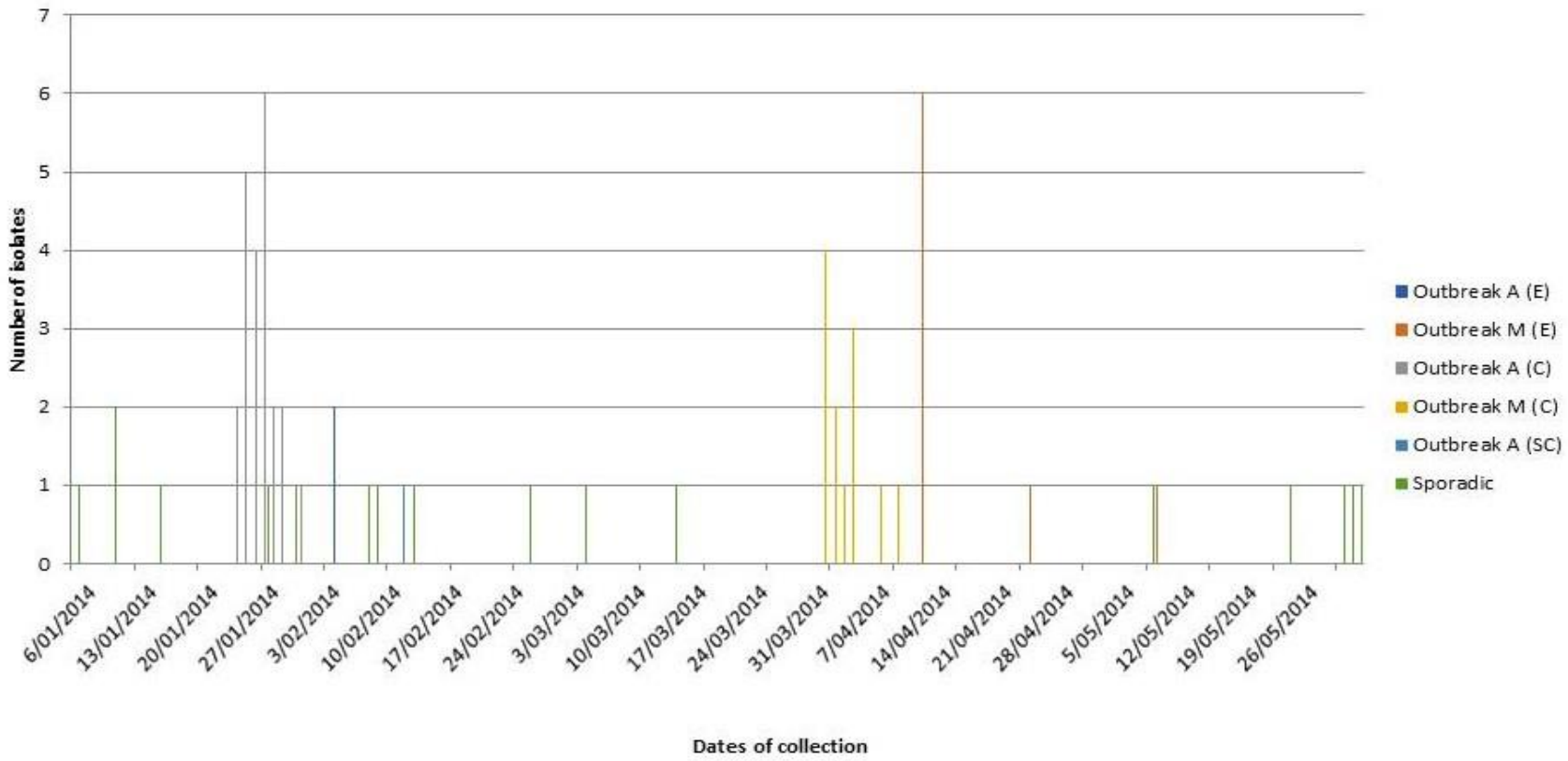


Figure 5.4 Distribution of outbreak related isolates by date of collection. Outbreak isolates are labelled "C" (case), "E" (environmental) and "SC" (secondary case).

Outbreak isolates clustered into two groups consistent with their known epidemiological sources on the basis of SNP differences (Figure 5.4). All case isolates from outbreak M had a unique non-synonymous SNP (C to T) at position 723663 (reference STM LT2 base position), which was also present in the environmental and food isolates obtained from outbreak M. All of the isolates from outbreak M also had a single nucleotide insertion (A) at position (A) at position 1789781.In addition, two isolates from outbreak M carried unique genomic variations, a deletion at position 2162287 and a synonymous SNP at position 3983630. Along with SNP differences described above, a seven base-pair deletion at position 2332558 was identified. This deletion was present in all isolates from Outbreak M but not those from Outbreak A. (see supplementary material for a complete SNPs table). All case and environmental isolates from outbreak A shared a unique single nucleotide insertion (A) at position 7408. In outbreak A, all of the case and environmental isolates from outbreak shared a unique single nucleotide insertion; only 1 of the case isolates also had a unique SNP at position 185058 and clustered separately from major outbreak A (Figure 5.4). Sporadic case isolates showed distinct SNP profiles compared to the isolates from Outbreak A and M, which are displayed independently in the MST as miniclusters detected by WGS. Most sporadic cases were more than two SNPs away from the outbreaks and from each other, with only a few exceptions (Figure 5.5).

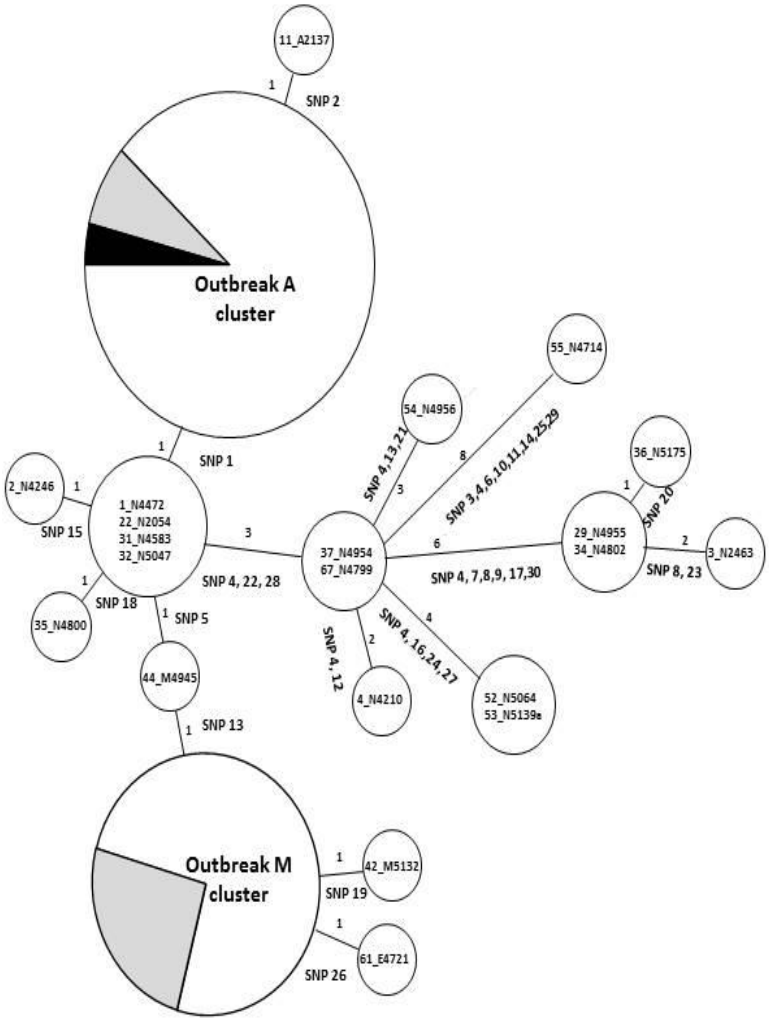


Figure 5.5 Minimum spanning tree of SNPs identified from the 52 STM isolates. Each circle represents isolates with indistinguishable genome; the size of circles corresponds to the number of isolates. The grey colour indicates the proportion of environmental isolates included in the analysis. The black corresponds to the isolate from the secondary case. Numbers above or on the left of the connecting lines between circles are the number of SNPs. SNPs as numbered and named accordingly in supplementary material.

5.6 Discussion

Our findings have illustrated how molecular typing adds important evolutionary perspective to contemporary serotyping of *Salmonella* enterica. The congruence between observed and expected results of MLST-7 indicates that MLST scheme should perform well when applied to *Salmonella* strains circulating in NSW and Australia and can be reliably employed for molecular serotyping of them, minimizing the considerable challenges of serotyping, which include low throughput and high cost (205). We note that our experience has been limited to a relatively small set of 100 isolates and should be verified on a larger set, however, we have been encouraged by recent findings described by Bale *et al* (2016), who reported excellent molecular serotyping ability of MSLT-7 after analysing 29 *Salmonella* isolates with uncommon antigenic formulae, referred to Public Health England (PHE) between 1994 and 2004 (206). A Danish group headed by Litrup *et al* (2010), also established an association between *Salmonella* enterica serovars by means of MLST typing. Though the main aim of their research was to assess the concordance between virulence gene content and MLST data, the assignment of ST types agrees with the ones found in our study for serovars Enteritidis in 10 evaluated strains (ST 11) (207).

Serovar Typhimurium (21 strains), which we identified mostly associated to ST19, ST34 and ST85 in the monophasic Typhimurium isolates, presented a larger diversity of ST types in the Danish study. Still, ST19 remained the most common one, with the addition of ST34, 35 and 376 for 3 individual isolates respectively. From these more uncommon ST types related to serovar Typhimurium, the only one present within our set of isolates was ST34 (n=1).

Our MLST results are similar to the ones described by Achtman *et al* (2012) in terms of relation ST type to designated serovars (205). For instance, within serovar Typhimurium, majority of isolates were grouped within ST19. This complex is the central one, with the largest proportion of

isolates. Soyer *et al* (2009) described STM monophasic variants, identifying independent genetic events, containing distinct variants grouped within ST34 (208). This last ST type appeared to be associated with antibiotic resistance in China (209). There have been 26 different ST types reported by Achtman *et al* (2012) for the Typhimurium serovar, which differs from the 4 found within our group of Typhimurium isolates (205). This can be explained by the difference in isolates number that each project considered. Our project included 91 different Typhimurium previously serotyped human isolates, whereas the mentioned research published in 2012 describes the analysis of 482 Typhimurium isolates. Our isolates corresponded to patients belonging only to Australia, particularly the state of NSW, much smaller when compared to the mentioned study, which included representatives from all continents. Also, our study did not consider other host than human, neither environmental isolates.

Australia differs from many other developed countries, where serovar Enteritidis is the dominant cause of sporadic cases as well as outbreaks, mostly due to sociocultural profiles and dietary habits (210). Isolates from our study belonging to Serovar Enteritidis, were mostly classified as ST11, the most common ST for this particular serovar according to literature that focused on analysing isolates belonging to serovar Enteritidis from Greece (211). Similarly to Typhimurium, our study grouped Enteritidis isolates in one primary complex, which consists of two closely related ST types, which differ by less than 2 loci. These limited loci difference findings correspond to other published results (212). MLST approach can provide discriminatory power when compared to traditional serotyping methods, reproducibility, reliability and portability. In addition, the increasing availability of MLST data in the public domain facilitates method harmonization between different laboratories and enables MLST to be used as a tool to accurately infer ancestral lineages. However, our experience also demonstrated the limits of MLST in

distinguishing between closely related outbreak isolates. In fact the seven-gene approach tends to cluster all Typhimurium isolates into a single sequence type, reason why the application of emergent technologies such as WGS can be used in a far more efficient way in the context of outbreak investigation. Currently, the Public Health England (PHE) is in the process of replacing the conventional MLST methodology with MOST, a method based on short read sequence data derived WGS (213). Since MLST can be inferred by WGS it is clear that the sequencing data generated by this platform is capable of providing more accurate results than the conventional MLST method.

There were differences regarding the serovars included in our study when compared to the ones included by previously published research, which referred to the analysis of the top 20 serovars among US human sources, other 20 among non-clinical non-human sources and an additional 20 human source isolates from other parts of the world (158). Australia's, particularly NSW, top 10 incident serovars differ from other areas worldwide, therefore we were only able to develop comparisons between serovars Typhimurium, including monophasic strains, and Enteritidis. England presents serovars such as *S.* Stanley, *S.* Kentucky and *S.* Virchow within its most recurrent ones whereas for NSW serovar Virchow and Stanley are present but distant apart from the top 10 incident.

Our results have demonstrated the superior resolution of whole genome sequencing. Using an example of one community outbreak, our study demonstrated unique genomic variations in each of the outbreak clusters, with one or two specific nucleotide variations in each outbreak group. This is consistent with a recent report examining 57 isolates of STM across five outbreaks, where within-outbreak isolates were genetically indistinguishable or differed by one or two SNPs (39). The findings also independently confirmed those of another recent study that analysed genomes

of 12 STM isolates and similarly demonstrated the presence of unique SNPs within outbreaks. In this research, isolates from distinct outbreaks differed from each other by more than 10 SNPs whereas in our study, isolates were more genetically similar, as it has been shown in other publications (196). Other studies have also found highly similar isolates within individual outbreaks, along with greater variability between outbreaks, where eggs from the same source were implicated (214, 131).

The MLVA method has been implemented and employed to distinguish variants of *S.* Typhimurium worldwide, providing a discriminatory power exceeding that of the MLST method (131, 93). Furthermore, the clustering of STM isolates using MLVA approach has been correlated well with epidemiological data adding value to public health investigations. This utility is especially important STM was responsible for over 76% of outbreaks in NSW, whereas United States describes an average of 64% of egg-associated outbreaks (215). The analysis of STM infections distribution based on MLVA profiles in NSW, with an average of over 350 different profiles individualized through MLVA typing, is likely to be explained by a relative stability of 'endemic' MLVA profiles grouped within PT170, associated with majority of confirmed outbreaks in NSW until 2012 (22, 27, 30,113-114). This magnitude of MLVA profiles seems not to be unique to Australian settings: the average of 372 distinct MLVA profiles identified through our study is similar to findings described in 2007 (119) in Denmark during a period of 2 years, and to some degree under the 414 distinct patterns described in Belgium (165). These studies including ours confirm the concept of stability of the MLVA loci in terms of reiterative presence of particularly endemic profiles involved in the majority of outbreaks present in different populations.

Just as we observed locally, seasonal peaks are broadly described by Wuyts *et al* (2013), associated to Belgium warmer months (165). Similar trend was previously reported for Australia,

particularly in NSW by Sintchenko *et al* (2012) and by Lal *et al* (2016) for New Zealand; the first one detailed the presence of three seasonal peaks between August 2007 and August 2010, rising from November until March, where high temperatures take place (34). The second authors found relatively high temperatures to be positively associated with risk of infection in Auckland and Christchurch (216). While analysing the presentation frequency of particularly successful MLVA profiles, our study identified endemic and therefore successful profiles, often accompanied by related isolates with minimum allelic variations. Niemann *et al* (2015) described similar conclusions, with loci variation within a particular set of isolates mostly observed at STTR5 and/or STTR6, rarely at STTR3 (217). Additional reports by Chiou *et al* (2010) also labelled loci STTR5 and STTR6 as hyper variable whereas STTR3 was considered as moderately variable (168). Our results highlight that locus STTR10pl has the greatest number of different alleles, followed by STTR5 and STTR6, suggesting a close genomic relatedness as well as the constant appearance of related MLVA profiles associated with outbreaks or large number of cases associated with STM infection.

Our observations that STTR3 and STTR9 are less variable among chosen loci are congruent with reports by Dyet *et al* (2010) (218) and Prendergast *et al* (2011) (219). Hopkins *et al* (2007) have argued that these single-locus variations detected in MLVA profiles can occur worldwide (220), and proposed to consider isolates with single locus differences as related, similarly to what Larsson *et al* (2009) described (163). The uniform descriptions of predominant variants of main persistent clonal lineages could lead to the assumption of them as potential parental clades, as described for clonal complexes 1 and 2 through our findings. Earlier research involving animal and human isolates by Best *et al* (2007) also identified major clusters including large numbers of MLVA profiles (215).

Kurosawa *et al* (2012) from Japan, demonstrated 3 major and 3 minor clusters based on MLVA typing for isolates from bovine Salmonellosis (221). Also, research from Finland by Lienemann (2015), supports the identification of "clonal clusters" based on the isolates MLVA profile, with over 80% of the isolates being identified as clustered, based on MLVA typing (222). Our data and previous findings could suggest that the rarely isolated variants might represent mutant off-springs which could potentially spread out through the community, enhancing a possible accumulation of them in the environment in particular and make contributions to the permanent circulation and persistence of *Salmonella*.

In conclusion, this chapter has examined the capacity of MLST to accurately identify and differentiate the most common serovars of *Salmonella enterica* co-circulating in NSW. Also, it has characterized the profile and temporal changes in STM populations in NSW based on the resolution power of MLVA typing for public health laboratory surveillance, highlighting existence of epidemic clades within specific seasons. We illustrated how WGS of STM associated with acute gastroenteritis can illuminate distinct foodborne community outbreaks amongst groups of isolates sharing the same MLVA profile. Genomic analyses can also further differentiate sporadic from outbreak cases and cluster sporadic isolates within endemic MLVA types of STM, which can significantly improve the resolution of public health laboratory surveillance.

Chapter 6: Temporal dynamics of STM in NSW

6.1 Introduction

In Chapters 5 and 6, results generated using the MLVA typing method has enabled temporal changes in STM populations to be identified. However, several important questions remain unanswered in terms of appearance of changes in STM population related to the increase or decrease in the severity of seasonal epidemics. Also, inquiries related to the real impact of established and new STM clades on the temporal dynamics of STM epidemics are relevant topics of discussion to provide the reader with a panoramic view in terms of temporal-dynamic events involved in STM infection presentation in NSW.

In this Chapter, we further examine seasonal epidemics of STM previously described in Chapter 5 intending to unveil why the severity of annual epidemics in NSW varies and whether there is an association between emergence, increase or decrease of particular STM clades and changes in seasonal severity, generating new insights about the key attributes of STM populations associated with seasonal increases of human STM disease in the most populous state of Australia (Chapter 1, research objective 3).

6.2 Seasonality of *Salmonella* Typhimurium (STM) infections

Seasonal trend of cases followed the expected STM seasonal pattern, presenting an increase of cases during warmer months. Two seasonal peaks were usually found: one around summer (Quarter 1, January – March) and a final one related to late spring –early summer (Quarter 4, late September - December). Refer to Chapter 5 for figures. While analysing the total STM cases in NSW during all years listed in Table 6.1 and correlating them with clustered cases (Chapter 4), seasonality is observed with high incidence peak of total number of cases as well as clustering

present between summer and early autumn months across the whole period. A decrease in cases and clustering occurred around winter and early spring 2011, decreasing the total number of cases as well as clusters by about 50% (Figure 6.1). To investigate changes in the magnitude of annual STM epidemics, Summer-Autumn seasons were divided into 'high' and 'low' based on the incidence of STM cases. We calculated the average number of cases for summer-autumn season in NSW, which equals 1,300 cases and defined a season as 'high' if more than 1,300 cases were reported, whereas "low season" was defined by a total number of cases less than 1,300. For details regarding the STM isolates considered in this section of the thesis, total number of STM isolates serotyped and then genotyped as well as the number of MLVA profiles presented every year, refer to Chapter 5. The increase of STM cases during summer months was reflected in increases in the number of MLVA types detected in the community, as showed in Table 6.1.

The proportions of STM isolates included in clusters showed seasonal fluctuations corresponding to changes in STM incidence. There was no significant change in the number of clusters or their average size over the period of this study (Figure 6.1). Regarding phage type (PT), this typing method was applied in NSW until year 2010 officially. For later years, it has been mostly inferred based on previous results and association to particularly common MLVA profiles. Year 2012 presents some isolates officially phage typed for research purposes (0.7%).

For the remaining years, STM spectrum in NSW, based only on PT inference, was dominated by a small number of phage types, particularly PT170 which was the most common in NSW through the period of our study and through the analysed set of isolates (99.1%). The arrival of PT 135 has been inferred based on recent appearance of other MLVA profiles since 2012. In 2014 we noted the arrival and gradually increasing activity of PT9 with characteristic MLVA

profiles. PTs exit as well as its relationship with particularly successful MLVA profiles can be review in figure 6.2.

Table 6.1 Total STM and MLVA profile count, NSW 2010 – 2015

Year	No. isolates (% of total)	Number of MLVA profiles
2010	2,054 (17.4)	436
2011	1,972 (16.7)	398
2012	1,571 (13.3)	295
2013	1,807 (15.3)	360
2014	2,547 (21.6)	353
2015	1,848 (15.7)	390
Total	11,799 (100)	Average MLVA types: 372

Figure 6.1 Counts of STM cases and clustered cases in NSW between December 2008 and March 2016

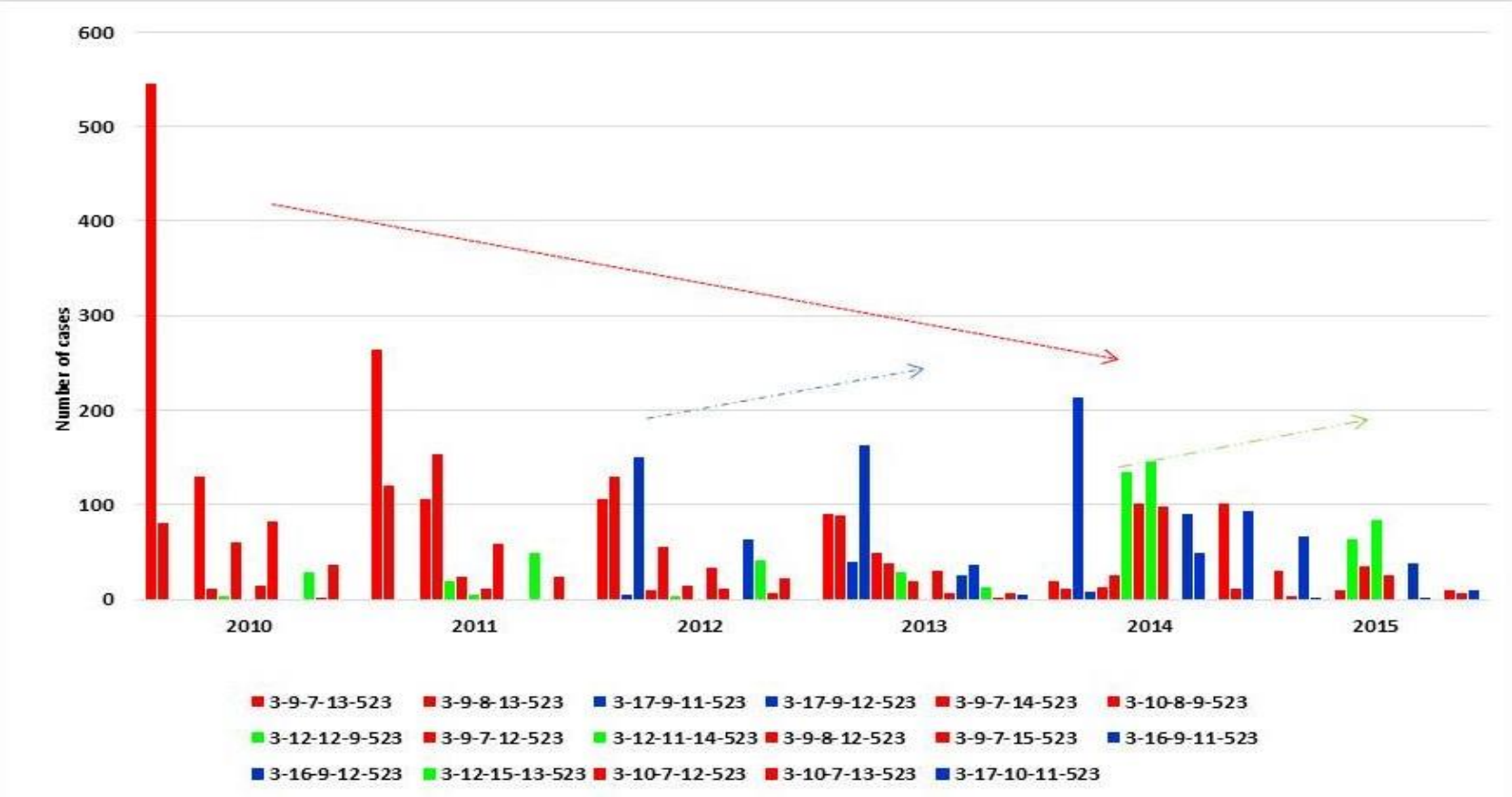


Figure 6.2 Temporal trends of MLVA clades and associated phage types, NSW 2010 – 2015. Each bar colour corresponds to specific MLVA clades. The red arrow represents the exit of PT170; blue dotted arrow represents the arrival of PT135 and the green dotted arrow represents arrival of PT9

6.3. Rise and replacement of successful MLVA clades

Focusing on the main representatives from STM clonal complex 1 and 2, previously described in Chapter 5 (subchapter 5.3) and their general distribution between mid-2008 and mid-2016, Figure 6.3 shows the seasonal epidemics of human STM disease in NSW associated with STM clonal complex 1, which were the predominant cause of STM infections until 2012.

During 2012 there was an evident change in terms of MLVA profiles causing STM infections; the MLVA profile 3-17-9-11-523 and its respective clades emerged and rapidly became the predominant cause of illness (Figure 6.3). MLVA clades belonging to STM clonal complex 2, presented a discrete initial appearance during 2012 (refer to Chapter 5, Table 5.5) and replacement of clades members of the initially endemic STM clonal complex 1. Neither of the individual MLVA profile members of complex 2 has presented the high number of identifications associated with STM infection that a single MLVA profile from STM clonal complex 1 displayed in 2010 and 2011 (Table 5.5, year 2010 and 2011).

From STM complex 2, MLVA profile 3-17-9-12-523 and its one repeat difference clade 3-17-9-11-523, were the ones with the highest association to STM infection cases. When STM complex 2 has dominated the STM population in terms of new cases, STM clonal complex 1 and its members, appeared less common, but it was still identified until mid-2016 (Figure 6.3). There was no apparent association between particular successful MLVA profiles and a specific epidemic season during the study period. All successful profiles fluctuated accordingly to the previously described high and low seasons. The majority of STM infections associated with these successful clades, regardless if they belonged to STM complex 1 or 2, occurred during warmer seasons, regardless of whether it was a high or low season in terms of number of cases as shown in Figure 6.4.

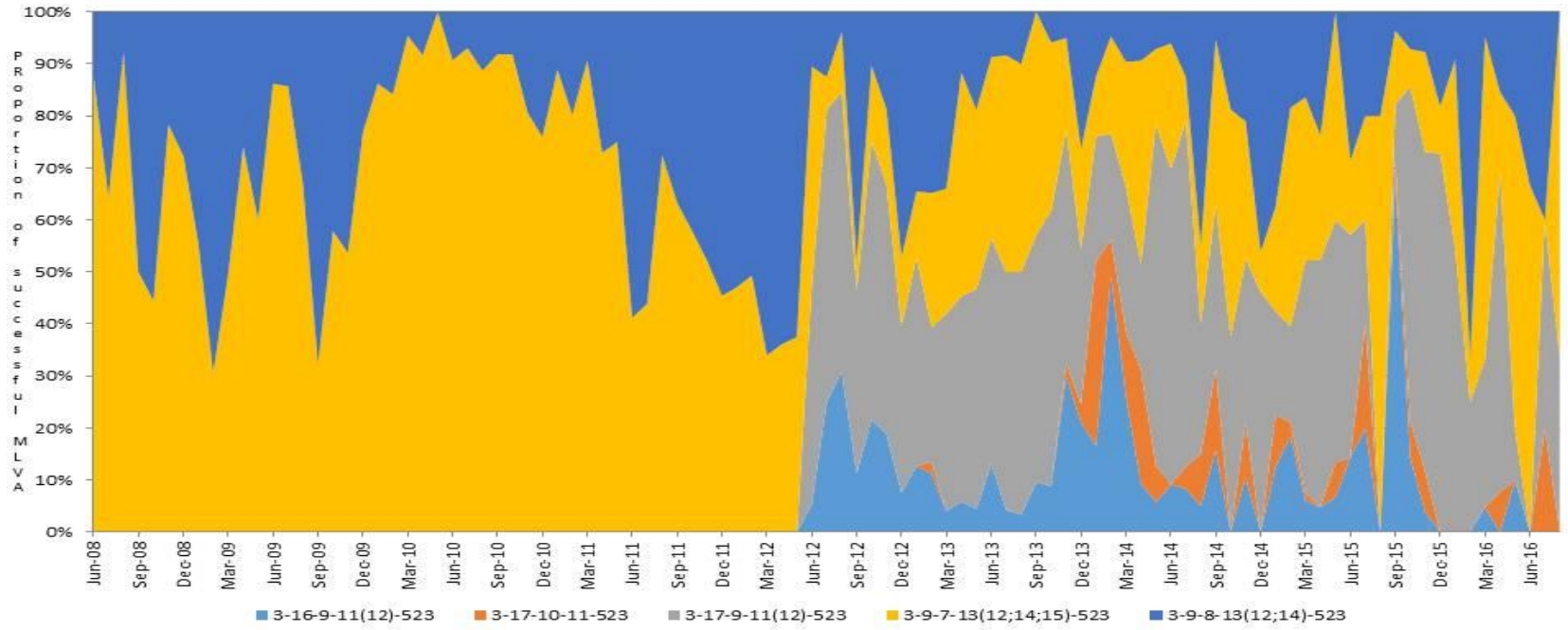
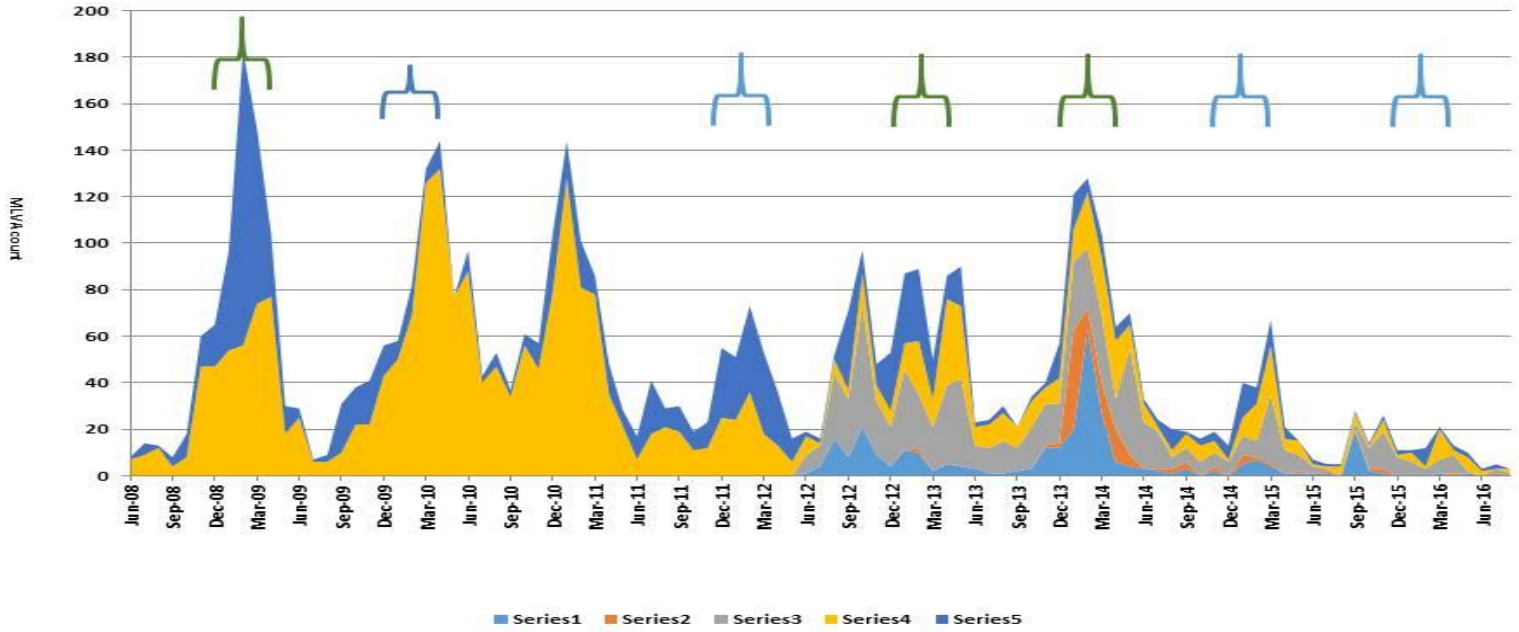


Figure 6.3 Temporal dynamics of a set of successful clades associated with STM infection cases in NSW, 2008 – 2016. The different colours represent proportion of presence for a set of successful MLVA profiles through time.



* High seasons (Summer 2010, 2011, 2014 and 2015) display a blue bracket; Low seasons (Summer 2009, 2012 and 2013 display a green bracket.

Figure 6.4 Seasonal peaks involving successful MLVA profiles and STM infection cases, NSW 2008 – 2016. Series 1 (light blue) corresponds to MLVA profiles 3-16-9-11 (12)-523; Series 2 (orange) corresponds to MLVA profile 3-17-10-11-523; Series 3 (grey) corresponds to MLVA profiles 3-17-9-11(12)-523; Series 4 (yellow) corresponds to MLVA profiles 3-9-7-12 (13;14;15)-523 and Series 4 (dark blue) corresponds to MLVA profile 3-9-8-13(12;14)-523.

By calculating a linear regression of the observed number of incident cases every summer, plotted against the expected number of incident cases related to new MLVA profiles seen in the preceding winter, we estimated that for each incident case associated with a new pattern during Q2-Q3 (autumn - winter season), 4.4 more incident cases are to be expected in the following Q4-Q1 period, (p= 0.032) for the 5 years studied. Though there was no association between particular MLVA profiles and high or low season, the diversity of infections in the preceding winter was a good predictor for a subsequent high season.

6.4 MLVA profile uniqueness

The concept of 'profile uniqueness' considers any MLVA profile present among a particular time period that causes less than 10 cases to be unique. Our results showed that close to 2% of all typed isolates are related to more than 50 STM cases in NSW, however 90% of all MLVA profiles produce less than 10 cases per year, (Table 6.2). Demonstrating the relatively large diversity of "unique" MLVA profiles across NSW, which are associated with yearly STM infections. When correlating the seasonality of the STM cases during summer and winter season, referring to the number of STM infection cases presented (see subchapter 6.2) and the presence of unique MLVA profiles, warmer seasons always displayed a higher proportion of unique MLVA profiles (U_{sp}) present when compared to the number of profiles present during colder seasons (U_{wi}) (Figure 6.5). The difference found between these two was statistically significant (p = 0.006). The proportion of unique profiles in summer looks the same for each year except 2010, regardless of high or low season. The proportion of unique MLVA profiles during summer seasons was always higher during the preceding winter months, progressively decreasing until 2015. In general, during high seasons, the proportion of unique profiles tended to decrease year after year; only during

summer period 2015, there was a minor increase in their proportion. During low seasons, the proportion of unique MLVA profiles presented yearly variations. During winter periods, unique patterns were slightly lower than proportions found in spring and summer.

Table 6.2 MLVA profiles uniqueness linked to STM cases presentation, 2010 – 2015

Years	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Unique patterns linked to < 10 cases	399 (91.5%	367 (92.2%	266 (90.2%	330 (91.7%	309 (87.5%	352 (90.3%
Patterns linked to 10 - 50 cases	30 (6.9%)	24 (6%)	22 (7.5%)	25 (6.9%)	37 (10.5%	33 (8.4%)
Patterns linked to over 50 cases	7 (1.6%)	7 (1.8%)	7 (2.3%)	5 (1.4%)	7 (2%)	5 (1.3%)
Total No. MLVA patterns	436	398	295	360	353	390

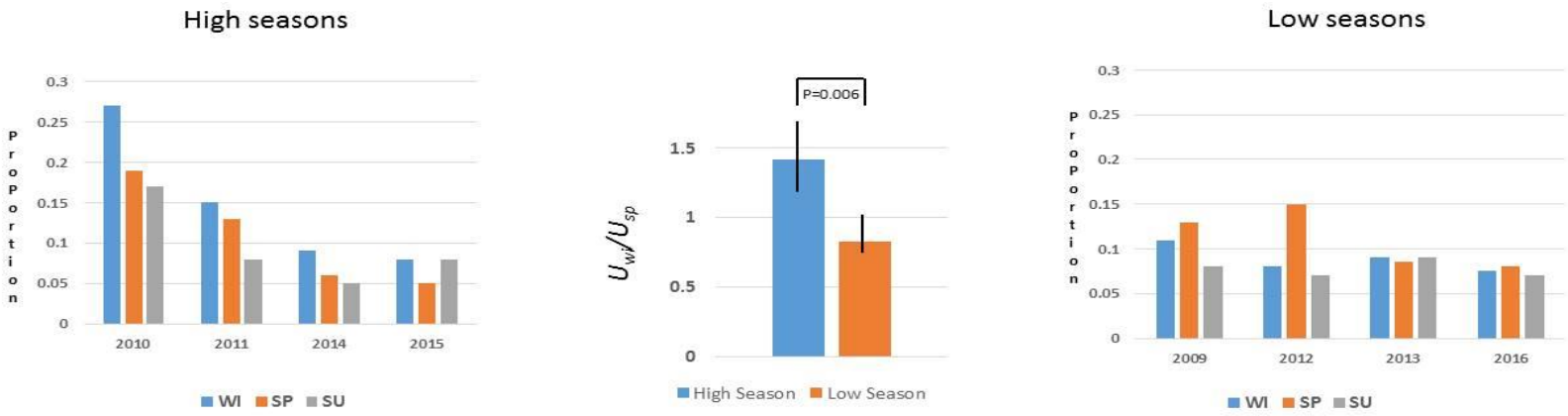


Figure 6.5 Unique MLVA profiles proportion among STM clades present during high and low season. Blue bars correspond to winter months; orange bars correspond to spring months; grey bars correspond to summer months. Middle chart represents difference in presence of unique profiles only in terms of high/low season (associated to the number of STM cases presented during summer when compared to winter season). U_{wi} = proportion of unique profiles during winter; U_{sp} = proportion of unique profiles during spring.

To examine the relationship between the number of MLVA profiles, presence of singletons (MLVA profiles present only once) and the number of MLVA profiles occurring more than once, we employed the concept described in 1988 by Marrugan of "richness of the population". This concept is generally used in ecological sciences to estimate the number of different species in an ecological community. In this case, it has been used to estimate the relative abundance and diversity that MLVA profiles, categorizing them in those present only once and those occurring more than once during particular time, within a particular STM population. It appeared that STM population in NSW was characterized by a high number of singletons, contributing to the season's peaks distributed through the year. Every year, the highest number of singletons was observed during summer months, followed by a progressive decrease through the following autumn-winter months, to then finally rise again during the period between spring and summer. Only during 2014, which recorded the highest number of STM cases during the study period, was the difference in number of singletons present during months related to summer and autumn less visible (Figure 6.6).

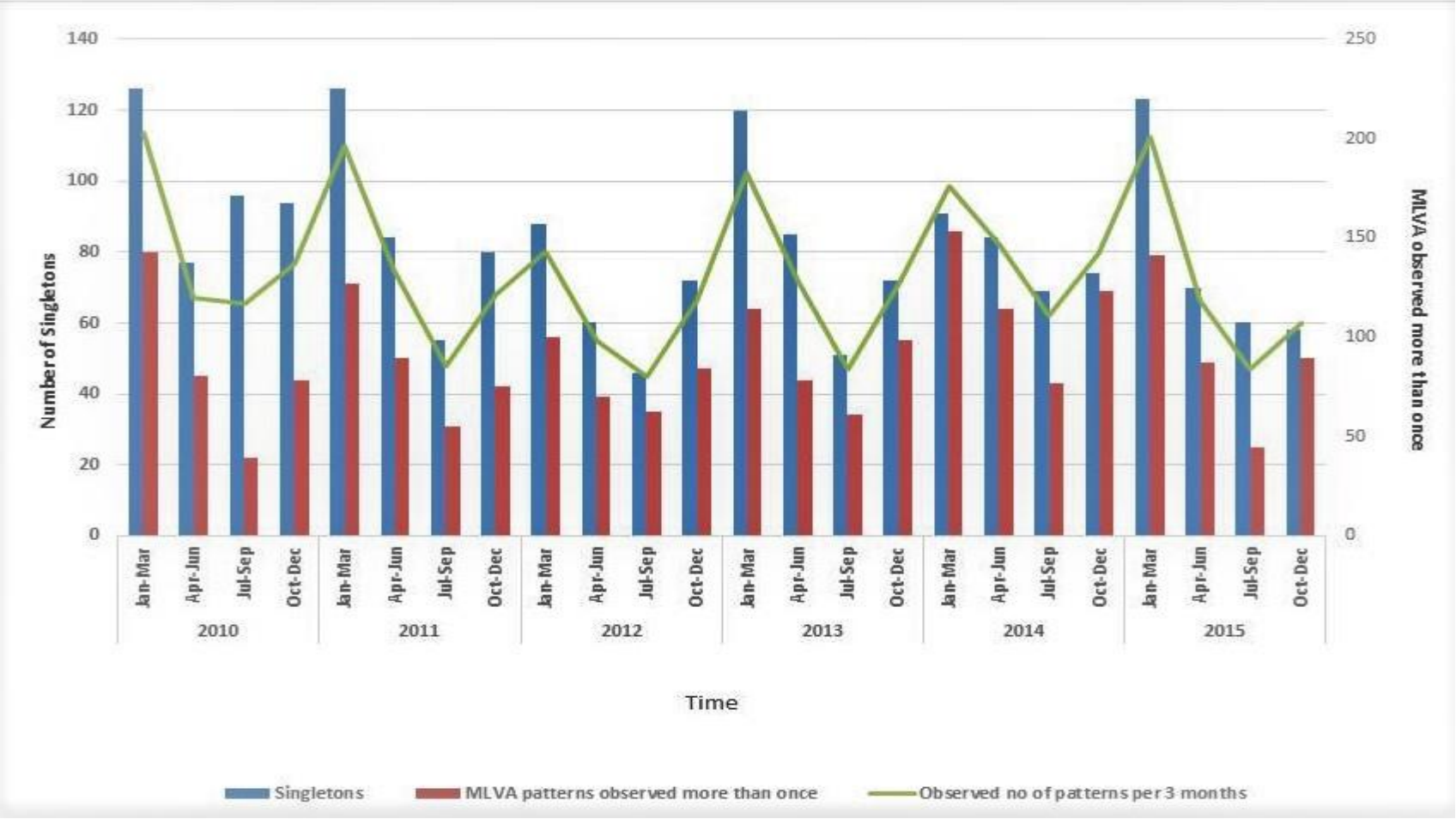


Figure 6.6 Trends of STM MLVA profiles richness based on proportion of singletons, patterns observed more than once and total No. of observed patterns quarterly, NSW 2010 - 2015

6.5 Reduction in newly identified MLVA profiles in winter and spring can predict salmonellosis epidemics

In order to quantify the association between the proportion of novel MLVA profiles and the total case numbers detected during the preceding winter season as a predictor of "cases abundance" for the following summer season, a "Population velocity index" (Pv) was generated. The Pv enables quantification of the relationship between the proportion of new MLVA patterns during April and September (Quarters 2 and 3 (Q2 and Q3)), and the total number of STM cases during the same period as follows:

Pv = (Proportion of new patterns during Q2-Q3 x Total No. of STM cases during Q2-Q3).

Based on the Pv formula, correlation proved that for a particular preceding winter (Q2-Q3) with a low expected number of incident new MLVA profiles, it is likely to expect a lesser number of cases as well as new MLVA profiles during the following summer (Q4-Q1). The actual correlation can be viewed in Figure 6.7.

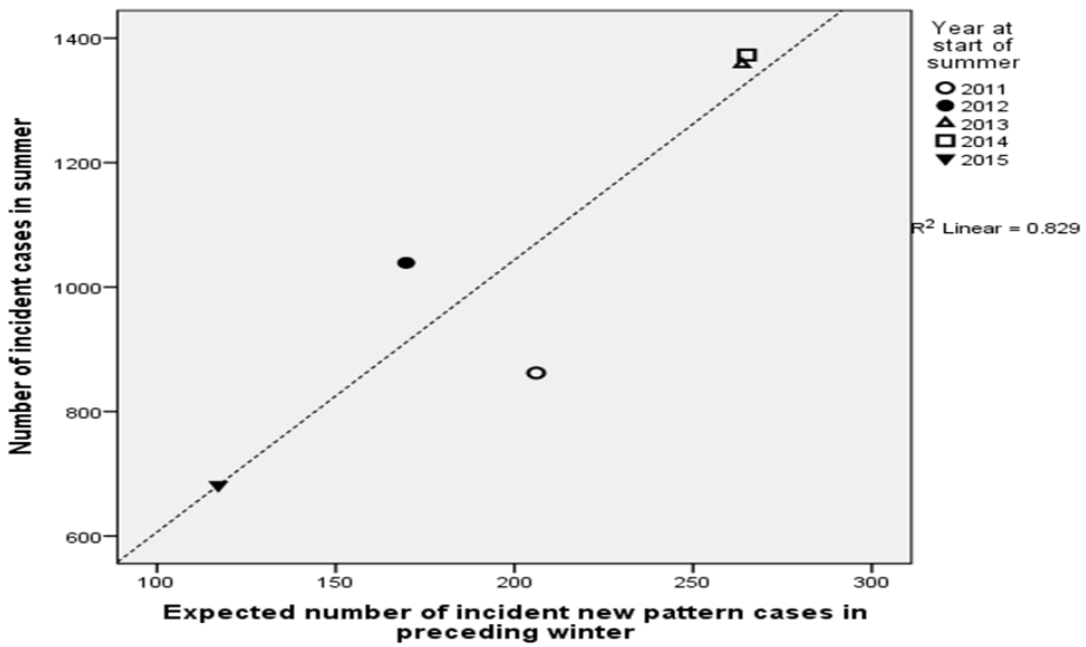


Figure 6.7 Correlation between number of incident cases (associated with presentation of new MLVA profiles) during summer (Q4-Q1 quarters) and the expected number of incident new pattern cases during the preceding winter (Q2-Q3), NSW 2010 - 2015

When estimating Spearman's rank correlation coefficient by associating the expected number of new profiles in winter and the number of new cases in the following summer, for 5 years, the indicator resulted in 0.9 (p=0.037), suggesting that whenever the number of cases associated to new pattern during winter season quarters increases, the number of cases associated to same kind of patterns during summer season quarters will also increase.

6.6 Discussion

Our results have shown how the severity of seasonal STM epidemics can be predicted from small changes in the STM populations recovered from human cases. While we postulated the significant role of newly identified STM MLVA profiles and the importance of the time of their appearance, it was remarkable that established measures of population diversity remained stable. For example, Simpson's diversity index, which measures the probability that two unrelated isolates characterized as being different by MLVA, was consistent over the time of the study. However, it didn't take into account the actual variation of the number of repeat differences at each locus, identifying loci, which showed the least or most diversity. Pendergarst *et al* (2011) calculated diversity within each locus showing high diversity associated to STTR6 and 10 (142). The results reported in this chapter support the initial hypothesis that polymorphism in STM genomes, translated into MLVA profile variations in number of tandem repeats, is a key attribute that, in conjunction with epidemiological evidence, can affect the recognition and investigation of STM transmission events and community outbreaks.

We observed a significant increase in the number of STM isolates through the years, with an average of over 350 different MLVA profiles **individualized** through typing. This increase is usually related to a relative stable presence of the so called endemic MLVA patterns grouped within PT170, which still managed to be associated with majority of confirmed outbreaks in NSW

until 2012 (35-37,127-128). The success of clades related to complexes 1 and 2 throughout the study period could be explained by constant shedding of new STM infection from environmental reservoirs to humans. The average of 372 distinct MLVA profile identified in our study is similar to the ones described by Torpdhal *et al* (2007) in Denmark during a period of 2 years (190) and to some degree also similar to the 414 distinct patterns described in Belgium by Wuyts *et al* (2013). This last study, similar to our approach, characterized Typhimurium populations associated with seasonal epidemics, showing involvement of specifically predominant MLVA profiles causing large numbers of STM cases (122). The descriptions of predominant variants of main persistent clonal lineages could lead to the assumption of them as potential parental clades, as described through our finding as clonal complexes 1 and 2. Data could suggest as well that the rarely isolated variants might represent mutant offsprings, which could potentially spread out through the community, enhancing a possible accumulation of them in the environment in particular, and make contributions to the permanent circulation and persistence of *Salmonella*.

There is a stability of the novel/endemic ratio MLVA profiles to all observed types, which supports the assumption of constant seeding from the earlier set STM infections. The continuous introduction of new STM variants through latest years through NSW collected isolates must be highlighted, most of them not associated with outbreak presentation but with series of individual number of STM infection cases. Though numbers still remain low, the introduction of these novel variants will directly impact the STM population profile; even more, if particularly successful novel patterns raise their incidence, there would be an impact within spread and incidence rates. As stated by Tein Ngoi *et al* (2013), it is expected, when STM is endemic in a region, genetic homogeneity of local STM strains will be found (230). The analysis of potential prediction indexes adds further insights into the natural history of STM infection, proposing a baseline to delineate

cases presentation forthcoming, as well as bringing the need of early warning systems using prospective STM typing approaches on public health outcomes.

These predictors require further analysis incorporating additional years and season. Predicting trends in terms of high/ low epidemic for upcoming warmer season, which will present higher incidence of STM cases, based on the appearance of new profiles and their relationship with total STM cases could be a simple but reliable tool aiming to direct actions as well as higher resolution analysis including WGS, for detection of cases and monitoring.

In conclusion, significant increases in seasonal epidemic of STM can be associated with a reduction in newly identified MLVA profiles in the preceding winter and spring, reflecting the parallel decrease in population diversity and the emergence of successful STM clades under selection pressure. Prospective surveillance of STM based on MLVA-5 can identify the reservoirs of diversity from which future epidemics emerge. These findings also draw support for the conclusion that, in the analysis of seasonal STM epidemic, both their magnitude in case numbers and the diversity of STM subtypes need to be considered for precise public health assessment.

Chapter 7: Genomic analysis of successful STM

7.1 Introduction

Salmonella Typhimurium (STM) has demonstrated remarkable diversity as a zoonotic 'generalist' serovar from which future epidemics and more 'specialist' high virulence strains might emerge. Evidence suggests that different serovars of *Salmonella* have been associated with different risks of invasive disease (218, 231) and diverse virulence traits underlie diverse clinical outcomes (232). Indeed, significant differences in invasiveness involving small numbers of virulence genes have been recently identified within serovars predominant in Europe (233-234) and the recent emergence of STM sequence type (ST)313 in Africa has been linked to invasive disease. Analysis of the ST313 isolates identified genome degradation concentrated in pathways associated with an enteric lifestyle, compared with the most widespread STM ST19 causing diarrhoea in humans. It appears that STM ST313 has been adapting towards a more host-restricted lifestyle typical of *S.* Typhi (225). However, the impact of genomic variation among epidemic STM isolates remains poorly understood. Only one recent study has applied WGS analysis and associated to MLVA typing in terms of presence of these variations in genes, but to *S.* Enteritidis isolates in the Belgian National Reference Laboratory of Foodborne Outbreaks (122,232). Findings from this study, described in previous chapters (Chapters 5 and 6), demonstrated several successful clades that have dominated STM populations in NSW. In this chapter we attempted comparative genomic analysis between NSW sporadic and epidemic STM clades based on their MLVA profile in order to explain their successfulness of particular clades by linking genomic variation with epidemic potential in order to support the translation of WGS into public health laboratory surveillance of salmonellosis.

7.2 Selection of representative *STM* isolates

Sixteen isolates representing three high incidence MLVA profiles across NSW throughout the years 2010 and 2015 were selected. The 3 MLVA profiles considered within this group were: STM-clade 3-9-7-12-523 (PT170), STM-clade 3-10-13-12-496 (PT9) and STM-clade 3-17-9-11-523 (PT9). For the STM clades 3-9-7-12-523 and 3-10-13-12-496, isolates identified by NSW ERL between 2010 and 2015 were included in the study (n=6/clade). For STM clade 3-17-9-11-523, due to its appearance since mid-2012, only 4 representatives were included (n=4). In addition, three strains were selected to contrast the features of successful STM clades. These non-successful isolates represented very uncommon MLVA profiles from the same phage types. One such isolate, STM MLVA 5-10-12-9-490, was observed very occasionally between February and July 2012 (n=5). The other two, STM MLVA 3-9-7-15-523 and STM MLVA 3-16-11-11-523, represented clades associated with low rates of human infection when compared to the epidemic clades. Though they were classified as non-successful due to the limited number of STM cases they were associated with in contrast with the successful clades, they presented one or two tandem-repeat differences in one or two loci when compared with two of the successful MLVA profiles selected. The list of STM isolates selected for this comparative study is summarized in Table 7.1.

Table 7.1 STM isolates selected for the comparative study (n=19). Year of collection and designation (‘successful’/’non-successful’) have been included.

Isolate Name	Year	MLVA profile	Designation
STM-01_S46 (STM1)	2010	3-9-7-12-523	Successful
STM-2_S27 (STM2)	2010	3-10-13-12-496	Successful
STM-3_S58 (STM3)	2011	3-9-7-12-523	Successful
STM-4_S48 (STM4)	2011	3-10-13-12-496	Successful
STM-5_S13 (STM5)	2012	3-9-7-12-523	Successful
STM-6_S16 (STM6)	2012	3-17-9-11-523	Successful
STM-7_S35 (STM7)	2012	3-10-13-12-496	Successful
STM-8_S31 (STM8)	2013	3-9-7-12-523	Successful
STM-9_S8 (STM9)	2013	3-17-9-11-523	Successful
STM-10_26 (STM10)	2013	3-10-13-12-496	Successful
STM-11_50 (STM11)	2014	3-9-7-12-523	Successful
STM-12_S23 (STM12)	2014	3-17-9-11-523	Successful
STM-13_S52 (STM13)	2014	3-10-13-12-496	Successful
STM-14_S19 (STM14)	2015	3-9-7-12-523	Successful
STM-15_S30 (STM15)	2015	3-17-9-11-523	Successful
STM-16_S7 (STM16)	2015	3-10-13-12-496	Successful
SRR2538292 (STM17)	2010	3-9-7-15-523	Non-successful
SRR2538293	2012	5-10-12-9-490	Non-successful
SRR2538304	2012	3-16-11-11-523	Non-successful

7.3 Main features of the core genome of STM clades

As stated in Chapter 3, the core genome includes mainly genes that are essential for the bacterial cell in terms of growth and survival, and are often related to code for enzymes involved on the biosynthetic pathway (152). It is highly conserved in terms of sequence identity as well as in terms of the presentation order through the genome. The remaining 10% of genes has been described as unique to each of the different lineages and has been labelled as accessory genome (152). Comparative revision focusing on the accessory genome of all isolates, successful and non-successful was performed. Preliminary analysis of the accessory genome annotations as well as by using graphical representations such as heat maps, indicated there is a stable segment of the genome, the core genome, which didn't present any variations when compared within successful clades. There were also no variations when core genome from successful isolates was compared against the non-successful ones, regardless the fact that based on number of tandem repeat differences from MLVA profile the 3 non- successful clades are genetically distant; core genome remained stable for all set of 19 isolates as can be observed in the following heat map generated by Roary® (Sanger-Pathogens) (230). (Figure 7.1).

7.4 Variations in accessory genomes

The accessory genome of all isolates demonstrated variations, not only when comparing successful group of isolates against non-successful group of isolates but also when comparing each isolate against each other. Still, within almost all of the successful isolates (STM 1 – 16) and within the 2 non-successful clades, (SRR2538304 and SRR2538292) whose MLVA profile is identifiable as member of the two larger successful clonal complexes described in chapters 5 and 6 as clonal

complex 1 and clonal complex 2, there is a segment of the accessory genome within all of them, where the absence of genes (white colour) and the presence of other (grey) remains similar (Figure 7.1).

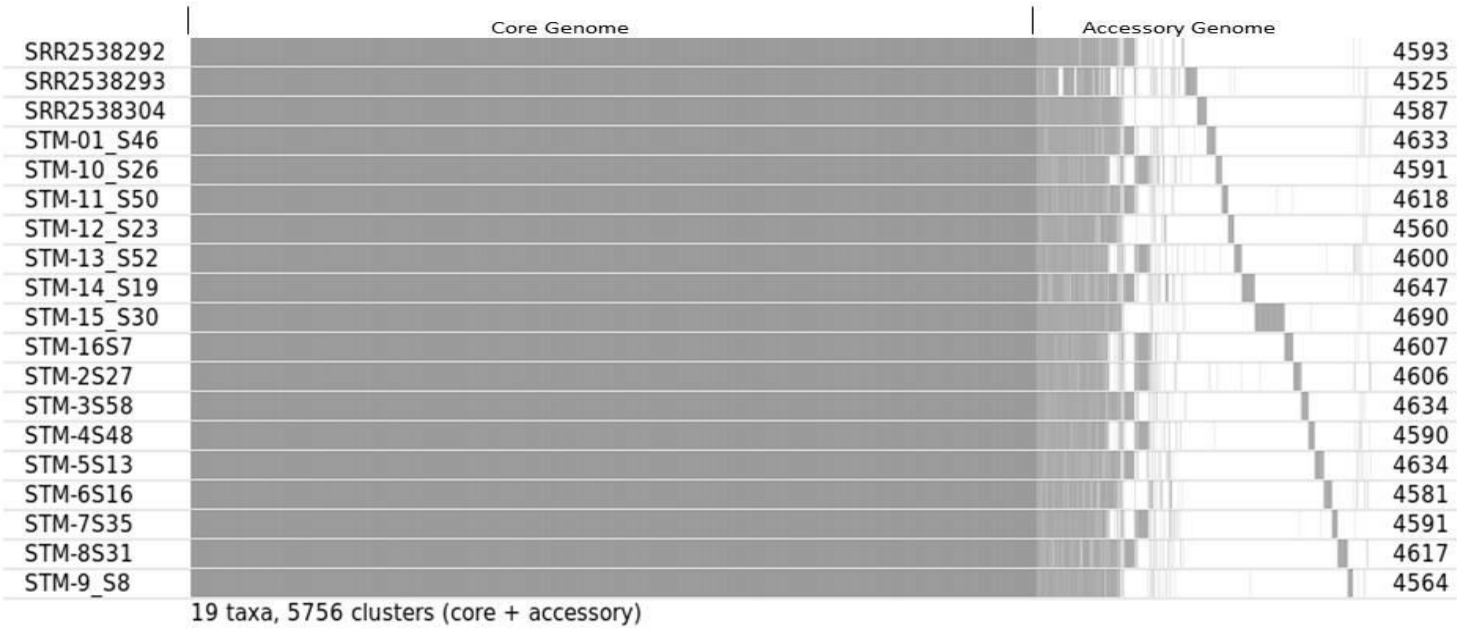


Figure 7.1 ROARY heat map displaying core and accessory genomes for 19 STM clades. Grey area displays detectable genes. Homogenous grey areas when comparing isolates indicate similarity of genomic composition. White colour corresponds to absence of particular genes. Isolates name SRR correspond to 'non-successful' ones. The remaining, STM1 to STM16 isolates, correspond to 'successful' isolates (see Table 7.1).

We hypothesized that epidemic potential can be associated and explained by the gene loss and/or acquisition within genomes of the unsuccessful sporadic clades and successful epidemic clades. Our data suggests that no genes were commonly absent in all three sporadic clades when compared to epidemic clades. Figure 7.2 illustrates genomic variation within each of the 19 analysed clades matched against a set of particular genes. Majority of successful clades retained their genes throughout the years studied. MLVA profile 3-10-13-12-496 strains (yellow) were particularly stable in time. This was observed for genes *hin* and *steC* among others. Genes such as *dinI* remained present within clades from the 3-10-13-12-496 complex. While comparing the genomic structure of the 3-selected successful MLVA profiles, there were a number of genes in common. For instance, profile 3-9-7-12-523 and 3-10-13-12-496 share 37 common genes hence profile 3-17-9-11-523 and 3-9-7-12-523 share 12 genes. For MLVA profiles 3-9-7-12-523 and 3-10-13-12-496 in terms of common genes presented, *yfgF* related to the bacteria motility, *xerD* related to chromosome stability and *ftsK* gene, essential for cell division are the relevant to mention (Figure 7.2).

Parallel comparison between non-successful profiles and each of the successful ones showed at least 190 genes in common between profiles 5-10-12-9-490 and 3-10-13-12-490. Additional comparisons were performed involving successful and non-successful profiles with limited number of tandem repeat differences; in both these cases the number of shared genes increased in at least 100 genes. Presence of *dinI* gene, associated with shutting down the response that facilitates bacterial replication was the most meaningful finding (Figure 7.3).

	STM STRAINS																				
Genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	Annotation	Gene function
glfT																				Galactoruranosyltransferase GlfT2	Cell wall biogenesis
gspA																				General stress protein A	Cellular stress response
lpfD																				Ong fimbrial protein LpfD	Specific adhesion system
yfgF																				Cyclic di-GMP phosphodiesterase yfgF	Bacterial motility (swimming)
Trg																				Methyl-accepting chemotaxis protein III	Regulatory gene
dinI																				DNA-damage-inducible protein I	Shuts off SOS response when overexpressed
rrrD																				Lysozyme RrrD	Lysozyme activity
phnT																				Putative 2-aminoethylphosphonate import ATP-binding protein PhnT	Energy coupling to the transport system
ftsK																				DNA translocase FtsK	Essential cell division protein
ybdO																				Putative HTH-type transcriptional regulator YbdO	Involved with virulence-related phenotypes
oadB																				Oxaloacetate decarboxylase beta chain	Lyase and sodium transporter
xerD																				Tyrosine recombinase XerD	Chromosome stability in bacteria
Hin																				DNA-invertase hin	Flagellar change variation promoter
gpfI																				Putative prophage major tail sheath	DNA packaging of the phage
aldB																				Aldehyde dehydrogenase B	Oxidation-reduction process
steC																				Secreted effector kinase SteC	Alters host cell physiology and promote bacterial survival in host tissue
dgaE																				D-glucosaminate-6-phosphate ammonia lyase	Catabolism of D-glucosaminate
clpP																				ATP-dependent Clp protease proteolytic subunit	Helps growth under stressful conditions
sopE																				Guanine nucleotide exchange factor SopE	Promoting entry into non-phagocytic cells
dnaC																				DNA replication protein DnaC	ATP binding
cfiA																				2-oxoglutarate carboxylase large subunit	Lipoyl binding

Figure 7.2. continued

	STM strains																				
Genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	**Annotation**	**Gene function**
yihV																				Sulfofructose kinase	Ribokinase activity
brnQ																				Branched-chain amino acid transport system 2 carrier protein	Structural gene
Dam																				DNA adenine methylase	Potential post replication repair function
dpp5																				Dipeptidyl-peptidase-5	Removes dipeptides from larger peptides
Pal																				Peptidoglycan-associated lipoprotein	Bacterial survival and pathogenesis
rspA																				Putative dehydratase	catalytic activity
rsxC																				Electron transport complex subunit rsxC	Electron transport at membrane level
garK																				Glycerate kinase	Organic acid phosphorilation
ftsK																				DNA translocate FtsK	Assembly of cell division
sspH																				E3 ubiquitin-protein ligase sspH2)	Bacterial survival and alters host cell physiology

Figure 7.2 Presence or absence of particular allele genes within the 19 analysed STM clades. Presence of gene = coloured box. Absence of gene = white box. Also, the 3 particular clonal complexes have been assigned similar coding colour (■ = STM complex 3-9-7-12-523; ■ = STM complex 3-17-9-11-523; ■ = STM complex 3-10-13-12-496; ■ = sporadic clade 5-10-12-9-490; ■ = sporadic clade 3-9-7-15-523; ■ = sporadic clade 3-16-11-11-523).

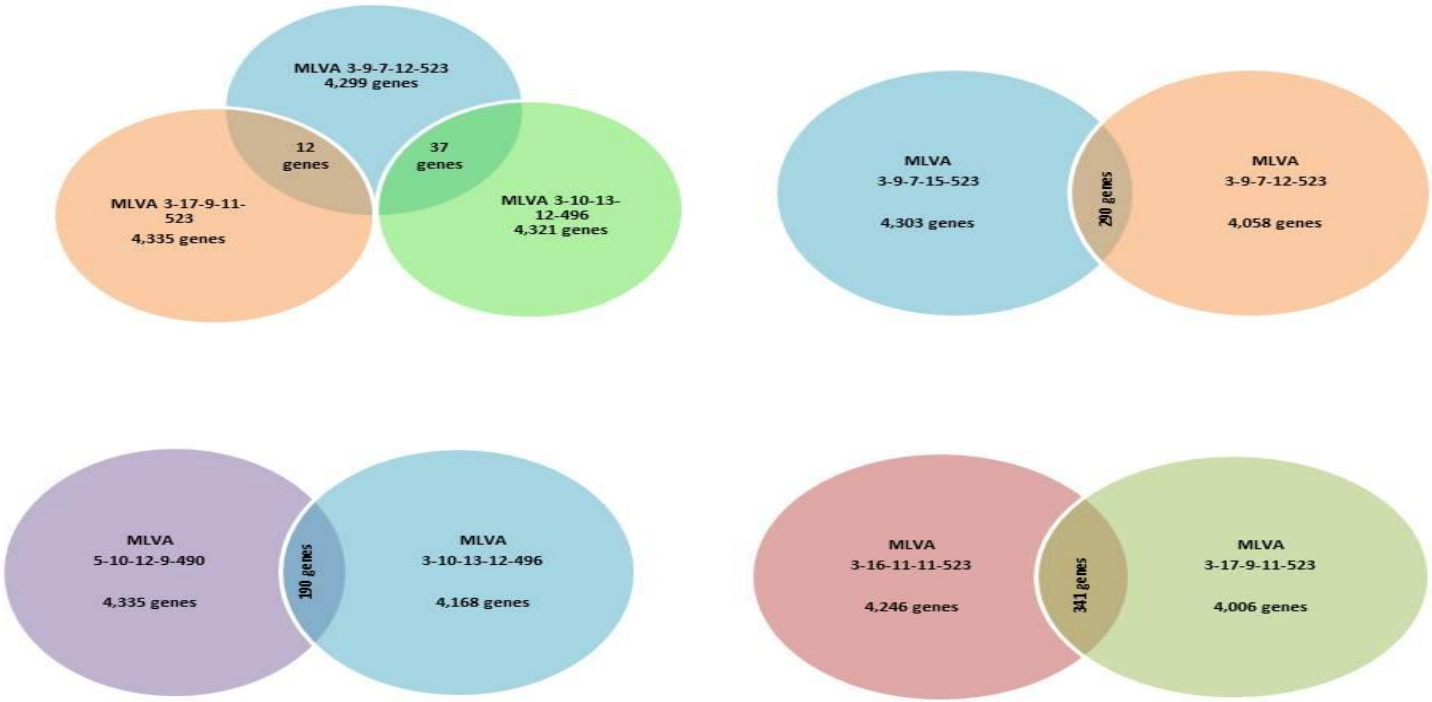


Figure 7.3 Venn diagrams summarizing numbers of genes and shared genes between three successful MLVA clades 3-9-7-12-523, 3-17-9-11-523 and 3-10-13-12-496; also number of genes shared between each of the successful clades compared to each of the 3 non-successful MLVA profiles. Successful clade 3-9-7-12-523 compared to non-successful 3-9-17-15-523, Successful clade 3-10-13-12-496 compared to non-successful 5-10-12-9-490 and successful clade 3-17-9-11-523 and non-successful 3-16-11-11-523.

Phylogenetic trees based on the accessory genome were used to cluster together the different successful and non-successful isolates (Figure 7.4). The tree clustered together isolates belonging to MLVA profile 3-10-13-12-496 (STM-2, STM-10, STM-16, STM-13, STM-4 and STM-7). Non-successful/sporadic clade 5-10-12-9-490 (SRR2538293), though it is not part of the mentioned cluster for profile 3-10-13-12-496, did share a common ancestor. For these same isolates, the ones recovered in 2014 and 2011 were grouped together; similar situation occurred for isolates 3-10-13-12-496from 2013 and 2015. 3-10-13-12-496 isolates from 2010 and 2012 were allocated within the same general cluster but in independent branches. As expected, isolates with MLVA profile 3-17-9-11-523 (STM-6, 12, 9 and 15) were clustered together and shared a common ancestor. Unsuccessful/sporadic clade 3-16-11-11-523 (SRR2538304), despite being close to a most recent ancestor of the successful clades, was positioned separately in the tree indicating significant differences between them in accessory genomes.

For STM MLVA profile 3-9-7-12-523, distribution within the tree was different. Two of the isolates (STM-11 and 8) collected in 2013 and 2014 were clustered together. Sporadic clade 3-9-7-15-523 (SRR2538292) was clustered independently, however with a common ancestor to the mentioned MLVA profile. Also, a common ancestor was shown between this sporadic clade and other ancestors for MLVA profile 3-17-9-11-523. The rest of the MLVA profile 3-9-7-12-523 clades (STM-5, 14, 3 and 1) collected in 2010, 2011, 2012 and 2015 were clustered separately. Figure 7.4 also showed how unsuccessful clades, i.e 3-16-11-11-523, clustered relative to the successful clades profile 3-17-9-11-523.

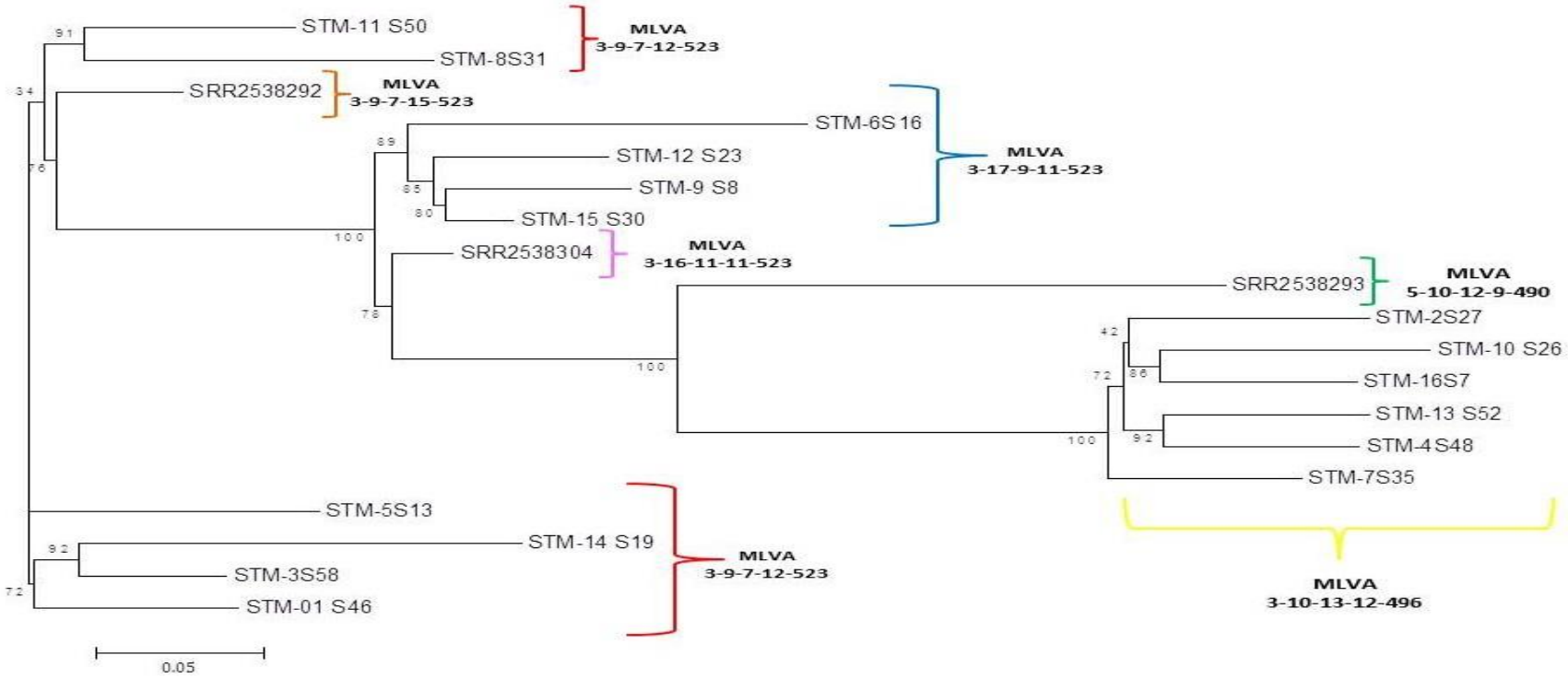


Figure 7.4 Maximum likelihood tree based on accessory genomes. The colours correspond to the identifying colours used for particular MLVA profiles in Figure 7.2. MLVA profile 3-9-7-12-523, 3-10-13-12-496 and 3-17-9-11-523 corresponded to the successful clades. MLVA profiles 5-10-12-9-490, 3-16-11-11-523 and 3-9-7-15-523 corresponded to the non-successful clades.

7.5 Genomic differences between successful and no-successful STM clades

When evaluating genomic differences within non-successful profile 3-9-7-15-523 when compared to successful ones, there is a single gene absent from its accessory genome. This one is present in all epidemic clades through the years and is gene *gspA*-2, a general stress protein A. It is usually present within the Enterobacteriaceae family and its function is associated with lipopolysaccharide 3-alpha-galactosyltransferase activity. Comparison of genomes from sporadic clade STM 3-9-7-15-523 with epidemic STM 3-9-7-12-523 identified three genes which are present in the epidemic clade but absent in the sporadic clade: *glfT*, *gspA* (previously described in general correlation against all successful clades in Figure 7.2) and *lpfD*. This last gene codes for the long polar fimbrial subunit lpfD located within DNA complement strand. It is involved in adhesion of *Salmonella* to either other cells or any extracellular matrix (231). The *lpfD*-2 gene is usually found in serovars Typhimurium and Enteritidis. Its loss could potentially lead to the reduction on STM ability to effectively colonize host's cells. Figure 7.5 displays the *lpfD*-2 location on the STM complement strand.

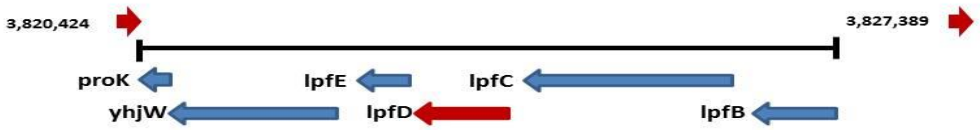


Figure 7.5 *lpfD* gene location on the complement strand of the STM genome (McClelland *et al*, 2001). The arrows towards left indicate this gene is located within the reverse strand. Red has been used to highlight its location.

Gene *glfT*-2 encodes the enzyme galactofuranose, which plays an important role in cell wall biogenesis involving the lipopolysaccharide (LPS) structure in gram-negative bacteria, and enabling the "entry to cell" associated activities. The absence of this gene within non-successful STM isolate's genome would translate into limited functionality within the non-successful STM clades to actively generate cell infection. Comparing MLVA profile 3-9-7-15-523, sporadic clade, with MLVA profile 3-17-9-11-523, classified as successful/epidemic based on case number presentation, the sporadic clade presents several missing genes when compared to the particular epidemic clade: *glfT*-2, *gspA*-2, *lpfD*-2, *yfgF*-3 and *trg*-2. They are located across the whole accessory STM genome, with no specific region involved. Two genes, *lpfD*-2 and *yfgF*-3 are important in STM adhesion to the host cell and the swimming behaviour, respectively (Figure 7.2).

Comparing non-successful clade 3-9-7-15-523 and successful STM-clade 3-10-13-12-496, regardless the year where the successful clade belongs to, results show there are 18 genes missing in the accessory genome of the non-successful clade; almost all these gene absences are different from the ones recognized when comparing non-successful clade 3-9-7-15-523 to STM-clade 3-17-9-11-523, with the exception of *gspA*-2. Five of the missing 18 genes contribute to the capability of a particular STM clade to effectively generate infection; these are *ptxA*, *ftsK*, *dinI*-1, *hin*-1 and *steC*-2. This last one, gene *steC*-2, is required for the formation of the *Salmonella* Pathogenicity Island (SPI)-2-dependent F-actin meshwork, responsible for the STM replication inside mammalian host cells (168) (Table 7.2).

Table 7.2 Genes lost in STM MLVA 3-9-7-15-523 in comparison with STM MLVA 3-10-13-12-496

Lost genes	**Annotation**	**Function/Special feature**
gspA	General stress protein	Bacterial stress response
dinI	DNA-damage-inducible protein I	plasmid: pSTM709
rrrD	Lysozyme RrrD	Lysozyme activity
phnT	Putative 2-aminoethylphosphonate import ATP-binding protein PhnT	Probably responsible for energy coupling to the transport system
ftsK	DNA translocase FtsK	Essential cell division protein
ybdO	putative HTH-type transcriptional regulator YbdO	Associated with virulence
oadB	Oxaloacetate decarboxylase beta chain	Lyase and sodium transporter
xerD	Tyrosine recombinase XerD	Stability of circular chromosomes in bacteria
dinI	DNA-damage-inducible protein I	Decreases replication of damaged DNA
rrrD	Lysozyme RrrD	DLP12 Prophage; Lysozyme activity
Hin	DNA-invertase hin	Flagellar change variation promoter
gpfI	Putative prophage major tail sheath protein	DNA packaging of the phage genome
aldB	Aldehyde dehydrogenase B	Oxidation-reduction process
steC	Secreted effector kinase SteC	Alters host cell physiology and promote bacterial survival in host tissues
dgaE	D-glucosaminate-6-phosphate ammonia lyase	Involved in the catabolism of D-glucosaminate
oadB	Oxaloacetate decarboxylase beta chain	Lyase and sodium transporter (sodium ion pump)

There were three genes missing within non-successful MLVA profile 5-10-12-9-490 accessory genome: *fsr*, s*spH*2-1and *gspA*-2. The last one was also absent in sporadic clade 3-9-7-15-523. Gene *fsr* corresponds to a trans membrane transport protein, which could have a role in virulence by activating and repressing multiple genes. Gene *sspH*-2 (E3 ubiquitin-protein ligase SspH2) codes an effector protein on SPI-2 that alters host cell physiology and to promote bacterial survival in host tissues. In terms of presence along the sporadic clade's genome, four genes can be identified as present within clade 5-10-12-9-490 only, *clcP, dnaC, yihV a*nd *sopE*.

All them appeared to be lost from epidemic clades, regardless individual isolate's year of collection. Special mention regarding relevant function for *sopE*, which has been described as part of a protein secretion system that translocate bacterial proteins into the host cell and for *clpP*, which presence would have a role in growth under stressful conditions.

Gene *dinI*, which main function has been associated to modulate the switching off of the SOS response needed in terms of facilitating bacterial replication, when present in a particular *Salmonella* organism was present in all non-successful clades including 5-10-12-9-490 and absent in both most successful MLVA profiles, 3-9-7-12-523 and 3-17-9-11-523. A list of this genes and a general summary of their name and function can be found in Table 7.3.

Table 7.3 Genes found in STM-clade 5-10-12-9-490, which are absent in epidemic STM clades

Gene	**Annotation**	**Function/Special feature**
Group 1091 (*dinI*)	DNA-damage-inducible protein I	Gifsy-1 prophage DinI
clpP	ATP-dependent Clp protease proteolytic subunit	Helps bacteria to grow under stressful conditions
gpfI	Putative prophage major tail sheath protein	DNA packaging of the phage genome
sopE	Guanine nucleotide exchange factor SopE	Promoting bacterial entry into non-phagocytic cells
dnaC	DNA replication protein DnaC	ATP binding
Group 176 (*cfiA*)	2-oxoglutarate carboxylase large subunit	Lipoyl binding
yihV	Sulfofructose kinase	Ribokinase activity

In terms of gene presence within non-successful clade 3-16-11-11-523 genome but absent in all successful profiles, there is a single relevant gene to refer to: *dinI*. The presence of *dinI* is common within all sporadic clades and it would reduce the STM clades facility to carry on with the required intracellular bacterial replication within pathogenesis (236). Additional genes that Roary software identifies as present within non-successful 3-16-11-11-523 clade but absent in all three successful profiles correspond to: *brnQ*, *xerD-*, *dam*, *dpp*5, *pal* and *rspA* (Table 7.4).

Table 7.4 Genes present in STM MLVA 3-16-11-11-523 but absent in any of the three successful STM clades 3-9-7-12-523, 3-17-9-11-523 and 3-10-13-12-496.

Gene	Annotation	Function/Special feature
dinI (group 1513)	DNA damage inducible protein I	Modulates SOS response
brnQ	Branched-chain amino acid transport system 2 carrier protein	Component of the LIV-II transport system for branched-chain amino acids. Structural gene
xerD	Tyrosine recombinase xerD	Catalysing the cutting and re-joining of the recombining DNA molecules during cell division
Dam	DNA adenine methylase	Post-replication mismatch repair function
*dpp*5	Dipeptidyl-peptidase-5	Removes dipeptides from the C-termini of N-blocked tripeptides, tetrapeptides and larger peptides
Pal	Peptidoglycan-associated lipoprotein	Essential for bacterial survival.
r*spA*	Putative dehydratase	Catalytic activity

7.6 Genomic variations within epidemic clades over time

Our findings showed that *sspH* gene, with a main role in altering host cell physiology and promotion of bacterial survival in host tissues, was initially present in the two isolates belonging to STM MLVA 3-9-7-12-523 from 2010 and 2011; it appeared lost and was not found in any of the four isolates recovered from patients diagnosed with STM disease since 2012. Correlating this information with the MLVA profile of associated human cases during each year (Chapter 6), 2012 was the period where the incidence of this clade was lowest (n=15) compared to 223 cases in 2009 and 89 identifications within STM cases in 2010.

The identification of this particular MLVA profile continued to decrease during 2012 and 2013, only to present a rise in recognition during 2014 (n=100) to again reduce its identification and association to STM cases in 2015 and 2016 (n=43 and 27 respectively). This increase on the STM cases in 2014 could not be related to any particular gene finding. For the same profile 3-9-7-12-523, there were ten genes that remained stable in accessory genome through the 6 years period: genes *glfT, gspA, lpfD, yfgF, phnT, oadB, xerD, cfiA, garK* and *ftsK*. Genes *yfgF* and *ftsK* have particular roles in motility and cell division respectively. For epidemic MLVA profile 3-10-13-12-496, genes remained constant for this particular clade's accessory genome between 2010 – 2015: genes *phnT, oadB* and *steC*. These four genes were absent in all 3 sporadic isolates. For epidemic clade 3-17-9-11-523, only 3 genes remained present in all isolates belonging to this profile between 2012 and 2015: *yfgF*, with a recognized role in suppressing motility defects, *trg* gene which has the ability to sense sugars, as part of its chemotaxis components involved directly into flagellum activity and *sspH*, with an active virulence role previously described for successful MLVA 3-9-7-12-523. The presence of these genes related to effective motility was related with the peak incidence of this clade in terms of STM cases in 2014 (n=200).

7.7 Comparison of present SNPs within the three successful MLVA profiles with the reference STM LT2

After performing a variant detection analysis with CLC genomics workbench ® for all the three successful clades, in compare to each other, the STM clade which presented the largest number of SNPs was clade 3-10-13-12-496, with an average of SNPs presentation of over 1,500 SNPs per year. Within these, approximately 70% of them, which corresponds to more than 1,000 SNPs, were related to synonymous SNPs with no evidence of amino acid change as a consequence

of the nucleotide mutation. This proportion remained constant through the 6 years of study. MLVA profile 3-9-7-12-523 was the successful clade, regardless the year, with the least synonymous SNPs (mean of 301). SNPs observed associated to MLVA profile 3-17-9-11-523 presented a similar mean of 302. In terms of trend of SNP presence for STM clade 3-9-7-12-523, there was a significant drop in numbers, from a total number of SNPs of 1,529 in 2010 to 716 in 2011; the proportion of synonymous SNPs for year 2010 was 19.3% whereas the following year, the proportion of synonymous SNPs increased to 42%. This drastic increase was only visible within STM clade 3-9-7-12-523; the other 2 remained constant through all the years, with between 30 – 40% of the total SNPs corresponding to synonymous (Table 7.5)

Table 7.5 Comparative analysis in terms of presence of SNPs within STM successful clades and reference strain. There is no data available for profile 3-17-9-11-523 during 2010 – 2011 since this clade was not identified as associated to STM infection until 2012.

	Number of SNPs (synonymous / non-synonymous)					
MLVA Profiles	2010	2011	2012	2013	2014	2015
3-9-7-12-523	295/421 (N=1,529)	302/414 (N=716)	297/443 (N=730)	302/434 (N=736)	305/431 (N=736)	305/429 (N=734)
3-17-9-11-523			302/428 (N=730)	303/417 (N=720)	305/416 (N=734)	302/399 (N=701)
3-10-13-12-496	459/1,070 (N=1,529)	459/1,052 (N=1,511)	458/1,056 (N=1,514)	467/1,095 (N=1,562)	457/1,049 (N=1,506)	470/1,097 (N=1,567)

7.8 Discussion

Our results improve the understanding of dynamic changes in clonal distribution of STM over time. All genomic differences within isolates were found in the accessory genome; this confers advantage to each particular isolate in a specific niche. Relevant differences involved presence of genes associated with pathogenicity and motility within the successful clades and absence in the non-successful. Some of these genes were also found strictly in the two more successful MLVA profiles when associating them with generation of STM cases. Comparisons of STM LT2, *S.* Enteritidis and *S.* Gallinarum have proven that genes conserved between serovars show approximately 99% identity at the nucleotide level (154). It is not surprising then that when we analysed a limited in number however diverse set of STM isolates according to MLVA profile, but all belonging to the same serovar, the similarities in terms of genome composition will be even greater.

There was a clear stability of core genome within our isolates, with some variability observed at the accessory genome level, particularly when comparing the three successful clades against the 3 non-successful ones; Suez *et al* (2013) mentioned the same conclusion (161).

Nevertheless, even between members of the same or closely related serovars, between 1 and 5% of the genes are strain-specific, corresponding mainly to large prophage elements and even other mobile genetic elements and plasmids (79). With this assumption in mind, although the *Salmonella* genome can be considered conservative when compared to other organisms such as *E.* coli, it is still highly plastic, which was observed through the analysis of the similar successful MLVA profile isolates through time. There was no large number of genes to support the plasticity concept based on deletions or insertions; in fact, a majority of the genes remained constant through the 4 to 6 years analysis of the 3 selected epidemic clades. Considering that the selection of these

isolates was random in terms of representing each year but targeted in terms of not being based on any criteria which could increase the chances of genomic adaptability taking place, but only focused on having similar MLVA typing result and not belonging to the same patient as published by Octavia *et al* (2015), it seemed that each isolate had not gone through individual adaptive changes, resulting on large amount of modifications affecting numerous genes through time (219). Particularly for the MLVA profile member of the most successful STM clonal complex in NSW, 3-9-7-12-523, there are genes whose presence correlated with highly epidemic years. This is the case of gene *sspH*, a phage remnant containing virulence genes commonly associated with type III effector proteins that are injected by the bacteria and which are important for virulence, as described by Brussow *et al*, (2004) (237). This gene was found in all of our isolates belonging to 2 out 3 STM successful MLVA profiles; these two corresponded to the profiles associated with the largest number of STM cases in NSW.

Its presence within the STM genome has been described by Bhavsar et at (2013) as one of the secretion systems on SPI-1 and SPI-2 capable of transporting effector proteins directly into the host's cells so they can interfere in particular cellular processes (165). These SPI-2 effectors are critical for *S.* Typhimurium pathogenesis, particularly for systemic infection (Kuhle and Hensen, 2004) (238). Pang *et al* (2013) also found this particular gene within their 6 different analysed strains, confirming not only its association to SPI-2 secretions but also its role in terms of virulence (239). *sspH* gene remained within the genome of all epidemic clade isolates analysed in this study between 2010 and 2011. This period of time is where STM cases associated with this particular MLVA profile presented higher incidence. Once the gene disappeared (from 2012 onwards), the number of cases associated with this particular profile decreased substantially, according to NSW ERL *Salmonella* dataset.

Gene *steC*, was also found in one of the successful MLVA profiles (3-10-13-12-496); this gene has been described by Poh *et al* (2007) and Imami *et al* (2013) as responsible for the active replication of STM in the host cells as a secreted effector bySPI-2 (240, 165). Thought this gene was not a generalized finding within all three successful MLVA profiles, MLVA profile 3-10-13-12-496 was the one with presence of STM cases which has presented more stability in time when compared with the other two successful profiles 3-9-7-12-523 and 3-17-9-11-523; these last two have been very successful at their own period of time, as shown in chapters 5 and 6, however their association with STM cases has presented fluctuations.

MLVA successful clade 3-10-13-12-496 has been not been associated with STM cases in as elevated numbers but its presence has remained more constant in time, when compared to the other two profiles. The disappearance during years 2012 - 2013 and later appearance of genes such as *rsxC* and *garK w*ithin the accessory genome of successful clade 3-9-7-12-523 could be explained as an adaptive measure to respond and regulate growth in a more efficient way, based on host and/or environmental characteristics. Lamichhane-Khadka *et al* (2011) have described the presence of gene *rsxC*, among others, as a tool for metabolic flexibility, allowing STM to survive in diverse environmental conditions both outside and within the host, as a way of increasing the survival capability of the bacteria either within macrophages and/or animal hosts (241).

Amongst the STM 3-9-7-12-523 isolates analysed in this study, the *garK* gene was present until 2012, absent in 2013 and 2014 only to reappear in 2015. Perhaps the gene stimulation expression by means of a modification of environmental variables (temperature, humidity), generated the suppression of this particular gene during those 2 years, where according to NSW Health Bulletins relate to lower presentation rate for this particular profile.

The presence of gene *yfgF* only within the three successful STM isolates can be explained by the description of the gene's function as stated by Girgis *et al* (2007), where this gene's contribution to the bacteria's capability of generating infection by means of been involved in motility-related infective functions is relevant (242). Similarities in terms of its association to the flagellum functions are found regarding gene *trg*. This gene was only present within the successful isolates belonging to MLVA profile 3-17-9-11-523, which happens to be an important member of the STM clonal complex that since 2012 has been leading the STM cases in NSW. Whenever there is a lack of *trg* presence, a measurable motility defect was present (242).

In terms of the non-successful-sporadic clades, the presence of gene *dinI* within their genome was a relevant finding. Our results found it within the genome of all non-successful-sporadic clades and also within successful MLVA profile 3-10-13-12-496. It was not present in isolates with successful profiles 3-9-7-12-523 or 317-9-11-523. Erickson *et al* (2003) and Yasuda *et al* (2001) have described the presence of the *dinI* gene associated with switching off the SOS response mechanism necessary for a successful bacterial DNA replication and repair within host cells (243, 236). This statement is easily translated to our results; the 2 MLVA profile, represented by a set of 10 isolates that didn't present it, happen to be the 2 more successful clades within the NSW STM cases spectrum. Clearly, for the successful MLVA profiles who did not have *dinI,* if this gene is absent, there will be no interference in terms of the battery of mechanisms involved in effectively invading the host cells, which will translate into more effective infection rates, and therefore higher number identifying this particular MLVA profiles which lack of the gene associated to cases. Regarding the presence of SNPs within the successful-epidemic clades, our results showed total number of SNPs of around 1,500 per year for successful MLVA profile 3-10-13-12-496 and around 700 SNPs for successful profiles 3-9-7-12-523 and 3-17-9-11-523. Phillips

et al (2015) referred to a SNP analysis within outbreak associated isolates: a single non-synonymous SNP was found within all related isolates, reaffirming the fact that increasing number of SNPs is associated with genetically distant isolates (45). It would be expected that in case of outbreak, the SNPs number should be reduced. Since our isolates don't belong to a particular outbreak, they are simple representatives from each year STM dynamics, it was expected to find large number of SNPs. Other studies evaluating STM in Passeriformes birds, humans and domestic animals in England described a median pairwise SNP difference of 130 (range of 18 – 406), considering the isolates genetically closely related (150).

The cut-off point in terms of up to how many SNP differences should be considered to state genetic relationships seems to vary from author to author. Even though there are large differences in terms the number of SNPs described prior to our research, there is one idea that is clear: WGS currently provides the highest resolution available to investigate the relatedness and gene content of any microorganism. Makendi *et al* (2016) described SNPs to over 2,600 for *S.* Weltevreden (57). Leekitcharoenphon *et al* (2016) reported a total of 4,619 SNPs for 315 STM DT104 isolates (66). This increase in the detected SNPs could be related to the isolates belonging to different countries worldwide.

Leekitcharoenphon *et al* (2016), when analysing isolates belonging to a particular MDR cluster with isolates belonging to STM DT104 isolates collected from different Danish farms, between 1997 – 2011, identified a total of 755 SNPs (66). Interestingly, those farms where direct relationship was proven, presented SNPs differences of less than 30, stating the fact that, as suspected, the number of mutations within close isolates should be expected to be small. For our isolates presenting the lowest number of SNPs (mean 301), belonging to MLVA profile 3-9-7-12-523 and 3-17-9-11-523, it would be then suggested that this high number of SNPs found within a

similar MLVA profile is related to the fact that this reduced set of isolates (n=6 for MLVA profile 3-9-7-12-523, n=4 for profile 3-17-9-11-523) have no relation with each other in terms of geographical location, time of collection or patient where samples came from. The probability that these sets of selected successful clades through time could have any genomic relationship is slight, which is translated into the large number of SNP difference found through the time follow up for each of the MLVA profiles. In terms of the major proportion of SNPs being classified as synonymous within our isolates, similar results were described by S. Fu *et al* (2016), where almost 70% of SNPs related to STM were synonymous and mostly related to metabolism functions (216). Genome of STM has been described as stable, therefore the presence of these SNPs was expected.

In conclusion, though we analysed a limited number of isolates belonging to successful MLVA profiles, associated with STM cases in a 6 years period, these isolates presented large genomic similarities, reflecting the general genomic clonality present within STM isolates. The particular genomic variations detected in epidemic clades through time were limited, however always related with the increase of the MLVA profile's ability to infect the host cells, which could be associated with the number of STM cases the period of time after the variation took place.

The limitation of our study due to a relatively small number of STM isolates should be acknowledged. This number reflects limited resources for our project. There are many S. Typhimurium draft genomes available in public repositories such as the NCBI Genbank and SRA, however, the main goal of the study was to examine STM epidemics in the NSW using a representative dataset, without any risk of inadequate metadata. In order to perform WGS we had to restrict the number of isolates to sequence, however the sample was selected including isolates randomly in terms of their representation for the particular period of time they belonged to, but

targeted considering that we selected the 2 MLVA successful profiles that really represent the epidemics of STM in NSW.

Chapter 8 intended to relate these genomic findings and translate them into an experimental animal model of infection in order to evaluate whether genomic variations taking place in STM genome can be associated with more/less effective infections in mice.

Chapter 8: Within and between host genomic variations of an acute and chronic STM animal model

8.1 Introduction

The aim of this study was to explore the variability of STM genomes over the natural course of infection. In order to minimize confounding effects that complicate human infection, we used established mouse models of experimental salmonellosis. We conducted two separate experiments using different mice strains in order to identify within- and between-host variations, and adaptation of STM genomes in acute and chronic STM infection (objective number 5). The results of this study revealed new insights and enabled better understanding of genomic variations will assist the interpretation of STM sequencing results in the context of different transmission pathways. This should be relevant for public health surveillance and interpretation of STM genomic variations in human cases and, potentially, the variation and adaptability occurring in chronic STM carriers.

8.2 The transmission pathway

Within-host and between host genome variations of STM were originally examined by infecting 6-week-old female BALB/c mice with an epidemic clade MLVA profile 3-9-7-12-523 isolate (Chapter 4; subchapter 4.7). The initial isolate was obtained from a human infection. Sets of five consecutive passages of STM from infected animal to a STM-naïve animal were performed in duplicate, using inoculum prepared from stool culture and intra-gastric inoculation. Symptomatic infection in mice was confirmed by testing weekly stool cultures. Animals were considered recovered when they stopped displaying any symptoms or sign of infection and had negative stool cultures for STM. The STM isolate from the recovered mouse was used as inoculum (i.e. passaged) to a healthy, STM-naïve mouse in order to reproduce another infection with the

same STM. This approach assured that the main variable in the experiment was the STM strain going through sequential animal hosts. The summary of the full transmission pathway network used for this described SNPs detection can be seen in Figure 8.1.

8.3 The experimental model of acute Salmonellosis

As described in Chapter 4, we employed an experimental model of acute salmonellosis to investigate emergence of polymorphisms in STM genomes. BALB/c mice with STM infection displayed systemic symptoms of infection such as ataxia, weight loss and hunched position. The addition of all these resulted on displaying severe systemic condition that led to the euthanasia of every compromised animal. Post mortem findings at tissue collection stage showed macroscopic signs of the infection. An evident splenomegalia was found in every one of the mice with STM-positive stools (Figure 8.2). Liver, intestine or mesenteric lymph nodes did not display macroscopic abnormalities that could have been potentially associated with the disease. The duration of culture-confirmed STM excretion in stools in infected animals remained stable at 3 weeks during passages 1, 2 and 3; passage 4 displayed 2 weeks of culture-confirmed STM excretion in stool of its respective group of mice.

Since this was a model of acute salmonellosis, minimal within-host STM genome variation was detected. However, non-synonymous (NS) as well as synonymous (S) single nucleotide polymorphisms (SNPs) were identified. Interestingly, the first passages were associated with synonymous SNPs but passages 2 and 3 led to non-synonymous mutations (Table 8.1).

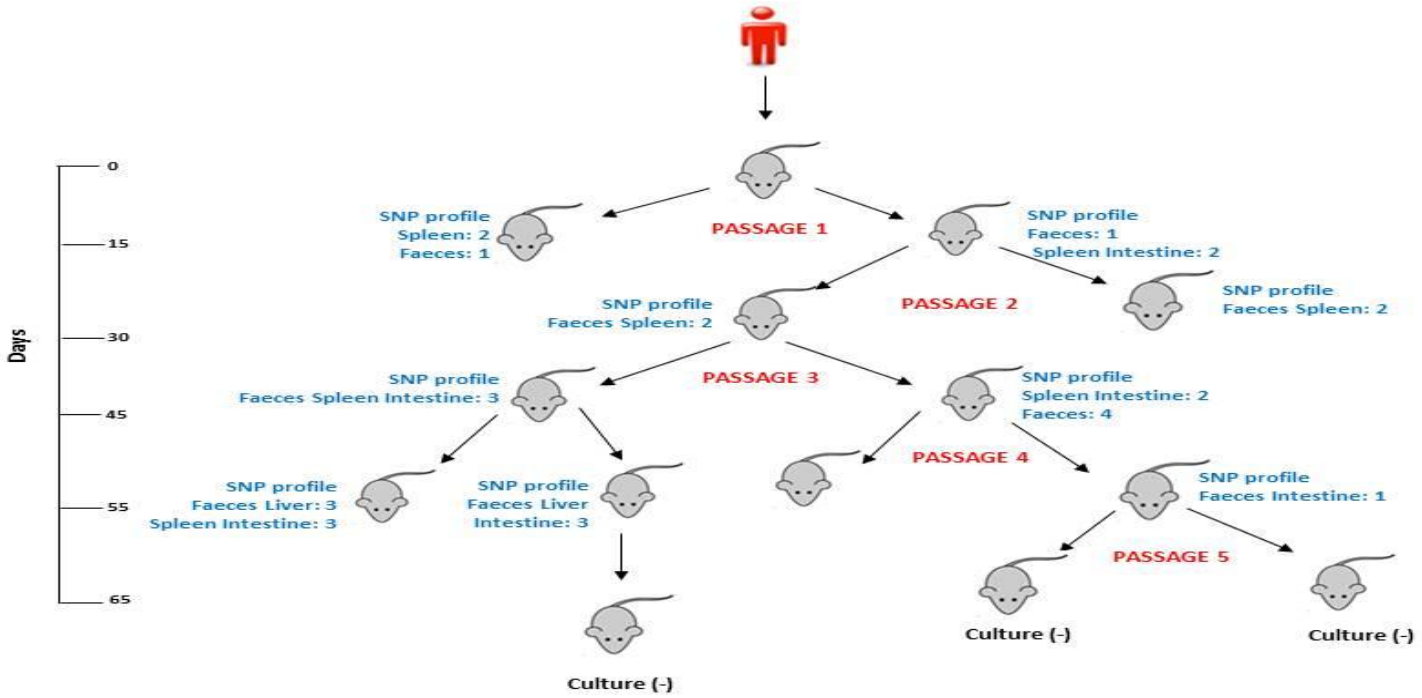


Figure 8.1 Complete transmission pathways for SNP detection used for the acute STM infection animal experiment. The SNP profiles are presented as associated to particular animal and the respective passage (colour blue). SNPs profile codes correspond to: 1= *rsxC, oadB, ftsK* genes; 2 = *rsxC, oadB, ftsK, mukB* genes; 3= *rsxC, oadB, ftsK, mukB, mobA* genes; 4= *rsxC, oadB, ftsK, mukB, yjhP* genes.

Table 8.1 Characteristics of SNPs observed in STM acute infection

Genes	**Annotation**	**Synonymous (S)/ non-synonymous (NS)**	**Amino acid change**	**Function**	**Passage where the SNP recorded the first time**
oadB	N-oxaloacetate decarboxylase subunit alpha	S	No change	Lyase + Sodium transporter	1
rsxC	Electron transport complex protein rsxC	S	No change	Involved in electron transport	1
ftsK	DNA translocase ftsK	S	No change	Essential cell division protein	1
mukB	Chromosome partition protein mukB	NS	Ala→Thr	Central role during cell division	2
mobA	Molybdenum cofactor guanil transferase	NS	Val→Ile	Growth – nutrient acquisition	3
yjhP	SAM-dependent methyl-transferase	S	-	Unknown	4

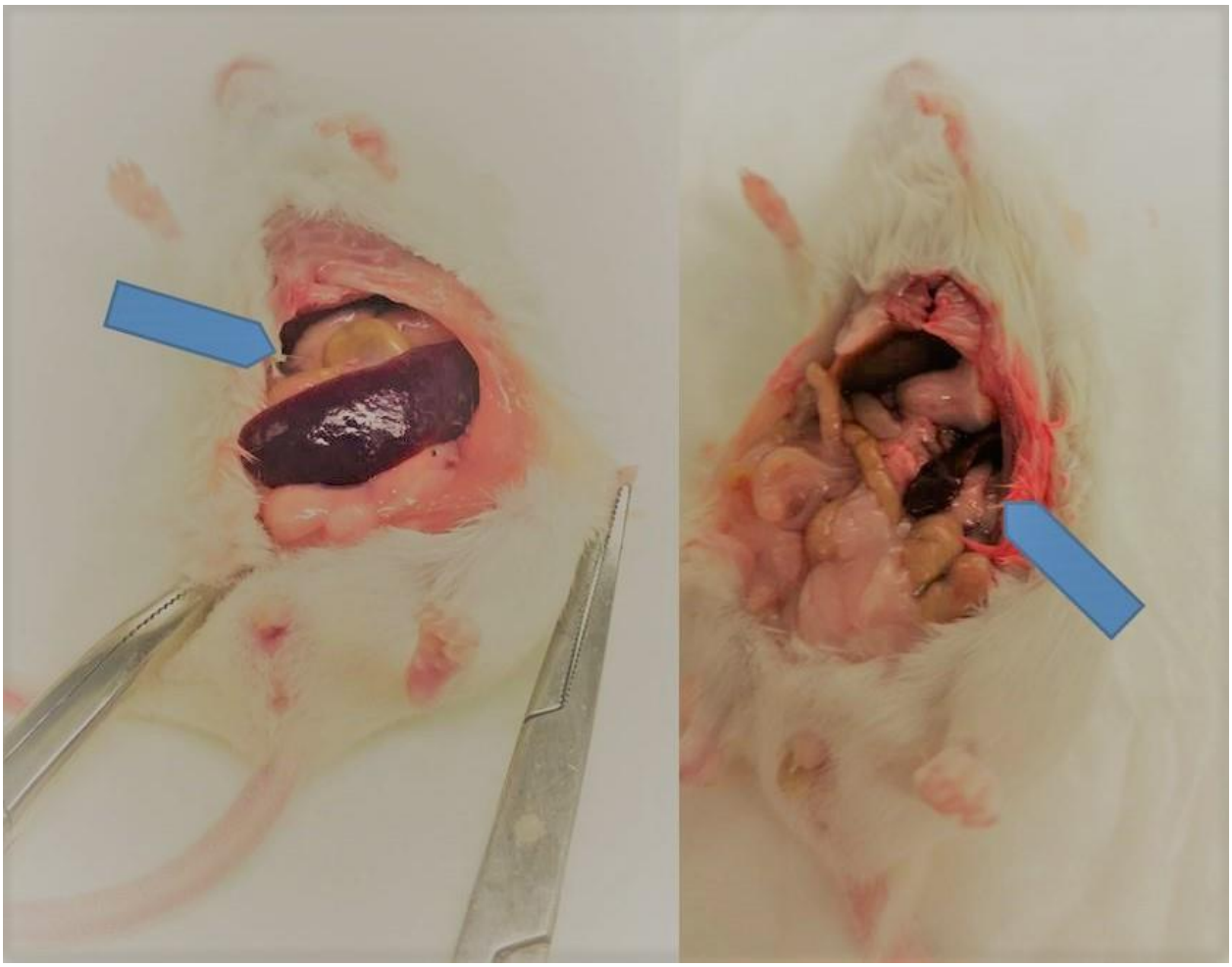


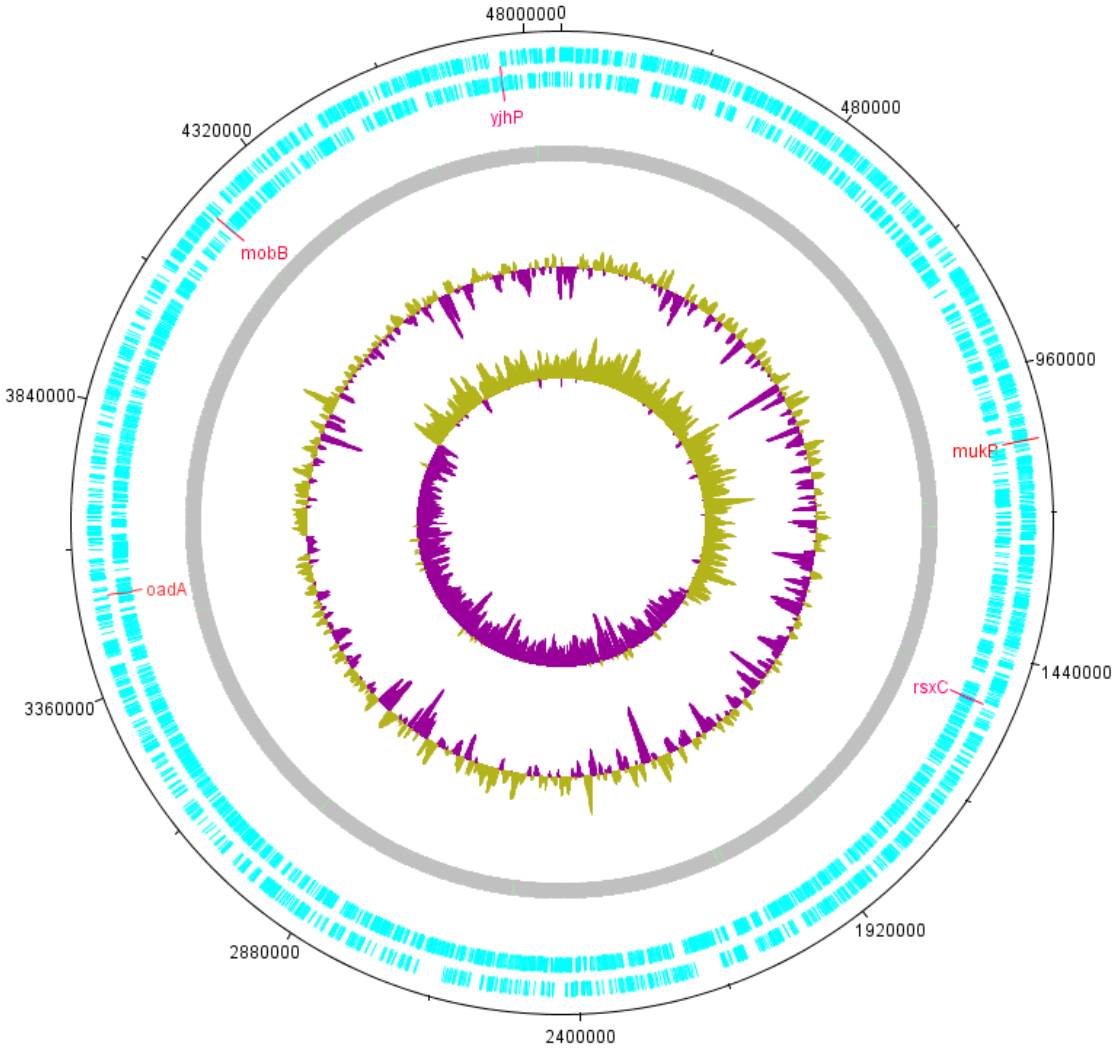
Figure 8.2 Spleen enlargement (splenomegalia) in a BALB/c mouse following acute infection with STM. An image of the spleen (red/purple structure) in a control healthy mouse is on the right. Blue arrows indicate position of the spleen.

After inoculation of the initial pair of mice (passage 1) with human STM strain, 4 different single SNPs were identified in both animals. The SNPs were located in four genes: *genes rsxC* (electron transport complex subunit RsxC), *oadB* (oxaloacetate decarboxylase beta chain), *ftsK* (DNA translocase FtsK) and *mukB* (chromosome partition protein MukB) (Figure 8.1). Differences were detected in terms of the type of sample where the SNPs were found. In one passage 1 mouse, SNPs were present in faecal samples and spleen tissue, approximately 10 days after inoculation took place. The second passage 1 mouse presented the exact same SNP profile in faecal samples, spleen and intestinal tissue. After performing 2nd inoculation of a second set of animals (passage 2), the same SNP profile present in passage 1 animals was identified in both passage 2 animals, remaining only associated to the same genes, *rsxC, oadB, ftsK* and *mukB* (Figure 8.1). In addition, the SNPs were not detected in intestinal tissue, only in faecal and spleen tissue samples.

Following the 3rd passage in a different pair of mice (passage 3), a different SNP profile emerged in one passage 3 mouse, which involved SNPs in genes *rsxC, oadB, ftsK, mukB* and *mobA* gene (molybdenum cofactor guanylyltransferase: Figure 8.1). A novel SNP profile was also detected in the second passage 3 mouse inoculated in parallel, consisting of SNPs in genes *rsxC, oadB, ftsK, mukB* and the *yjhP* gene. For the animals inoculated with stools with genes *rsxC, oadB, ftsK, mukB* and *mobA* (passage 3), there was no evidence of change in SNP profile. The STM culture from liver tissues obtained from autopsy of both mice at the end of experiment presented the mentioned SNP profile. Passage 4 was carried out by inoculation a set of 2 additional animals. These two mice were inoculated at passage 4 by using stools samples with SNP profile including SNPs in genes *rsxC, oadB, ftsK, mukB a*nd *yjhP*, previously described in passage 3.

One passage 4 mouse failed to develop any signs of the infection and her stool cultures were negative for STM, even after serial stool samples were collected. The second mouse did develop the disease, however, the STM isolate subsequently cultured from her was distinct from that in the inoculum. Whole genome sequencing revealed that the isolate recovered from the symptomatic passage 4 mouse had lost the SNP in the *mukB* gene from passage 2 and also gene *yjhP* from passage 3. The final SNP profile for the symptomatic passage 4 animal was associated with genes *rsxC, oadB* and *ftsK*, similar to the presence of SNPs that was detected within passages 1 and 2 (Figure 8.1). A final 5th passage was carried out for one mouse using an isolate with SNPs present in the *rsxC, oadB, ftsK, mukB* and *mob* genes from passage 3, and a second passage 5 mouse using an isolate with SNPs present in the *rsxC, oadB* and *ftsK* genes. For the three remaining inoculated mice (passage 5), the culture results were negative for STM. Series of stool samples were collected after the additional week, however, none of the collected stools grew STM on selective culture media. Genes *rsxC, oadB a*nd *ftsK* presented different SNPs that became fixed and carried through all four passages. In terms of amino acid change as a result of the mutation, 2 SNPs were identified as non-synonymous: *mobA a*nd *mukB* (Table 8.1). The observed polymorphisms were evenly distributed across the genome, with no apparent hot spots or associations (Figure 8.3).

Figure 8.3 The location of genome polymorphisms in comparison to the reference STM LT2 genome (Artemis DNA Plotter (Sanger Institute)). Light blue colour represents two reverse and forward DNA sequence strands of STM LT2 genome with genes containing SNPs (identified in acute animal experiment) with their names in red. Grey circle corresponds to the condensed STM LT2 reference genome. Purple inner circles represent G-C content through the genome.



8.4 The experimental model of chronic Salmonellosis

In order to examine the variations and adaptations within genomes of successful STM clades simulating a chronic STM infection, a separate animal experiment using mouse strain 129X1/SvJ was conducted (Chapter 4; subchapter 4.7). The initial inoculation of two animals was performed using a successful STM strain of human origin (MLVA 3-9-7-12-523), similar to the one used for acute infection (Passage 1, Week 0). Stool samples from infected animals were initially collected on a daily basis, but collection decreased to twice a week following XX number of weeks/days. During passage 1, STM infection was confirmed by positive STM cultures of stools collected 2 weeks post inoculation from a single mouse, which remained positive until week 17 when mice were culled. Both passage 1 mice had positive stool cultures by week 5 post-inoculation (Figure 8.4). positive STM tissue cultures from spleen, liver and intestines collected at week 17. Multiple tissue samples were positive by week 17.

Passage 2 was carried out at week 4 after beginning the experiment. The incubation period for passage 2 mice was longer than that of passage 1 mice. Passage 2 animals continued excreting STM in stools for 8 weeks from the initial inoculation. Tissue collection took place during week 16; STM positive cultures were obtained from the spleen and liver of both passage 2 animals. Differences in time to detection of positive cultures were observed during Passage 3, which was performed during week 11 of the experiment (Figure 8.4). Positive stool samples were obtained one-week post inoculation in both animals (week 12 from initial inoculum). Mice remained stool culture positive until week 9 post Passage 3 inoculation (week 20 from initial inoculum), at which time tissues were also collected. Systemic infection was confirmed by sampling of spleen and intestine tissue that grew STM on selective media.

Passage 4 was performed during week 17 of the experiment, and STM positive stool samples were also detected 1-week post inoculation in both animals. Stools remained positive until week 23 of the experiment (week 6 of passage 4), but only in 1 of the 2 passages 4 mice. The second animal excreted STM positive stool for only 4 weeks after inoculation (Figure 8.4). Tissue collection occurred during week 6 of passage 4 (week 23 from initial inoculation/passage 1), with spleen tissue STM positive for both animals and liver tissue positive for one animal (Figure 8.4). No positive cultures were obtained from the intestine tissue cultures. Regardless the passage, animal culling took place due to detected weight loss affecting the STM positive animals. No other particularly evident clinical sign was observed.

Passage number 5 was performed using the STM isolate recovered from one of the STM positive passage 4 animals. There were no cultures showing confirmed infection in any of the inoculated animals within passage 5. Re-inoculation using a fresh preparation of the inoculum was performed. After more than 10 days, there were no positive stool samples collected to reflect the infectious status of the animals. Mention that for this reason, passage 5 data is not included in Figure 8.4.

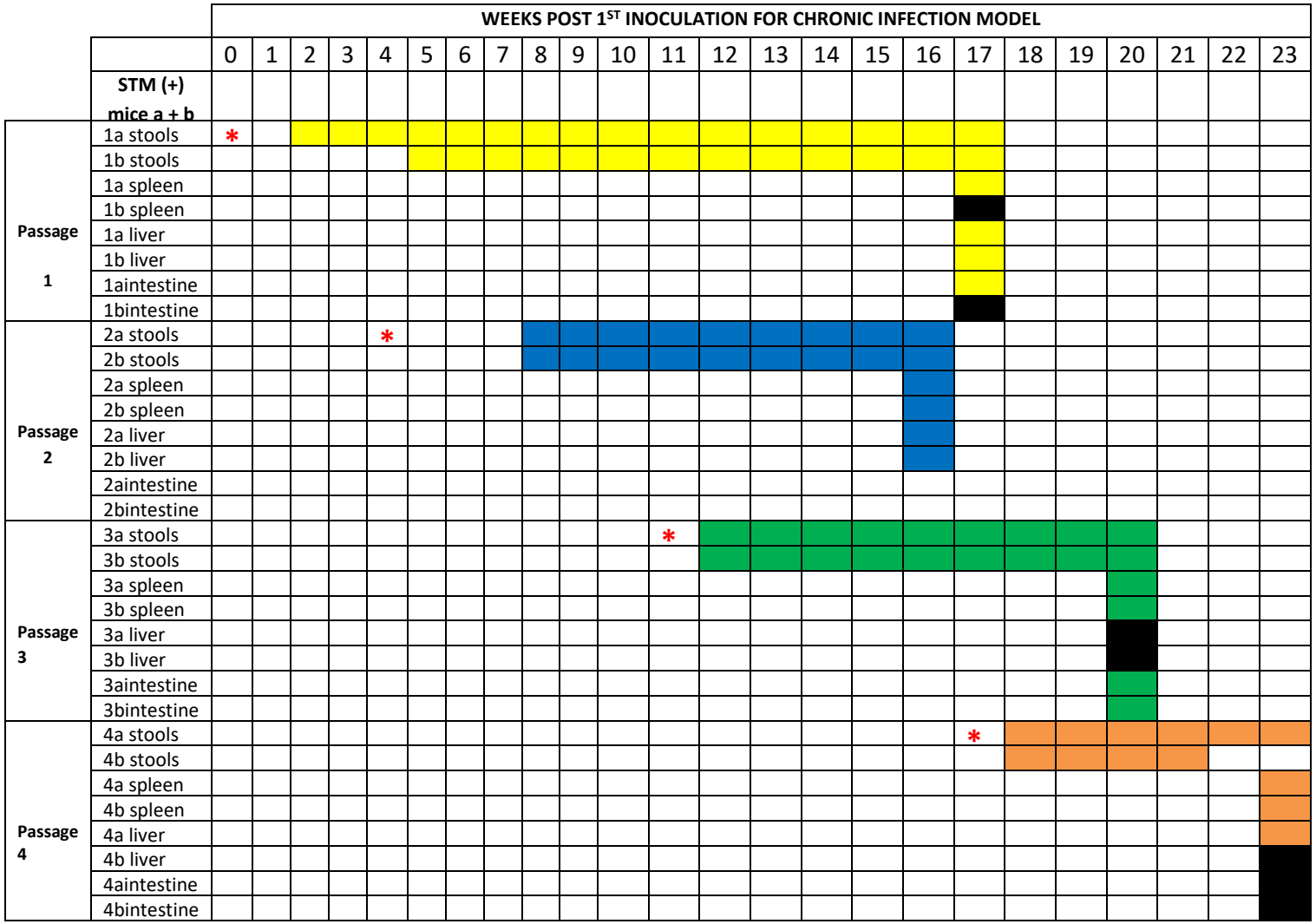


Figure 8.4 STM culture positivity from samples obtained from mice with chronic salmonellosis. Each colour corresponds to particular week and positive cultures detected. Black cells = negative culture. * = week where inoculation for that passage occurred. Inoculated mice have been labelled passage number and "a" and "b".

When comparing the duration of STM positivity in weeks, the acute and chronic models presented differences. Due to the acute presentation of clinical signs for STM in the balb/c acute mice, which displayed progressive deterioration (severe weight loss and general poor condition) resulting in euthanasia, positive cultures were obtained no longer than 4 weeks post inoculation. Initial passages, one and two, presented a faster appearance of systemic compromise when compared with passages three and four. Chronic infection model presented a larger number of weeks where STM positive cultures were obtained, confirming the shedding of STM for longer period of time as displayed in Figure 8.6. For instance, animals related to passage one, where displaying positivity for STM in stool cultures for 14 weeks average. Following passages reduced the time frame where positive cultures were obtained at least doubling the weeks recorded for the acute infection model. Positivity of culture during all passages can be reviewed in Figure 8.5(a).

Regarding weight (in grams), acute model mice displayed a progressive decrease during the 9 weeks of the experiment. The mean body weight for mice pre-inoculation was around 17 grams (Figure 8.5). After onset of infection, the weight loss was evident, with over 10 percent weight loss recorded for majority of mice. In chronic infection model mice, the mean body weight was around 20 gms; though there were weight gain and loss during time, these were minimal These findings were related to the general clinical status of the mice; none of them experienced evident clinical signs of infection, in contrast with the acute and evident signs displayed by the balb/c mice in the acute model. Based on the standard deviation (SD) error bars, there were no significant differences between weight gain and loss during each model of infection (Figure 8.5).

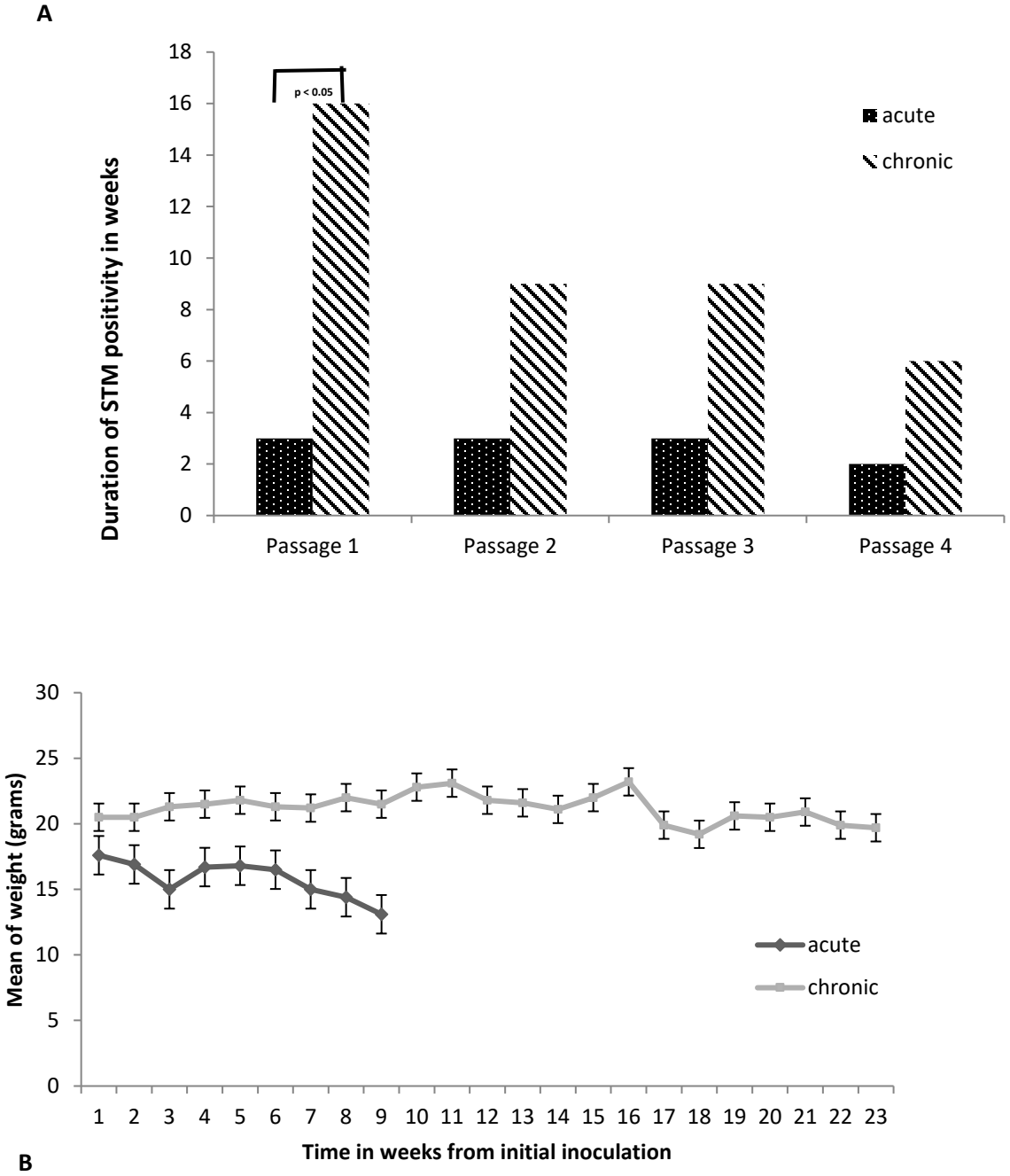


Figure 8.5 A: Duration of STM cultures positivity for the acute and chronic animal models of infection through consecutive passages. Passage 1 showed statistical significance between the number of positive STM cultures obtained in acute and chronic models. **B:** Mean of weight from mice included in both animal models.

The allelic differences between STM genomes of the isolate used in Passage 1 for the chronic model and the isolates recovered from infected animals during subsequent passages were recorded. The first polymorphisms were observed at week 4 of the first passage. The differences found involved 2 particular genes and these were single synonymous SNPs involving genes *ftsk* (DNA translocase ftsK) and *ssb* (Single-stranded DNA-binding protein 1), which play essential roles in cell division and DNA replication respectively (see Chapter 7 and Table 8.2 for details regarding the genes). Both findings were only present in one of the 2 inoculated animals. Both of them were synonymous SNPs with no amino acid changes induced. An interesting finding was that gene *ftsK* and the later to describe *oad* genes, were also associated with SNPs during the acute STM animal experiment. Further SNP was detected on stools from both animals, 2 weeks after detection of the previous one (week 6); this one corresponded to gene *rrsH* that is described as a 16S ribosomal RNA gene.

There was a persistent presence of the SNP related to *ftsk* gene across isolates involved in the 4 infective passages (fixed SNP); this was found not only in stool isolates but also among tissue, particularly in liver and spleen. The presence of differences within gene *oadA* (oxaloacetate decarboxylase) accompanied by gene *ssb* (single-stranded DNA-binding protein 1), was detected through carriage originated from passage 2, both of them detected in stool samples from the 2 inoculated animals. This last gene plays an important role in DNA replication, recombination and repair. After passage number 3 was performed, there was constant presence of some of the already mentioned SNPs such as *ftsK*. Particular findings involved only one of the 2 inoculated animals at this stage, where genes *yfdH* and *phrB* were involved, approximately around week 6 post passage 3 inoculation. Differences related to gene *dcoA* were also detected exclusively in one of the inoculated animals, which also didn't reflect the 2 previously mentioned genes. Passage 4 was

performed based on stools culture from the previously infected generation of mice. The analysis of this point in the transmission pathway showed carriage of 2 previously described genes, *ftsK* and *oadA*, both of them involving both of the inoculated mice. Analysis of the STM positive liver tissue from one of the mice inoculated during passage 4, reflected absence of gene *dcoB* (oxalacetate decarboxylase: beta chain) and presence of *cysN* gene (sulfate adenylyltransferase subunit 1).

In terms of amino acid changes, there were none present in the STM chronic infection model. None of the detected SNPs was non-synonymous. Figure 8.6 summarizes the transmission pathway utilized for this section of the project as well as presenting the SNP profile detected through passages. Table 8.2 summarizes the SNPs in STM chronic model.

Table 8.2 Characteristics of the SNPs (all synonymous) observed in STM chronic infection

Genes	**Annotation**	**Function**	**Passage where the SNP recorded the first time**
ftsK	DNA translocase ftsK	Essential cell division protein	1
Ssb	Single-stranded DNA binding protein 1	Role in DNA replication, recombination and repair	1
rrsH	16S ribosomal RNA	Coding gene for RNA	1
oadA	N-oxaloacetate decarboxylase subunit alpha	Lyase + Sodium transporter	2
phrB	deoxyribodipyrimidine photo-lyase	Involved in repair of UV radiation-induced DNA damage	3
yfdH	Putative glycosyltransferase	Regulatory gene	3
dcoB	Oxaloacetate decarboxylase beta chain	Sodium ion transport	4
cysN	Sulfate adenylyltransferase subunit 1	May be the GTPase, regulating ATP sulfurylase activity	4

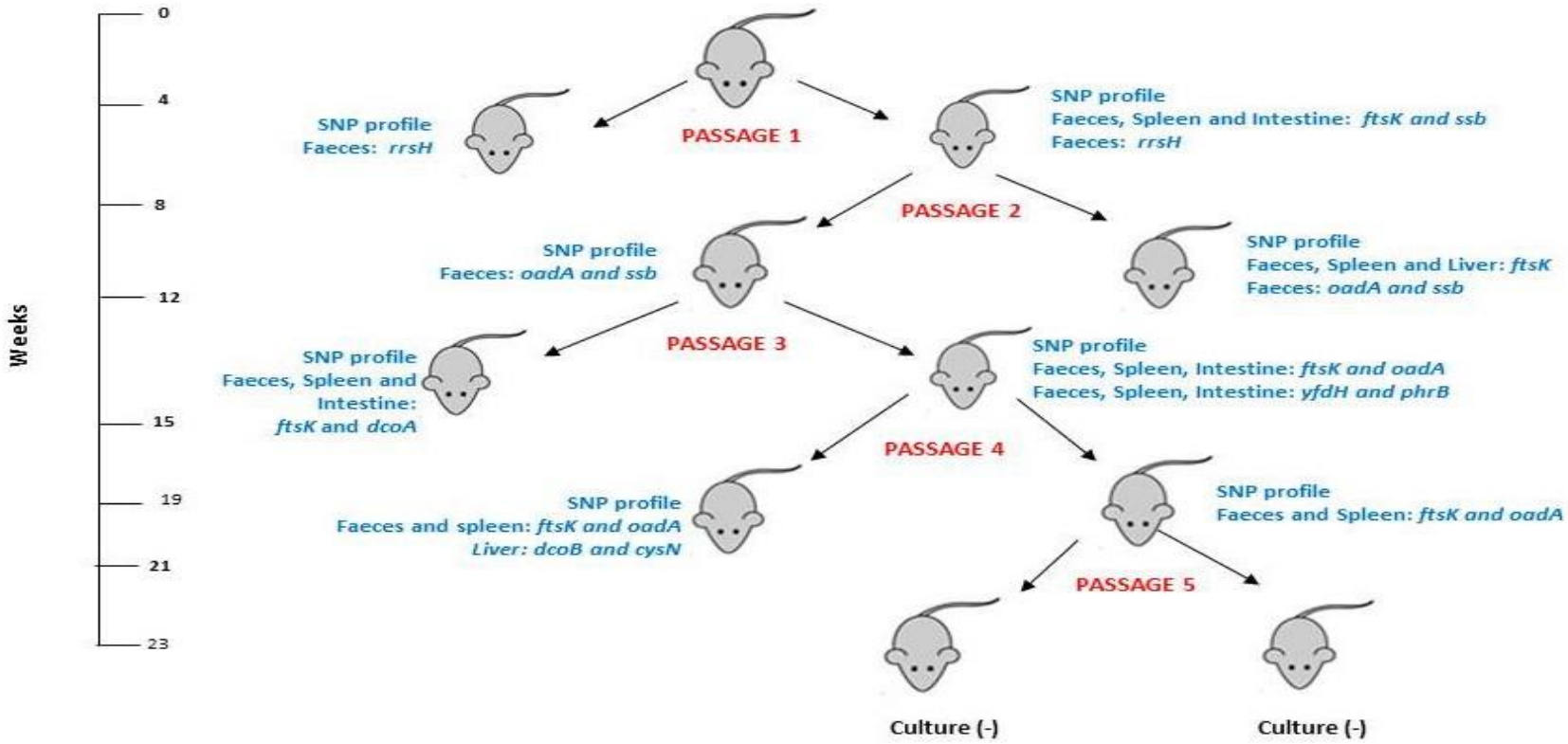


Figure 8.6 Complete transmission pathways for SNP detection for the chronic STM infection animal experiment. The SNP profiles are presented as associated to particular animal and the respective passage (colour blue).

8.5 Discussion

We have reproduced systemic STM infection in two strains of mice in order to measure the emergent genomic variation associated with presence of STM polymorphisms during acute and chronic disease. This approach was highly novel, and we are aware of one similar study (Sondberg and Jelsbak, 2016), which was published after completion of these experiments (244).

In contrast to the large body of literature describing animal models of typhoid, experimental models of non-typhoidal *Salmonella* have not been extensively studied. A mouse model of infection with serovar Typhimurium has been used to investigate enteric fever rather than non-typhoidal salmonellosis (11). Microbiologists used these models to examine the mechanisms and duration of bacterial excretion. Lam and Monack (2014) studied the changes present in a set of isolates in time, where the greater dynamic changes were observed at early time points in infection, showing a dramatic decrease in the number of *Salmonella* strains detected in the faeces when reaching the end of the experiment (62). Studies such as that presented here, which are focused on exploring the adaptive evolution of *Salmonella* in association with its host, remain uncommon.

Our observations correlate well with the recent study by Sondberg and Jelsbak (2016), which investigated STM adaptation during experimental chronic infection of mice. They documented a number of mutations in STM during a short-term experimental chronic infection of STM, and analogous to our findings, identified allele differences present in stools and organs such as liver, spleen and large intestine (244). The timeliness of these polymorphisms was also similar to our study. Once a variant was detected in stools, it was likely to continue being identified over time until reaching a particular moment in time, over four weeks, where the gene differences are no longer found. Sondberg and Jelsbak (2016) described not only distinct SNPs but also samples

from stools and tissues belonging to different animals sharing similar SNPs, which indicated transmission between the animals in their study (244). However, no horizontal transmission between animals under our care was observed. This may have been due to the in vitro characteristics present during the experiments, where cleaning of the housing was constantly performed, minimizing an increase on the exposure to excreted STM within the cage.

The focus of this study was on within-host variations and adaptations in STM genomes. Interestingly, we observed polymorphism in the *ftsK* gene, which became fixed in subsequent passages in both the acute and also the chronic transmission pathways. This gene has been described as an essential cell division protein, with a confirmed role conferring increased resistance to DNA damage (245). Our results suggest that this gene can be important in defining virulence phenotype of STM association and in the success of STM clades as human pathogens despite the fact that in terms of actual consequence generated from the presence of this SNP, no amino acid change took place. Gene *ftsK* still presents a relevant function in terms of active cell division, which could promote the ability of STM to remain actively infectant (Chapter 7). Our experiments highlighted the remarkable stability of the STM genome in acute and chronic infections. However, at the same time our findings indicate that acute STM infection presents a higher likelihood of mutations in STM genomes, including non-synonymous SNPs leading to amino acid changes. These were not present through the chronic infection model; all SNPs detected were synonymous. Perhaps these changes that took place through longer period of time (23 weeks) may be explained mostly in terms of host factor's influence rather than the infecting organism itself. The idea behind selecting the particular mice strain used for developing a chronic model of infection was to effectively mimic the chronic trend in terms of carriage that chronic human salmonellosis displays. Other mice strains, even the previously used BALB/c would not have

generated the chronicity that we observed. This chronic carriage displayed a number of fixed SNPs passing on through generations of mice in time without presenting amino acid changes and retaining the SNPs through the passages. A future direction of this work would be to continue to elucidate the actual adaptation strategies employed by the pathogen in terms of interactions with the host. By approaching an understanding these adaptations within STM genome, contributions towards STM infections reduction in human population can be achieved.

We acknowledge the relatively short duration of our models, particularly the acute one, and the effect it could have on the interpretation of our findings. Further studies of STM excretion over longer periods of time may warrant reconfirming our conclusions. However, our results are supported by genome sequencing experiments of STM strains cultured from human carriers showing limited genomic variation (5 SNPs or fewer) associated with short- and long-term carriage (173,216,246-248). Though the models encountered issues to extend the analysis in terms of the inability of the inoculum to generate more than 5 and 4 passages within acute and chronic model respectively, the results were sufficient to address the main objective of the study, which was to generate findings applicable to the interpretation of WGS data from human infections, so that human STM transmission pathways may be better understood (249-256).

In conclusion, the results of controlled experimental models of acute and chronic salmonellosis suggest acute infection is more likely to lead to genomic polymorphisms than chronic infection. Chronic infection model in animals revealed high stability of the genome through time, becoming persistently present through time and also through infected individual. The difference in genomic polymorphisms may reflect the size of microbial load during acute and chronic infections and can be an important factor in the interpretation of possible transmission pathways when STM carrier hosts are involved.

Chapter 9: Conclusions

9.1. Drawing the findings together

The main aim of this thesis was to investigate the molecular epidemiology of STM in New South Wales, Australia in order to gain new insights into the dynamics of seasonal epidemics. This region presented a unique setting for this study because of its access to quality pathology services and a well-established public health surveillance system. The subtyping techniques routinely performed for diagnostic testing in NSW, such as MLVA typing, have performed accurately and with sufficient discriminatory power to individualize isolates and, in most cases, associate them with particular settings and circumstances. Typing of STM infections in NSW is typically uncomplicated as the majority of STM infections in developed countries such as Australia are caused by a single strain, in contrast to developing countries where salmonellosis cases are often associated with co-infection by multiple strains, which complicates laboratory surveillance and deciphering of transmission pathways.

The emerging fields of genomic epidemiology and genomics-enhanced surveillance bridge the fields of microbiology, epidemiology and genomics in order to enhance understanding of the dynamics infectious disease epidemics, including STM (257-261). The application of enhanced genomic surveillance methods, such as whole genome sequencing, will certainly improve public health laboratory surveillance by providing a better understanding of the nature and molecular markers of dominant and emerging STM clades (262-265). Our main goal was to present a series of analyses to support the implementation of genomic-based surveillance, to improve the recognition of transmission pathways, which will ultimately lead to better understanding and therefore control of foodborne diseases such as STM.

Several countries such as England, US and some states in Australia, have already implemented genomic diagnosis and surveillance, transforming the diagnostic practice in clinical microbiology and reducing the turnaround time of testing (266-268). Implementation of WGS involves a series of modifications to the current laboratory workflows including acquisition of high performance computers and up skilling of laboratory personnel involved in the generation and analysis of the data produced by molecular subtyping and whole genome sequencing (269-270). The results presented in this thesis have added significant body evidence and new insights that may improve the quality of surveillance and control of STM disease in developed countries and elsewhere. This chapter draws together the findings from each of the thesis's chapters and discusses encountered limitations and the implications for future research in this area.

9.1.1. Objective 1 - To explore the capacity of multi-locus sequence typing to identify and differentiate the most common serovars of *Salmonella enterica* co-circulating in NSW

We compared the resolution of multi-locus sequence typing (MLST) with conventional serotyping to infer *Salmonella* serovars (271-272). Results generated using conventional serotyping of *Salmonella* and molecular serotypes inferred from MLST were congruent when applied to *Salmonella* strains circulating in NSW and Australia. These observations suggest that the use of molecular serotyping to characterize *Salmonella* in NSW and Australia can minimize the considerable challenges of traditional serotyping. Although the number of isolates included in the analysis was limited, the results confirmed the ability of the MLST-7 technique to infer serotypes. Based on the findings of this study, it is recommended that diagnostic laboratories consider implementing MLST-7 as a good and reliable supplement or alternative to serotyping. It

is clear that serotyping is still routinely used in many diagnostic laboratories Whilst moving away from traditional serotyping would require a change in the workflow of public health microbiology laboratories, the routine use of MLST or another similarly high throughput discriminatory sequence-based typing method would provide a clearer picture of short-term and long-term epidemiology and transmission routes of *Salmonella*, enabling data comparison and disease control in a global context. It is important to mention though, that this recommendation is only based upon the comparison between serotyping and MLST, even though as earlier discussed, the latter presents serious discriminatory issues. Ideally, all public health reference laboratories should be heading towards an organized and systematic whole genome sequence-based analysis. With better access to sequencing and bioinformatics resources, the serotype inference and more detail analysis using wgMLST for common serotypes should be based upon conduction of WGS.

9.1.2. – Objective 2 – To investigate the structure and mechanisms of seasonal epidemics of human STM disease in NSW

Our findings confirmed that the power and resolution of MLVA characterization was sufficient to enable detection of temporal changes in STM populations. This approach can be used to identify and quantify the existence and persistence of epidemic clades within specific seasons; either associated with large number of individual cases, community outbreaks or for general public health surveillance. An increase in STM cases during summer months as well as an increase in the number and diversity of MLVA profiles detected in the community during warms months was described and measured. There are several possible explanations to these observations. First, experts such as Akil et al (2014) have warned about an inevitable rise in STM cases due to global

warming. This might even be playing a role on the general decrease in presentation of typhoid related *Salmonella*s, becoming gradually replaced by a wide host range of non-typhoidal *Salmonella* serovars which, due to their multiple transmission routes, are expected to difficult the already highly incident salmonellosis profile. Second, it is the increased selection pressure driven by the extensive farming and food production practices which is most probably generating different modification, mostly in the environment, generating an impact in the ecological niches involved in presentation and sustainability over time of the organisms (275-278).

Substantial proportions of STM isolates included in clusters showed seasonal fluctuations corresponding to changes in STM incidence, with a relatively constant STM population diversity over the study period, reassuring the stable profile with predominance of well identified STM MLVA profiles. These findings reveal and confirm, as stated in 2016 by Earle et al, that not only the stability of the loci within the STM genome targeted by MLVA typing, but also the need to consider the existence of permanent links between human infections and animal reservoirs that could be involved in the constant dissemination of particular endemic MLVA. There have already been several suggestions in terms of the role the animal population has in the dissemination and persistence of STM through the food chain.

Our retrospective sequencing of STM isolates confirmed the high resolution of WGS and enhanced genomics surveillance, not only for the accurate identification and subtyping of pathogens, but also as a reliable tool for outbreak investigation. It has proven to be the most effective technology in order to prospectively identify outbreaks as well as providing information in terms of transmission events. Its ability to complement existing epidemiological tools by means of reconstructing transmission chains have allowed the identification of otherwise unrecognizable epidemiological links. We have managed to prove the added value of genome sequencing in the

investigation of point source community outbreaks associated with gastroenteritis caused by STM. However, the transition from MLVA based surveillance to genomic surveillance should be carefully planned, as the MLVA profiles cannot be inferred from WGS data.

9.1.3. Objective 3 – To characterize attributes of STM populations that are associated with seasonal epidemics of human STM disease in NSW

Despite public health efforts, *Salmonella* remains the most frequent cause of severe foodborne gastroenteritis in Australia and worldwide. We aimed to determine whether comparatively high burden seasonal STM epidemics could be predicted by STM activity in the preceding months, specifically the total number of STM cases and the appearance of new MLVA profiles. The aim of this was to develop a simple but reliable tool to act as an early warning system, to inform public health control and prevention efforts ahead of the peak summer months. We were able to determine that the severity of STM epidemics in NSW was associated with an increase in particular successful STM clades in the cooler months preceding peak summer months. The presence of gene variations associated with the tandem repeat numbers appeared concentrated within individual MLVA clonal complexes, which are the main characters throughout the numerous STM infections yearly. These potentially shedding reservoirs of STM strains in the environment, based upon their constant presence as the identified cause linked to several salmonellosis infections, might enable STM to become successfully persistent in the environment.

Endemic/novel STM MLVA profiles remained quite stable until 2012. There was an unexplained and sudden increase in the population diversity of STM in 2012, possibly related to genomic recombination events as well as environmental conditions. However, the introduction of

particularly uncommon profiles has not been seen in relation to specific outbreaks, only with individual human cases, where maybe particular circumstances related to either the sociocultural background or specific situations that could have increased the risk of exposure for the affected patients triggering the appearance of those novel MLVA profiles. Further analysis of host-pathogen interactions and STM populations in specific environmental reservoirs can explain the mechanisms of these shifts in STM population diversity.

9.1.4. Objective 4 – To compare the resolution power of MLST, MLVA and genome sequencing for public health laboratory surveillance

The results presented in Chapters 5 and 6 compared the resolution of three different but complimentary approaches to STM genotyping. MLST was found to be capable of inferring serovars accurately, supporting recent evidence from studies conducted in other geographical locations. Although the number of *Salmonella* serovars on our data set was not large, it captured the diversity of the top 10 most common serovars circulating in NSW. In addition, MLST was easy and fast to perform, and MLST results correlated with serotyping without exception, even correctly identifying genetic variants of STM, such as the monophasic strains. The reliable identification of the latter, often associated with multiple markers of antibiotic resistance, is of particular importance. However, the resolution of assigned Sequence Types (ST) appeared to be insufficient for public health surveillance purposes as there were numerous MLVA profiles identified within the ST19, the most common STM sequence type. In contrast to MLST, the MLVA approach was able to identify and define epidemiologically relevant clades of STM through the characterization described in the previous objective. Both subtyping methods have

become complementary tools with which to study the genetic relatedness of STM strains in routine analysis as well as during epidemiological investigations (278-280). Even though MLVA cannot be inferred based on WGS, it is still a reasonably fast, reliable and not highly expensive surveillance tool, with enough discriminatory power to preliminarily relate cases of Typhimurium infection.

Whole genome sequencing, as the third approach for STM identification and characterization, provides superior resolution that can enable precise identification of the microorganism down to the ST level, define outbreaks and enable identification of, and linkage to, particular source(s) as well as identification of previously unidentified epidemiological links that were not detected using conventional epidemiological and laboratory methods. Public health investigation of STM related outbreaks using only MLVA typing have underestimated the size and duration of the outbreaks in some instances. This situation could be avoided by widespread adoption of WGS for outbreak investigation.

9.1.5. Objective 5 – To examine variations in core and accessory genomes of successful STM clades

Results generated from experiments performed to address Objectives 1 to 4 enabled us to approach more fundamental questions about genomic differences to explain the epidemic capacity of specific STM clades, and to extend recent reports regarding determination of *Salmonella* invasiveness by a small number of virulence genes. We examined the genomic variation among epidemic STM isolates using core and accessory genomes established from complete sequences of STM isolates, and performed comparative genomic analysis between sporadic and epidemic

STM clades based on their MLVA and genomic polymorphisms. The findings emphasized the stability of accessory genomes, which was previously underappreciated and requires further in vivo validation (281-282). Interestingly, we have also identified gene loss in epidemic clades as a potential contributor to the pathogenic success of an STM MLVA profile, evidenced by an increase in the number of cases caused by particular profiles following gene loss.

9.1.6. Objective 6 – To identify within- and between-host variation and adaptation in STM genomes in models of acute and chronic infection

The aim of Objective 6 was to improve understanding of the genomic variations associated with different transmission pathways in order to accurately interpret STM genomic variations in human cases. Furthermore, it would be interesting to focus on the variation and adaptability occurring in STM chronic carriers, where the adaptive evolution of *Salmonella* in association with its host can become a critical bottleneck. In our model of chronic salmonellosis, we were able to explore STM adaptation strategies evidenced by the appearance of a limited number of mutations that did not compromise the ability of STM to maintain infection within the host. Rather, the STM isolate was able to successfully generate a persistent infection without severely compromising the host, which secured its presence in the experimental environment. The persistence in time of particular fixed SNPs related to essential cell functions as well as a conferred resistance to DNA damage were clear examples of adaptive genomic modifications. .

Compared to the chronic infection mode, more genomic changes occurred in the acute STM infection model, with the isolate displaying enhanced pathogenicity in terms of compromised animal health (283). Furthermore, some of the SNPs detected within the acute infection generated

amino acid changes, result that was not observed with the SNPs identified in the chronic infection model. There have been recent reports of STM strains cultured from human carriers showing similar limited genomic variation associated with short- and long-term carriage. Our results demonstrated that the STM core genome is absolutely stable; accessory genome is relatively stable, however capable of generating limited number of mutations in order to secure a niche within the environment and also as a way of adapting in order to establish chronic infection (285). These findings would be of relevance in the study of particular situations in patients classified as chronic carriers. It would be interesting to follow the model published by Im *et al*, (2016) where stools were collected randomly from African population. Findings showed presence of NTS in almost 25% of the stool samples, with none of the participants displaying signs associated with illness. Since developing countries are expected to present illness, the idea of taking this concept to developing regions with high NTS incidence would be of interest.

9.2. Study limitations

Several limitations of our research have to be acknowledged. We studied epidemics of STM in one jurisdiction of a large country, with relatively limited number of cases under genomic study, a specific laboratory testing and surveillance environment and microbial ecology for relatively short period of time. The small number of selected genomes could bias our identification of genomic changes within hosts and between hosts. However, we believe that the complexity of multi-factorial epidemiology of STM and cost of subtyping made, for this thesis purposes, an overarching study of genome dynamics unfeasible and the research of genomic variations under controlled conditions of experimental models and defined clusters is the way to tackle the relationships between genomic markers and public health outcomes. A larger study is needed to verify our observations of STM clonal changes and the relationships between novel MLVA

profiles and seasonal epidemics of human infections. If we add to this the fact that the each one of these isolates belonged to independent human STM infection cases which might have gone through their own individual adaptation processes, our findings might be too limited in terms of their ability to be generalized. Future lines should consider increasing the number of analysed genomes per years, as well as keeping a larger diversity of profiles considered for the analysis; only that way we might be able to, with certainty, extrapolate our results to the larger spectrum of STM illness.

9.3. Concluding remarks: implications for future research

The findings presented in this thesis have improved the understanding of the dynamics of salmonellosis epidemics in NSW, particularly related to STM. Also, we established the key role that genomic variability in STM genome presents as a way of interpretation of possible transmission pathways when STM carrier hosts are involved. The results have highlighted several areas that have practical implications for future research in molecular epidemiology of salmonellosis and public health surveillance. These include the complex structure of STM epidemics, represented not by actively changing STM populations but quite on the contrary, by highly dominant clades, with a limited level of diversity. This characteristic in terms of endemicity and genomic stability, makes us wonder whether, as shown through the successful / non-successful comparison and by means of the animal models, this region's local successful clades have become adapted to their environment, generating slight modifications within their genome so they can ensure a constant presence throughout the years to predominate in NSW, as well as taking advantage of beneficial environmental conditions during peak seasons. This adaptation could even become, if it hasn't already occurred, a host-adaptation, which would certainly change the public health approach towards the control of foodborne illness associated with STM. Further studies,

specifically analysis of a larger and more diverse set of isolates, are required in order to fully understand the epidemiological impact of STM genomic variability in NSW and beyond.

The generalizability of our results to other developed country settings with less endemic serovars could be a useful tool to better understand the nature of the MLVA profiles for particular regions associated with STM cases. The prediction indexes would become handy for the public health authorities' decision-making processes.

In conclusion, our research findings provided important insights into the dynamics of community epidemics of *Salmonella* Typhimurium and identified genomic polymorphisms that contribute to the establishment of epidemic clades. Our results and analyses have generated important evidence to guide the interpretation of *Salmonella* Typhimurium public health laboratory surveillance results and the integration of whole genome sequencing into control of foodborne diseases. We anticipate that this research will provide a starting point for future experiments, and the collection and analysis of larger genomic datasets in order to fully realize the power of genomic surveillance.

References

(1) Kirk MD, Angulo FJ, Havelaar AH, Black RE. Diarrhoeal disease in children due to contaminated food. *Bull World Health Organ*. 2016; 95(3):233–4.

(2) Kalyoussef S, Feja KN. Foodborne illnesses. *Adv Pediatr*. 2014; 61(1):287-312.

(3) Braden CR, Tauxe RV. Emerging trends in foodborne diseases. *Infect Dis Clin North Am*. 2013; 27(3):517-33.

(4) de Jong B, Ekdahl K. The comparative burden of salmonellosis in the European Union member states, associated and candidate countries. *BMC Public Health*. 2006; 6:4. doi: 10.1186/1471-2458-6-4.

(5) Collier SA, Stockman LJ, Hicks LA, et al. Direct healthcare costs of selected diseases primarily or partially transmitted by water. *Epidemiol Infect*. 2012; 140(11):2003-13.

(6) Havelaar AH, Haagsma JA, Mangen M-JJ, *et al*. Disease burden of foodborne pathogens in the Netherlands. *Int J Food Microbiol*. 2009; 156(3):231–8.

(7) Ford L, Miller M, Cawthorne A, Fearnley E, Kirk M. Approaches to the Surveillance of Foodborne Disease: A Review of the Evidence. *Foodborne Pathog Dis*. 2015; 12(12):927-36.

(8) Havelaar AH, Kirk MD, Torgerson PR *et al*. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. *PLoS Med*. 2015; 12(12):e1001923. doi: 10.1371/journal.pmed.1001923. eCollection 2015 Dec.

(9) Kirk MD, Pires SM, Black RE *et al*. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLoS Med*. 2015;12: e1001921. doi: 10.1371/journal.pmed.1001921. eCollection 2015 Dec.

(10) Berger CN, Sodha SV, Shaw RK, *et al*. Fresh fruit and vegetables as vehicles for the transmission of human pathogens. *Environ Microbiol*. 2010; 12(9):2385–97.

(11) Santos AC, Roberts JA, Cook AJ *et al*. *Salmonella* Typhimurium and *Salmonella* Enteritidis in England: costs to patients, their families, and primary and community health services of the NHS. *Epidemiol Infect*. 2011; 139(5):742-53.

(12) Havelaar AH, Kirk MD, Torgerson PR, *et al*. World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. Seidlein von L, editor. *PLoS Med*. 2015; 12(12): e1001923. doi: 10.1371/journal.pmed.1001923. eCollection 2015 Dec.

(13) Chen PL, Li CY, Hsieh TH *et al*. Epidemiology, disease spectrum and economic burden of non-typhoidal *Salmonella* infections in Taiwan, 2006-2008. *Epidemiol Infect*. 2012; 140(12):2256-63.

(14) Gal-Mor O, Boyle EC, Grassl GA. Same species, different diseases: how and why typhoidal and non-typhoidal *Salmonella* enterica serovars differ. *Front Microbiol*. 2014; 5:391. doi: 10.3389/fmicb.2014.00391.

(15) Majowicz SE, Musto J, Scallan E, *et al*. The Global Burden of Nontyphoidal *Salmonella* Gastroenteritis. *Clin Infect Dis.* 2010; 50(6):882-9.

(16) Barrow P, Methner U. *Salmonella* in Domestic Animals (internet). CAB International, UK 2013. (cited 2015 May 30th). 547p. Available from: https://library.sydney.edu.au/.

(17) Amajoud N, Bouchrif B, Maadoudi El M, et al. Prevalence, serotype distribution, and antimicrobial resistance of *Salmonella* isolated from food products in Morocco. *J Infect Dev Ctries*. 2017; 11(2):1–7.

(18) Coburn B, Grassl GA, Finlay BB. *Salmonella*, the host and disease: a brief review. *Immunol Cell Biol*. 2007; 85(2):112-8.

(19) Pires SM, Vieira AR, Hald T, Cole D. Source attribution of human salmonellosis: an overview of methods and estimates. *Foodborne Pathog Dis*. 2014; 11(9):667-76.

(20) Bäumler AJ, Winter SE, Thiennimitr P, Casadesús J. Intestinal and chronic infections: *Salmonella* lifestyles in hostile environments. *Environ Microbiol Rep*. 2011; 3(5):508-17.

(21) Van Den Brandhoff WE, De Wit GA, De Wit M, Van Duynhoven Y. Costs of gastroenteritis in The Netherlands. *Epidemiol Infect*. 1999; 132:211-21.

(22) Amagliani G, Brandi G, Schiavano GF. Incidence and role of *Salmonella* in seafood safety. *Food Research International*. 2012; 45:780-788.

(23) Wheeler JG, Sethi D, Cowden JM *et al*. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *Brit Med J.* 1999; 318(7190):1046-50.

(24) Kirk M, Ford L, Glass K, Hall G. Foodborne Illness, Australia, Circa 2000 and Circa 2010. *Emerg Infect Dis*. 2014; 20(11):1857–64.

(28) Voetsch AC, Van Gilder TJ, Angulo FJ, *et al*. (2006) FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. *Clin Infect Dis*. 2004;38 Suppl 3: S127-34.

(25) Mullner P, Spencer SE, Wilson DJ *et al*. Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. *Infect Genet Evol.* 2009; 9:1311-19.

(26) Chousalkar KK, Sexton M, McWhorter A *et al*. *Salmonella* Typhimurium in the Australian Egg Industry: Multidisciplinary Approach to Addressing the Public Health Challenge and Future Directions. *Crit Rev Food Sci Nutr*. 2017; 57(12):2706-11.

(27) Fearnley E, Raupach J, Lagala F, Cameron S. *Salmonella* in chicken meat, eggs and humans; Adelaide, South Australia, 2008. *Int J Food Microbiol*. 2011; 146(3):219-27.

(28) Ford L, Glass K, Veitch M *et al*. Increasing Incidence of *Salmonella* in Australia, 2000-2013. *PLoS One.* 2016;11: e0163989.

(29) Hall G, Yohannes K, Raupach J, Becker N, Kirk M. Estimating community incidence of *Salmonella*, Campylobacter, and Shiga toxin-producing *Escherichia coli* infections, Australia. *Emerg Infect Dis*. 2008; 14(10):1601-09.

(30) Gantois I, Ducatelle R, Pasmans F *et al*. Mechanisms of egg contamination by *Salmonella* Enteritidis. *FEMS Microbiol Rev*. 2009; 33(4):718-38.

(31) Sánchez-Vargas FM, Abu-El-Haija MA, Gómez-Duarte OG. *Salmonella* infections: an update on epidemiology, management, and prevention. *Travel Med Infect Dis*. 2011; 9(6):263-77.

(32) Graziani C, Luzzi I, Owczarek S, Dionisi AM, Busani L. *Salmonella* enterica Serovar Napoli Infection in Italy from 2000 to 2013: Spatial and Spatio-Temporal Analysis of Cases Distribution and the Effect of Human and Animal Density on the Risk of Infection. *PLoS One.* 2015; 10(11): e0142419. doi: 10.1371/journal.pone.0142419. eCollection 2015.

(33) Ao TT, Feasey NA, Gordon MA, Keddy KH, Angulo FJ, Crump JA. Global burden of invasive nontyphoidal *Salmonella* disease, 2010. *Emerg Infect Dis.* 2015; 21(6):941-49. https://dx.doi.org/10.3201/eid2106.140999.

(34) The integrated surveillance of *Salmonella* in Denmark and the effect on public health. Volume 3: Risk management strategies, monitoring and surveillance. In: Food Safety Assurance and Veterinary Public Health. Smulders FJM, Collins JD (eds.). Wageningen, The Netherlands: Wageningen Academic Publishers, 2005, pp. 213–38.

(35) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2015 (cited 2016 September 30). 30p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(36) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2015 (cited 2016 September20). 39p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(37) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2011 (cited 2013 September 30). 30p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(38) Ashton PM, Nair S, Peters TM *et al*. Identification of *Salmonella* for public health surveillance using whole genome sequencing. *Peer J*. 2016; 4:e1752. doi: 10.7717/peerj.1752. eCollection 2016.

(39) Marzel A, Desai PT, Goren A *et al*. Persistent Infections by Nontyphoidal *Salmonella* in Humans: Epidemiology and Genetics. *Clin Infect Dis*. 2016; 62(7):879-86.

(40) Gilbert GL. Using MLVA to type strains of *Salmonella* Typhimurium in New South Wales. *NSW Public Health Bull*. 2008; 19(2):29–31.

(41) Wang Q, Chiew R, Howard P, Gilbert GL. *Salmonella* typing in New South Wales: current methods and application of improved epidemiological tools. *NSW Public Health Bull*. 2008; 19(2):24–5.

(42) Sintchenko V, Wang Q, Howard P, *et al*. Improving resolution of public health surveillance for human *Salmonella* enterica serovar Typhimurium infection: 3 years of prospective multiple-locus variable-number tandem-repeat analysis (MLVA). *BMC Infect Dis*. 2012; 12(1):78. doi: 10.1186/1471-2334-12-78.

(43) Lindstedt BA, Vardund T, Aas L, Kapperud G. Multiple-locus variable-number tandem-repeats analysis of *Salmonella* enterica subsp. enterica serovar Typhimurium using PCR multiplexing and multicolour capillary electrophoresis. *J Microbiol Methods*. 2004; 59(2):163-72.

(44) Lindstedt BA, Heir E, Gjernes E, Kapperud G. DNA fingerprinting of *Salmonella* enterica subsp. enterica serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. *J Clin Microbiol*. 2003; 41(4):1469-79.

(45) Phillips A, Sotomayor C, Wang Q, Holmes N, Furlong C, Ward K, *et al*. Whole genome sequencing of *Salmonella* Typhimurium illuminates distinct outbreaks caused by an endemic multi-locus variable number tandem repeat analysis type in Australia, 2014. BMC Microbiology. *BMC Microbiology*. 2016; 14:1–9. doi: 10.1186/s12866-016-0831-3.

(46) Brenner FW, McWhorter-Murlin AC. Identification and serotyping of *Salmonella*. Atlanta: Centers for Disease Control and Prevention, 1998. *J Clin Microbiol*. 2000; 38(7):2465-67.

(47) Bachmann NL, Petty NK, Ben Zakour NL, Szubert JM, Savill J, Beatson SA. Genome analysis and CRISPR typing of *Salmonella* enterica serovar Virchow. *BMC Genomics.* 2014; 15:389. doi: 10.1186/1471-2164-15-389.

(48) Bell RL, Jarvis KG, Ottesen AR, McFarland MA, Brown EW. Recent and emerging innovations in *Salmonella* detection: a food and environmental perspective. *Microb Biotechnol.* 2016; 9(3):279-92.

(49) Bekal S, Berry C, Reimer AR *et al*. Usefulness of High-Quality Core Genome Single-Nucleotide Variant Analysis for Subtyping the Highly Clonal and the Most Prevalent *Salmonella* enterica Serovar Heidelberg Clade in the Context of Outbreak Investigations. *J Clin Microbiol*. 2016; 54(2):289-95.

(50) Snitkin ES, Zelazny AM, Thomas PJ *et al*. Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med.* 2012; 148ra116. doi: 10.1126/scitranslmed.3004129.

(51) Liu F, Kariyawasam S, Jayarao BM *et al*. Subtyping *Salmonella* enterica serovar enteritidis isolates from different sources by using sequence typing based on virulence genes and clustered regularly interspaced short palindromic repeats (CRISPRs). *Appl Environ Microbiol*. 2011; 77(13):4520-26.

(52) Urwin R, Maiden MCJ. Multi-locus sequence typing: a tool for global epidemiology. *Trends in Microbiol*. 2003; 11(10):479-87.

(53) Sharma P, Dahiya S, Balaji V *et al*. Typhoidal *Salmonella*e: Use of Multi-Locus Sequence Typing to Determine Population Structure. *PLoS One*. 2016; 11(9): e0162530. https://doi.org/10.1371/journal.pone.0162530.

(54) Papadopoulos T, Petridou E, Zdragas A, Mandilara G, Nair S, Peters T. Comparative study of all *Salmonella* enterica serovar Enteritidis strains isolated from food and food animals in Greece from 2008 to 2010 with clinical isolates. *Eur J Clin Microbiol Infect Dis*. 2016; 35(5):741-6.

(55) Thomson NR, Clayton DJ, Windhorst D, *et al*. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Res*. 2008; 18(10):1624–37.

(56) M Ngoi ST, Thong KL. Molecular characterization showed limited genetic diversity among *Salmonella* Enteritidis isolated from humans and animals in Malaysia. *Diagn Microbiol Infect Dis*. 2013; 77(4):304-11.

(57) Makendi C, Page AJ, Wren BW, Le Thi Phuong T, Clare S, Hale C *et al*. A Phylogenetic and Phenotypic Analysis of *Salmonella* enterica Serovar Weltevreden, an Emerging Agent of Diarrheal Disease in Tropical Regions. *PLoS Negl Trop Dis*. 2016 Feb 11;10(2): e0004446. https://doi.org/10.1371/journal.pntd.

(58) Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet.* 2011; 52(4):413-35.

(59) Yoshida C, Gurnik S, Ahmad A *et al*. Evaluation of Molecular Methods for Identification of *Salmonella* Serovars. *J Clin Microbiol*. 2016; 54(8):1992-98.

(60) Gallego B, Sintchenko V, Wang Q, Hiley L, Gilbert GL, Coiera E. Biosurveillance of emerging biothreats using scalable genotype clustering. *J Biomed Inform*. 2009; 42(1):66-73.

(61) Loman N, Watson M. So, you want to be a computational biologist? *Nat Biotechnol*. 2013; 31(11):996-8.

(62) Octavia S, Wang Q, Tanaka MM, Kaur S, Sintchenko V, Lan R. Delineating community outbreaks of *Salmonella* enterica serovar Typhimurium by use of whole-genome sequencing: insights into genomic variability within an outbreak. *J Clin Microbiol.* 2015; 53(4):1063-71.

(63) Le VT, Diep BA. Selected insights from application of whole-genome sequencing for outbreak investigations. *Curr Opin Crit Care*. 2013; 19(5):432-9.

(64) Leopold SS. Editorial: Our love affair with technology and the choices we make. *Clin Orthop Relat Res*. 2014; 472(10):2907-8.

(65) Staden R. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Res*. 1979; 6(7):2601-10.

(66) Leekitcharoenphon P, Hendriksen RS, Le Hello S *et al*. Global Genomic Epidemiology of *Salmonella* enterica Serovar Typhimurium DT104. *Appl Environ Microbiol*. 2016; 82(8):2516-26.

(67) Miyamoto M, Motooka D, Gotoh K, Imai T, Yoshitake K, Goto N *et al*. Performance comparison of second- and third-generation sequencers using a bacterial genome with two chromosomes. *BMC Genomics*. 2014; 15:699. doi: 10.1186/1471-2164-15-699.

(68) Lampa S, Dahlö M, Olason PI, Hagberg J, Spjuth O. Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data. *Gigascience*. 2013; 25:2(1):9.

(69) Patel AK, Singhania RR, Pandey A, Joshi VK, Nigam PS, Soccol cr. Enterobacteriaceae, coliforms and E. coli. Encyclopaedia of Food Microbiology. Elsevier; 2014. p. 659-666.

(70) McClelland M, Sanderson K, Spieth J, *et al*. Complete genome sequence of *Salmonella* enterica serovar Typhimurium LT2. *Letters to Nature* 2001; 413(6858):852-56.

(71) Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis*. 2011; 17(1):7-15.

(72) Kubota K, Iwasaki E, Inagaki S *et al*. The human health burden of foodborne infections caused by Campylobacter, *Salmonella*, and Vibrio parahaemolyticus in Miyagi Prefecture, Japan. *Foodborne Pathog Dis*. 2008; 5(5):641-648.

(73) Korsgaard H, Madsen M, Feld NC, Mygind J, Hald T. The effects, costs and benefits of *Salmonella* control in the Danish table-egg sector. *Epidemiol Infect*. 2009; 137(6):828-36.

(74) Centres for Disease Control and Prevention (CDC). (internet) National *Salmonella* Surveillance Overview. Atlanta, Georgia: US Department of Health and Human Services, CDC, 2011. (cited October 25, 2013)
Available at:
http://www.cdc.gov/nationalsurveillance/PDFs/NationalSalmSurveillOverview_508.pdf

(75) Hohmann EL. Nontyphoidal salmonellosis. *Clin Infect Dis* 2001; 32(2):263-9.

(76) Trevejo RT. Epidemiology of Salmonellosis in California, 1990-1999: Morbidity, Mortality, and Hospitalization Costs. *Am J Epidemiol*. 2003; 157(1):48-57.

(77) Mahmoud B. *Salmonella* - A Dangerous Foodborne Pathogen (internet). USA: Intech: 2012 (cited 2014 May). Available from:
https://www.intechopen.com/books/*Salmonella*-a-dangerous-foodborne-pathogen/studies-on-pcr-based-rapid-detection-systems-for-*Salmonella*-spp-

(78) Barry EM, Gomez-Duarte O, Chatfield S *et al*. Expression and immunogenicity of pertussis toxin S1 subunit-tetanus toxin fragment C fusions in *Salmonella* Typhi vaccine strain CVD 908. *Infect Immun*. 1996; 64(10):4172-81.

(79) Octavia S, Sara J, Lan R. Characterization of a large novel phage-like plasmid in *Salmonella* enterica serovar Typhimurium. *FEMS Microbiol Lett*. 2015; 362: fnv044. doi: 10.1093/femsle/fnv044. Epub 2015 Mar 20.

(80) Qi XL, Wang HX, Bu SR, Xu XG, Wu XY, Lin DF. Incidence rates and clinical Symptoms of *Salmonella*, Vibrio parahaemolyticus, and Shigella infections in China, 1998-2013. *J Infect Dev Ctries*. 2016; 10(2):127-33.

(81) Levy BS, McIntire W. The economic impact of a food-borne salmonellosis outbreak. *JAMA*. 1974; 230(9):1281-82.

(82) Nadon CA, Trees E, Ng LK *et al*. Development and application of MLVA methods as a tool for inter-laboratory surveillance. *Euro Surveill*. 2013; 18(35):20565.

(83) Tauxe RV. Surveillance and investigation of foodborne diseases; roles for public health in meeting objectives for food safety. *Food Control*. 2002; 13:363-69.

(84) Tauxe RV, Doyle MP, Kuchenmüller T, Schlundt J, Stein CE. Evolving public health approaches to the global challenge of foodborne infections. *Int J Food Microbiol*. 2010;139 Suppl 1: S16-28.

(85) Cronquist AB, Mody RK, Atkinson R *et al*. Impacts of culture-independent diagnostic practices on public health surveillance for bacterial enteric pathogens. *Clin Infect Dis*. 2012;54 Suppl 5: S432-9. doi: 10.1093/cid/cis267.

(86) David JM, Sanders P, Bemrah N *et al*. Attribution of the French human Salmonellosis cases to the main food-sources according to the type of surveillance data. *Prev Vet Med*. 2013; 110(1):12-27.

(87) Majowicz SE, Edge VL, Fazil A *et al*. Estimating the Under-reporting Rate for Infectious Gastrointestinal Illness in Ontario. *Canadian Journal of Public Health* / *Revue Canadienne de Santé Publique.* 2005; 96(3):178-81.

(88) Sangal V, Harbottle H, Mazzoni CJ *et al*. Evolution and population structure of *Salmonella* enterica serovar Newport. *J Bacteriol*. 2010; 192(24):6465-76.

(89) Gillespie IA, O'Brien SJ, Adak GK, Ward LR, Smith HR. Foodborne general outbreaks of *Salmonella* Enteritidis phage type 4 infection, England and Wales, 1992-2002: where are the risks? *Epidemiol Infect*. 2005; 133(5):795-801.

(90) Fricke, WF., Mammel, MK., McDermott, PF., *et al*. Comparative genomics of 28 *Salmonella* enterica isolates: evidence for CRISR-mediated adaptive sub lineage evolution. *Journal of Bacteriology*. 2011; 193(14):3556–68.

(91) Grimont, P.A.D., Grimont, F. and Bouvet, P. (2000) Taxonomy of the genus *Salmonella*. In: Wray, C. and Wray, A. (eds) *Salmonella* in Domestic Animals. CAB International, Wallingford, UK.

(92) Kauffmann, F. The Bacteriology of Enterobacteriaceae. Collected Studies of the Author and his Co-Workers. The Bacteriology of Enterobacteriaceae. Collected Studies of the Author and his Co-Workers. Copenhagen: Munksgaard; 1966.

(93) Kauffmann, F. Judicial Commission of the International Committee on Systematics of Prokaryotes. The type species of the genus *Salmonella* Lignieres 1900 is *Salmonella* enterica (ex Kauffmann and Edwards 1952) Le Minor and Popoff 1987, with the type strain LT2T, and conservation of the epithet enterica in *Salmonella* enterica over all earlier epithets that may be applied to this species. Opinion 80. *Int J Syst Evol Microbiol*. 2005; 55(Pt 1):519-20.

(93) Crosa JH, Brenner DJ, Ewing WH, Falkow S. Molecular relationships among the *Salmonella*e. *J Bacteriol.* 1973; 115(1):307-15.

(94) Penner JL. International Committee on Systematic Bacteriology Taxonomic Subcommittee on Enterobacteriaceae. *Int J Syst Bacteriol*. 1988; 38:223-4.

(95) Ewing WH. The nomenclature of *Salmonella*, its usage, and definitions for the three species. *Can J Microbiol.* 1972; 18(11):1629-37.

(96) Lin-Hui Su, Cheng-Hsun Chi. *Salmonella*: Clinical Importance and Evolution of Nomenclature. *Med J.* 2007; 30(3):210-19.

(97) Tindall BJ, Grimont PAD, Garrity GM, Euzeby JP. Nomenclature and taxonomy of the genus *Salmonella*. *Internat J Syst Evol Microbiol.* 2005; 55(Pt 1):521–24.

(98) Cox JM, Pavic A. *Salmonella*. Encyclopaedia of Food Microbiology. Elsevier; 2014. p. 322-331.

(99) Poppoff, MY, Le Minor L. Antigenic formulas of the *Salmonella* serovars. 8th revision. WHO Collaborative Centre for Reference and Research on *Salmonella*. Paris, Institut Pasteur, 2001.

(100) Chen, H. M., Wang, Y., Su, L. H., & Chiu, C. H. Nontyphoid *Salmonella* infection: microbiology, clinical features, and antimicrobial therapy. *Pediatr Neonatol.*2013; 54(3): 147-52.

(101) Lynch M and Tauxe R. Bacterial infections in humans. 2009. Springer US.

(102) Blaser MJ, Newman LS. A review of human salmonellosis: I. Infective dose. *Rev Infect Dis*. 1982; 4(6):1096-106.

(103) Bell C, Kyriakides A. *Salmonella*: A Practical Approach to the Organism and its Control in Foods (internet). Wiley Online Library; 2007 (cited 2014, May). Available from: http://onlinelibrary.wiley.com/book/10.1002/9780470999455

(104) Uzzau, S., Brown, D.J., Wallis, T., Rubino, S., Leori, G., Casadusus, J., Platt, D.J. and Olsen, J.E. Host adapted serotypes of *Salmonella* enterica. *Epidemiol Infect.* 2000; 125(2):229–55.

(105) Stevens, M.P., Humphrey, T.J. and Maskell, D.J. Molecular insights into farm animal and zoonotic *Salmonella* infections. *Philos Trans R Soc London B Biol Sci*. 2009; 364(1530):2709–23.

(106) Bell RL, Jarvis KG, Ottesen AR, McFarland MA, Brown EW. Recent and emerging innovations in *Salmonella* detection: a food and environmental perspective. *Microb Biotechnol*. 2016; 9(3):279-92.

(107) Ferrato C, Chui L, King R, Louie M. Utilization of a molecular serotyping method for *Salmonella* enterica in a routine laboratory in Alberta Canada. *J Microbiol Methods*. 2017; 135:14-19. doi: 10.1016/j.mimet.2017.01.018.

(108) Coburn B, Grassl GA, Finlay BB. *Salmonella*, the host and disease: a brief review. *Immunol Cell Biol*. 2007; 85(2):112-18.

(109) Crump JA, Sjölund-Karlsson M, Gordon MA, Parry CM. Epidemiology, Clinical Presentation, Laboratory Diagnosis, Antimicrobial Resistance, and Antimicrobial Management of Invasive *Salmonella* Infections. *Clin Microbiol Rev*. 2015; 28(4):901-37.

(110) Jones TF, Ingram A, Cieslak PR, Vugia DJ, *et al*. Salmonellosis outcomes differ substantially by serotype. *J Infect Dis* 2008; 198(1):109-14.

(111) Sukhnanand S, Alcaine S, Warnick LD *et al*. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella* enterica serotypes. *J Clin Microbiol*. 2005; 43(8):3688-98.

(112) Grimont PAD, Weill FX. Antigenic Formulae of the *Salmonella* Serovars (ninth ed.) Paris: WHO Collaborating Centre for Reference and Research on *Salmonella*, Institut Pasteur; 2007.

(113) Sanderson KE, Roth JR. Linkage map of *Salmonella* Typhimurium, edition VII. *Microbiol Rev*. 1988; 52(4):485–532.

(114) Porwollik S. *Salmonella*: from genome to function (Internet) Horizon Scientific Press; 2011. 300p

(115) Frye J, Karlinsey JE, Felise HR *et al*. Identification of new flagellar genes of *Salmonella* enterica serovar Typhimurium. *J Bacteriol*. 2006; 188(6):2233-43.

(116) Matheson N, Kingsley RA, Sturgess K *et al*. Ten years' experience of *Salmonella* infections in Cambridge, UK. *J Infect.* 2010; 60:21-25.

(117) Taunay AE, Fernandes SA., Tavechio AT, *et al*. The role of public health laboratory in the problem of salmonellosis in São Paulo, Brazil. *Rev. Inst. Med. trop*. S. Paulo. 1996;38: 119-27.

(118) Park SH, Aydin M, Khatiwara A *et al*. Current and emerging technologies for rapid detection and characterization of *Salmonella* in poultry and poultry products. *Food Microbiol*. 2014; 38:250-62. doi: 10.1016/j.fm.2013.10.002.

(119) Torpdahl M, Skov MN, Sandvang D, Baggesen DL. Genotypic characterization of *Salmonella* by multilocus sequence typing, pulsed-field gel electrophoresis and amplified fragment length polymorphism. *J Microbiol Methods*. 2005; 63(2):173-84.

(120) Linnarsson S. Recent advances in DNA sequencing methods - general principles of sample preparation. *Exp Cell Res*. 2010; 316(8):1339-43.

(121) Wegener H, Hald T, Wong D, *et al*. *Salmonella* Control Programs in Denmark. *Emerging Infectious Diseases* 2003; 9(7): 774-80.

(122) Wuyts V, Mattheus W, De Laminne de Bex G *et al*. MLVA as a tool for public health surveillance of human *Salmonella* Typhimurium: prospective study in Belgium and evaluation of MLVA loci stability. *PLoS One*. 2013; 8(12):e84055. doi: 10.1371/journal.pone.0084055. eCollection 2013.

(123) Li J, Maclehose R, Smith K, Kaehler D, Hedberg C. Development of a *Salmonella* screening tool for consumer complaint-based foodborne illness surveillance systems. *J Food Prot.* 2011; 74(1):106-10.

(124) Hur J, Jawale C, Lee JH. Antimicrobial resistance of *Salmonella* isolated from food animals: A review. *Food Res Int*. 2012; 45(2):819-30.

(125) Aslam M, Checkley S, Avery B *et al*. Phenotypic and genetic characterization of antimicrobial resistance in *Salmonella* serovars isolated from retail meats in Alberta, Canada. *Food Microbiol*. 2012; 32(1):110-17.

(126) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2013 (cited 2014 September 5). 30p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(127) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2012 (cited 2014 September 5). 30p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(128) Annual NSW OzFoodNet Network Report. (Internet). NSW (AU); 2009 (cited 2014 September 5). 30p. Available from: http://www.health.nsw.gov.au/Infectious/foodborne/Pages/ozfo

(129) Wattiau P, Boland C, Bertrand S. Methodologies for *Salmonella* enterica subsp. enterica subtyping: gold standards and alternatives. *Appl Environ Microbiol*. 2011; 77(22):7877-85.

(130) Zhang S, Yin Y, Jones MB *et al*. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J Clin Microbiol*. 2015; 53(5):1685-92.

(131) Köser CU, Ellington MJ, Cartwright EJ *et al*. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog*. 2012;8(8): e1002824.

(132) Köser CU, Fraser LJ, Ioannou A *et al*. Rapid single-colony whole-genome sequencing of bacterial pathogens. *J Antimicrob Chemother*. 2014; 69(5):1275-81.

(133) Aanensen DM, Spratt BG. The multilocus sequence typing network: mlst.net. *Nucleic Acids Res*. 2005; 33: 728-33.

(134) Achtman M, Wain J, Weill F, Nair S, Zhou Z, Sangal V *et al*. Multilocus sequence typing as a replacement for serotyping in *Salmonella* enterica. *Plos Pathogens*. 2012; (8)6:1-19.

(135) Ranieri ML, Shi C, Moreno Switt AI, den Bakker HC, Wiedmann M. Comparison of typing methods with a new procedure based on sequence characterization for *Salmonella* serovar prediction. *J Clin Microbiol*. 2013; 51(6):1786-97.

(136) Harbottle H, White DG, McDermott PF, Walker RD, Zhao S. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella* enterica serotype Newport isolates. *J Clin Microbiol.* 2006; 44(7):2449-57.

(137) Maiden M, Bygraves J, Feil E, Morelli G, Russell J, Urwin R *et al*. Multilocus sequence typing: a portable approach to the identification of clades within populations of pathogenic organisms. *Proc Natl Acad Sci*. 1998; 95:3140-45

(138) Turner KM, Feil EJ. The secret life of the multilocus sequence type. *Int J Antimicrob Agents*. 2007; 29:129-35.

(139) Cooper JE, Feil EJ. Multilocus sequence typing--what is resolved. *Trends Microbiol*. 2004; 12(8):373-77.

(140) Hu H, Lan R, Reeves PR. Adaptation of multilocus sequencing for studying variation within a major clade: evolutionary relationships of *Salmonella* enterica serovar Typhimurium. *Genetics*. 2006; 172(2):743-50.

(141) Larsen MV, Cosentino S, Rasmussen S *et al*. Multilocus sequence typing of total-genome-sequenced bacteria. *J Clin Microbiol*. 2012; 50(4):1355-61.

(142) Prendergast DM, O'Grady D, Fanning S *et al*. Application of multiple locus variable number of tandem repeat analysis (MLVA), phage typing and antimicrobial susceptibility testing to subtype *Salmonella* enterica serovar Typhimurium isolated from pig farms, pork slaughterhouses and meat producing plants in Ireland. *Food Microbiol*. 2011; 28(5):1087-1094.

(143) Litrup E, Torpdahl M, Nielsen EM. Multilocus sequence typing performed on Campylobacter coli isolates from humans, broilers, pigs and cattle originating in Denmark. *J Appl Microbiol*. 2007;103(1):210-8.

(144) Ranjbar R, Ahmadi M, Memariani M. Multiple-locus variable-number tandem repeat analysis (MLVA) for genotyping of *Salmonella* enterica subspecies enterica serotype Infantis isolated from human sources. *Microb Pathog*. 2016; 100:299-304. doi: 10.1016/j.micpath.2016.10.012. Epub 2016 Oct 18.

(145) Best EL, Lindstedt BA, Cook A, *et al.* Multiple-locus variable-number tandem repeat analysis of *Salmonella* enterica subsp. enterica serovar Typhimurium: comparison of isolates from pigs, poultry and cases of human gastroenteritis. *J Appl Microbiol*. 2007; 103:565-72.

(146) Larsson JT, Torpdahl M, Petersen RF, *et al*. Development of a new nomenclature for *Salmonella* Typhimurium multilocus variable number of tandem repeats analysis (MLVA). *Euro Surveill*. 2009;14(15). pii: 19174.

(147) Roetzer A, Diel R, Kohl TA *et al*. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. *PLoS Med*. 2013;10: e1001387.

(148) Huehn S, Bunge C, Junker E, Helmuth R, Malorny B. Poultry-associated *Salmonella* enterica subsp. enterica serovar 4,12:d:- reveals high clonality and a distinct pathogenicity gene repertoire. *Appl Environ Microbiol*. 2009; 75:1011-20.

(148) Gong J, Zhang J, Xu M *et al*. Prevalence and fimbrial genotype distribution of poultry *Salmonella* isolates in China (2006 to 2012). *Appl Environ Microbiol*. 2014; 80:687-93.

(149) Löfström C, Hintzmann AS, Sørensen G, Baggesen DL. Outbreak of *Salmonella* enterica serovar Typhimurium phage type DT41 in Danish poultry production. *Vet Microbiol*. 2015; 178:167-72.

(150) Horton RA, Wu G, Speed K *et al*. Wild birds carry similar *Salmonella* enterica serovar Typhimurium strains to those found in domestic animals and livestock. *Res Vet Sci*. 2013; 95:45-48.

(151) Kornacki, J. Principles of Microbiological Troubleshooting in the Industrial Food Processing Environment (internet). Food Microbiology and Food Safety. Springer Science, 2010 (cited December 2014). 193p.
Available from:
https://link-springercom.ezproxy1.library.usyd.edu.au/book/10.1007%2F978-1-4419-551

(152) Jansen AM, Hall LJ, Clare S, Goulding D, Holt KE, Grant AJ. A *Salmonella* Typhimurium-Typhi genomic chimera: a model to study Vi polysaccharide capsule function in vivo. *PLoS Pathog*. 2011; 7(7): e1002131.

(153) Hedberg CW, Greenblatt JF, Matyas BT, Lemmings J, Sharp DJ, Skibicki RT *et al*. Timeliness of enteric disease surveillance in 6 US states. *Emerg Infect Dis*. 2008; 14(2):311-3.

(154) Hawkey J, Edwards DJ, Dimovski K, Hiley L, Billman-Jacobe H, Hogg G, Holt KE. Evidence of microevolution of *Salmonella* Typhimurium during a series of egg-associated outbreaks linked to a single chicken farm. *BMC Genomics*. 2013; 14:800. doi: 10.1186/1471-2164-14-800.

(155) Palmer, S, Bauerfeind, R. Oxford textbook of zoonoses: biology, clinical practice, and public health control (internet). Oxford textbooks in public health, 2011, 2nd ed. (cited 2013 November 3). Available from: http://opac.library.usyd.edu.au/record=3

(156) Hyeon JY, Chon JW, Park JH *et al*. A Comparison of Subtyping Methods for Differentiating *Salmonella* enterica Serovar Enteritidis Isolates Obtained from Food and Human Sources. *Osong Public Health Res Perspect*. 2013; 4(1):27-33.

(157) Bennett SD, Littrell KW, Hill TA, Mahovic M, *et al*. Multistate foodborne disease outbreaks associated with raw tomatoes, United States, 1990-2010: a recurring public health problem. *Epidemiol Infect*. 2015; 143(7):1352-59.

(158) Niemann J, Tietze E, Ruddat I *et al*. Epidemiological analysis of the dynamic and diversity of *Salmonella* spp. in five German pig production clusters using pheno- and genotyping methods: an exploratory study. *Vet Microbiol*. 2015; 176:190-95.

(159) Bhatnagar D. Microbial foodborne diseases: mechanisms of pathogenesis and toxin synthesis (internet). Lancaster, Pa. Technomic Pub. Co., Inc., 2000 (cited 2013 November 3). 550p. Available from: http://opac.library.usyd.edu.au/record=2395579

(160) Rabsch W. *Salmonella* enterica Serotype Typhimurium and Its Host-Adapted Variants. *Infect Immun*. 2002; 70(5):2249-55.

(161) Suez J, Porwollik S, Dagan A *et al*. Virulence gene profiling and pathogenicity characterization of non-typhoidal *Salmonella* accounted for invasive disease in humans. *PLoS One*. 2013;8: e58449.

(162) Fàbrega A, Vila J. *Salmonella* enterica serovar Typhimurium skills to succeed in the host: virulence and regulation. *Clin Microbiol Rev*. 2013; 26(2):308-41.

(163) Didelot X, Bowden R, Street T, Golubchik T, Spencer C, McVean G, *et al*. Recombination and Population Structure in *Salmonella enterica*. Casadesús J, editor. *PLoS Genet*. 2011; 7: e1002191.

(164) Hensel M. Evolution of pathogenicity islands of *Salmonella* enterica. *Int J Med Microbiol*. 2004; 294(2-3):95-102.

(165) Imami K, Bhavsar AP, Yu H *et al*. Global impact of *Salmonella* pathogenicity island 2-secreted effectors on the host phosphoproteome. Mol Cell Proteomics. 2013;12(6):1632-1643.

(166) Nieto PA, Pardo-Roa C, Salazar-Echegarai FJ *et al*. New insights about excisable pathogenicity islands in *Salmonella* and their contribution to virulence. *Microbes Infect*. 2016; 18(5):302-9.

(167) Che D, Hasan MS, Chen B. Identifying pathogenicity islands in bacterial pathogenomics using computational approaches. *Pathogens*. 2014; 3(1):36-56.

(168) Dieye Y, Ameiss K, Mellata M, Curtiss R. The *Salmonella* Pathogenicity Island (SPI) 1 contributes more than SPI2 to the colonization of the chicken by *Salmonella* enterica serovar Typhimurium. *BMC Microbiol*. 2009; 9:3. doi: 10.1186/1471-2180-9-3.

(169) Baily JL, Foster G, Brown D *et al*. *Salmonella* infection in grey seals (Halichoerus grypus), a marine mammal sentinel species: pathogenicity and molecular typing of *Salmonella* strains compared with human and livestock isolates. *Environ Microbiol*. 2016; 18(3):1078-87.

(171) Fournier P, Drancourt M, Raoult D. Bacterial genome sequencing and its use in infectious diseases. *Lancet Infect Dis* 2007; 7(11): 711-23

(172) Sabbagh SC, Forest CG, Lepage C, Leclerc JM, Daigle F. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella* enterica serovars Typhimurium and Typhi. *FEMS Microbiol Lett.* 2010; 305(1):1-13.

(173) Octavia S, Wang Q, Tanaka MM, Sintchenko V, Lan R. Genomic Variability of Serial Human Isolates of *Salmonella* enterica Serovar Typhimurium Associated with Prolonged Carriage. *J Clin Microbiol*. 2015; 53(11):3507-14.

(174) Reeves, M.W., Evins, G.M., Heiba, A.A., Plikaytis, B.D. and Farmer, J.J. III. Clonal nature of *Salmonella* Typhi and its genetic relatedness to other *Salmonella*e as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella* bongori comb.nov. *J Clin Microbiol*. 1989; 27(2), 313–20.

(175) Holt KE, Parkhill J, Mazzoni CJ, *et al*. High-throughput sequencing provides insights into genome variation and evolution in *Salmonella* Typhi. *Nat Genet* 2008; 40(8):987–93.

(176) Orsini M, Mangone I, DiPasquale A *et al*. Draft Genome Sequences of 19 *Salmonella* enterica Serovar Typhimurium [4,5:i:-] Strains Resistant to Nalidixic Acid from a Long-Term Outbreak in Italy. *Genome Announc*. 2015;3(4). pii: e00911-15. doi: 10.1128/genomeA.00911-15.

(177) Zhang S, Yin Y, Jones MB *et al*. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. *J Clin Microbiol*. 2015; 53(5):1685-92.

(178) Kjeldsen MK, Torpdahl M, Pedersen K, Nielsen EM. Development and comparison of a generic multiple-locus variable-number tandem repeat analysis with pulsed-field gel electrophoresis for typing of *Salmonella* enterica subsp. enterica. *J Appl Microbiol*. 2015; 119(6):1707-17.

.

(179) Dimovski K, Cao H, Wijburg OL *et al*. Analysis of *Salmonella* enterica serovar Typhimurium variable-number tandem-repeat data for public health investigation based on measured mutation rates and whole-genome sequence comparisons. *J Bacteriol*. 2014; 196(16):3036-44.

(180) Kruy SL, van Cuyck H, Koeck JL. Multilocus variable number tandem repeat analysis for *Salmonella* enterica subspecies. *Eur J Clin Microbiol Infect Dis.* 2011; 30(4):465-73.

(181) Ktari S, Ksibi B, Gharsallah H *et al*. Molecular epidemiological characteristics of *Salmonella* enterica serovars Enteritidis, Typhimurium and Livingstone strains isolated in a Tunisian university hospital. *APMIS*. 2016; 124(3):194-200.

(183) Lindstedt BA, Torpdahl M, Nielsen EM, Vardund T, Aas L, Kapperud G. Harmonization of the multiple-locus variable-number tandem repeat analysis method between Denmark and Norway for typing *Salmonella* Typhimurium isolates and closer examination of the VNTR loci. *J Appl Microbiol*. 2007; 102(3):728-35.

(184) Dimovski K, Cao H, Wijburg OL *et al*. Analysis of *Salmonella* enterica serovar Typhimurium variable-number tandem-repeat data for public health investigation based on measured mutation rates and whole-genome sequence comparisons. *J Bacteriol*. 2014; 196(16):3036-44.

(185) Larsson JT, Torpdahl M; MLVA working group., Møller Nielsen E. Proof-of-concept study for successful inter-laboratory comparison of MLVA results. *Euro Surveill*. 2013; 18(35):20566.

(186) Larsson JT, Torpdahl M; MLVA working group., Møller Nielsen E. Proof-of-concept study for successful inter-laboratory comparison of MLVA results. *Euro Surveill*. 2013; 18(35):20566.

(187) Chiou CS, Hung CS, Torpdahl M *et al*. Development and evaluation of multilocus variable number tandem repeat analysis for fine typing and phylogenetic analysis of *Salmonella* enterica serovar Typhimurium. *Int J Food Microbiol*. 2010; 142(1-2):67-73.

(188) Malorny B, Junker E, Helmuth R. Multi-locus variable-number tandem repeat analysis for outbreak studies of *Salmonella* enterica serotype Enteritidis. *BMC Microbiol*. 2008; 8:84. doi: 10.1186/1471-2180-8-84.

(189) Bertrand S, De Lamine de Bex G, Wildemauwe C *et al*. Multi locus variable-number tandem repeat (MLVA) typing tools improved the surveillance of *Salmonella* enteritidis: a 6 years retrospective study. *PLoS One*. 2015; 10: e0117950.

(190) Torpdahl M, Sørensen G, Lindstedt BA, Nielsen EM. Tandem repeat analysis for surveillance of human *Salmonella* Typhimurium infections. *Emerg Infect Dis*. 2007; 13(3):388-95.

(191) Guibert S, Weber M. Epigenetics: Erase for a new start. 2012; *Nature* 492, 363–364.

(192) Feng S, Jacobsen, SE Reik W. Epigenetic Reprogramming in Plant and Animal Development. *Science* 2010; (330):622–27.

(193) Litrup E, Torpdahl M, Malorny B, Huehn S, Christensen H, Nielsen EM. Association between phylogeny, virulence potential and serovars of *Salmonella* enterica. *Infect Genet Evol*. 2010; 10:1132-39.

(194) Kirsten J.M. van Nimwegen, Ronald A. van Soest,Joris A. Veltman, Marcel R. Nelen, Gert Jan van der Wilt, Lisenka E.L.M. Vissers, and Janneke P.C. Grutters. Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clin Chem* 2016; 62(11):1458-64.

(195) Elizabeth Robilotti, Mini Kamboj. Integration of Whole-Genome Sequencing into Infection Control Practices: the Potential and the Hurdles. *J Clin Microbiol* 2015; 53(4): 1054-55.

(195) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *PNAS*. 1977; 74(12):5463-7.

(196) Vincent A, Derome N, Boyle B, Culley A, Charette S. Next-generation sequencing (NGS) in the microbiological world: How to make the most of your money. *J Microbiol Methods*. 2017; 138:60-71. doi: 10.1016/j.mimet.2016.02.016.

(197) Kwong JC1, McCallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. *Pathology*. 2015; 47(3):199-210.

(198) Almeida F, Medeiros MI, Rodrigues DP *et al*. Draft Genome Sequences of 40 *Salmonella* enterica Serovar Typhimurium Strains Isolated from Humans and Food in Brazil. *Genome Announc*. 2016;4 (5): e00892-16.

(199) Chandra Shekhar Pareek & Rafal Smoczynski & Andrzej Tretyn. Sequencing technologies and genome sequencing. *J Appl Genetics* 2011; 52(4):413-35.

(200) Feasey N, Dougan G, Kingsley R, Heyderman R, Gordon N. Invasive non-typhoidal *Salmonella* disease: an emerging and neglected tropical disease in Africa. *Lancet* 2012; 379(9835): 2489–99.

(201) Earle SG, Wu C-H, *et al*. Identifying lineage effects when controlling for population structure improves power in bacterial association studies. *Nat Microbiol* 2016; 1:16041. doi: 10.1038/nmicrobiol.2016.41.

(202) Illumina. Nextera® DNA Library Prep Reference Guide. Illumina Inc. USA; January 2016. 36p.

(203) Illumina. Sequencing coverage calculation methods for human-whole genome sequencing. Illumina Inc. USA; 2014. 3p.

(204) Illumina. Quality scores for next generation sequencing. Illumine Inc. USA; 2011. 2p.

(205) Fricke WF, Rasko DA. Bacterial genome sequencing in the clinic: bioinformatic challenges and solutions. *Nat Rev Genet*. 2014; 15(1):49-55.

(206) Schlaberg R, Chiu CY, Miller S *et al*. Validation of Metagenomic Next-Generation Sequencing Tests for Universal Pathogen Detection. *Arch Pathol Lab Med*. 2017;141(6):776-86.

(207) Baily JL, Foster G, Brown D *et al*. *Salmonella* infection in grey seals (Halichoerus grypus), a marine mammal sentinel species: pathogenicity and molecular typing of *Salmonella* strains compared with human and livestock isolates. *Environ Microbiol*. 2016; 18(3):1078-87.

(208) Zankari E, Hasman H, Kaas RS, Seyfarth AM, Agersø Y, Lund O *et al*. Genotyping using whole-genome sequencing is a realistic alternative to surveillance based on phenotypic antimicrobial susceptibility testing. *J Antimicrob Chemother*. 2013; 68(4):771-7.

(209) Knobloch JK, Niemann S, Kohl TA, Lindner U, Nitschke M, Sayk F. Whole-genome sequencing for risk assessment of long-term Shiga toxin-producing Escherichia coli. *Emerg Infect Dis*. 2014; 20(4):732-3.

(210) van Nimwegen KJ, van Soest RA, Veltman JA, Nelen MR, van der Wilt GJ, Vissers LE *et al*. Is the $1000 Genome as Near as We Think? A Cost Analysis of Next-Generation Sequencing. *Clin Chem*. 2016; 62(11):1458-64.

(207) Joensen K. G., Tetzschner A. M., Iguchi A., Aarestrup F. M., Scheutz F. Rapid and easy in silico serotyping of Escherichia coli using whole genome sequencing (WGS) data.
J. Clin. Microbiol. 2015; (53): 2410–26.

(208) Fierer J, Guiney DG. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. *J Clin Invest*. 2001; 107(7):775-80.

(209) Connor TR, *et al*. What's in a name? Species-wide whole-genome sequencing resolves invasive and non-invasive lineages of *Salmonella* enterica ParaTyphi B. *mBio* 2016; 7(4):e00527-16.

(210) Gong J, Zhang J, Xu M *et al*. Prevalence and fimbrial genotype distribution of poultry *Salmonella* isolates in China (2006 to 2012). *Appl Environ Microbiol*. 2014; 80(2:687-93.

(211) Torres MI, Lewis P, Cook L, Cook P, Kardamanidis K, Shadbolt C. An outbreak of *Salmonella* Typhimurium linked to a kebab takeaway shop. *Commun Dis Intell Q Rep.* 2012; 36(1):101-6.

(212) Li B, Jackson SA, Gangiredla J, Wang W, Liu H, Tall BD *et al*. Genomic evidence reveals numerous *Salmonella* enterica serovar Newport reintroduction events in Suwannee watershed irrigation ponds. *Appl Environ Microbiol*. 2015; 81(24):8243-53.

(213) Kurosawa A, Imamura T, Tanaka K *et al*. Molecular typing of *Salmonella* enterica serotype Typhimurium and serotype 4,5,12:i:- isolates from cattle by multiple-locus variable-number tandem-repeats analysis. *Vet Microbiol*. 2012; 160:264-68.

(214) Mather AE, Lawson B, de Pinna E *et al*. Genomic Analysis of *Salmonella* enterica Serovar Typhimurium from Wild Passerines in England and Wales. *Appl Environ Microbiol*. 2016; 82(22):6728-35.

(215) Lee KE, Lee DY, Choi HW *et al*. Characterization of *Salmonella* enterica serovar Typhimurium and *Salmonella* enterica serovar 4,[5],12:i:- isolates from pigs presenting with diarrhoea in Korea. *J Vet Med Sci*. 2015; 77(11):1511-15.

(216) Fu S, Octavia S, Wang Q *et al*. Evolution of Variable Number Tandem Repeats and Its Relationship with Genomic Diversity in *Salmonella* Typhimurium. *Front Microbiol.* 2016; 7:2002. doi: 10.3389/fmicb.2016.02002.

(217) Ido N, Iwabuchi K, Sato'o Y *et al*. Molecular typing of *Salmonella* enterica serovar 4,[5],12:i:- isolates from humans, animals and river water in Japan by multilocus variable-number tandem repeat analysis and pulsed-field gel electrophoresis. *J Vet Med Sci*. 2015; 77(5):609-13.

(218) Ikumapayi UN, Antonio M, Sonne-Hansen J *et al*. Molecular epidemiology of community-acquired invasive non-typhoidal *Salmonella* among children aged 2 29 months in rural Gambia and discovery of a new serovar, *Salmonella* enterica Dingiri. *J Med Microbiol*. 2007; 56(Pt 11):1479-84.

(219) Fu S, Octavia S, Tanaka MM, Sintchenko V, Lan R. Defining the Core Genome of *Salmonella* enterica Serovar Typhimurium for Genomic Surveillance and Epidemiological Typing. *J Clin Microbiol*. 2015; 53(8):2530-38.

(220) Page AJ, Cummins CA, Hunt M *et al*. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015; 31(22):3691-93.

(221) Bhavsar AP, Brown NF, Stoepel J *et al*. The *Salmonella* type III effector SspH2 specifically exploits the NLR co-chaperone activity of SGT1 to subvert immunity. *PLoS Pathog*. 2013; 9:e1003518. doi: 10.1371/journal.ppat.1003518.

(222) Poh J, Odendall C, Spanos A *et al*. SteC is a *Salmonella* kinase required for SPI-2-dependent F-actin remodelling. *Cell Microbiol*. 2008; 10(1):20-30.

(223) Girgis HS, Liu Y, Ryu WS, Tavazoie S. A comprehensive genetic characterization of bacterial motility. *PLoS Genet*. 2007; 3(9):1644-60.

(224) Tewolde R, Dallman T, Schaefer U *et al*. MOST: a modified MLST typing tool based on short read sequencing. *PeerJ*. 2016; 4:e2308. doi: 10.7717/peerj.2308. eCollection 2016.

(225) Okoro CK, Barquist L, Connor TR *et al*. Signatures of adaptation in human invasive *Salmonella* Typhimurium ST313 populations from sub-Saharan Africa. *PLoS Negl Trop Dis.* 2015; 9:e0003611. https://doi.org/10.1371/journal.pntd.0003848.

(226) Okoro CK, Kingsley RA, Connor TR *et al*. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet.* 2012; 44(11):1215-21.

(227) Moffatt CR, Musto J, Pingault N *et al*. *Salmonella* Typhimurium and Outbreaks of Egg-Associated Disease in Australia, 2001 to 2011. *Foodborne Pathog Dis*. 2016; 13(7):379-85.

(228) Dyet KH, Turbitt E, Carter PE. Multiple-locus variable-number tandem-repeat analysis for discriminating within *Salmonella* enterica serovar Typhimurium definitive types and investigation of outbreaks. *Epidemiol Infect.* 2011; 139(7):1050-9.

(229) Lienemann T, Kyyhkynen A, Halkilahti J, Haukka K, Siitonen A. Characterization of *Salmonella* Typhimurium isolates from domestically acquired infections in Finland by phage typing, antimicrobial susceptibility testing, PFGE and MLVA. *BMC Microbiol.* 2015; 15:131. doi: 10.1186/s12866-015-0467-8.

(231) Van TT, Nguyen HN, Smooker PM, Coloe PJ. The antibiotic resistance characteristics of non-typhoidal *Salmonella* enterica isolated from food-producing animals, retail meat and humans in South East Asia. *Int J Food Microbiol*. 2012; 154(3):98-106.

(232) Van Boxstael S, Dierick K, Van Huffel X *et al*. Comparison of antimicrobial resistance patterns and phage types of *Salmonella* Typhimurium isolated from pigs, pork and humans in Belgium between 2001 and 2006. *Food Res Int*. 2012; 45:913-918.

(233) Torpdahl M, Lauderdale TL, Liang SY, Li I, Wei SH, Chiou CS. Human isolates of *Salmonella* enterica serovar Typhimurium from Taiwan displayed significantly higher levels of antimicrobial resistance than those from Denmark. *Int J Food Microbiol*. 2013; 161(2):69-75.

(234) European Food Safety Authority (EFSA) Annual Report. (Internet) (EU); 2010 (cited 2013 September 5). 68p. Available from: https://www.efsa.europa.eu/en/corporate/pub/ar10

(235) Chiou CS, Hung CS, Torpdahl M *et al*. Development and evaluation of multilocus variable number tandem repeat analysis for fine typing and phylogenetic analysis of *Salmonella* enterica serovar Typhimurium. *Int J Food Microbiol*. 2010; 142(1-2):67-73.

(236) Hopkins KL, Maguire C, Best E, Liebana E, Threlfall EJ. Stability of multiple-locus variable-number tandem repeats in *Salmonella* enterica serovar Typhimurium. *J Clin Microbiol.* 2007; 45(9):3058-61.

(237) Espinoza RA, Silva-Valenzuela CA, Amaya FA, Urrutia ÍM, Contreras I, Santiviago CA. Differential roles for pathogenicity islands SPI-13 and SPI-8 in the interaction of *Salmonella* Enteritidis and *Salmonella* Typhi with murine and human macrophages. *Biol Res*. 2017; 50(1):5. doi: 10.1186/s40659-017-0109-8.

(238) Gopinath G, Jean-Gilles Beaubrun J, Grim C *et al*. Whole-Genome Sequences of Six *Salmonella* enterica Serovar Bovismorbificans Isolates Associated with a 2011 Multistate Hummus-Borne Outbreak. *Genome Announc*. 2014; 2 (2): e01239-13. doi: 10.1128/genomeA.01239-13.

(239) Hughes LA, Wigley P, Bennett M, Chantrey J, Williams N. Multi-locus sequence typing of *Salmonella* enterica serovar Typhimurium isolates from wild birds in northern England suggests host-adapted strain. *Lett Appl Microbiol*. 2010; 51(4):477-79.

(240) Jarvik T, Smillie C, Groisman EA, Ochman H. Short-term signatures of evolutionary change in the *Salmonella* enterica serovar Typhimurium 14028 genome. *J Bacteriol.* 2010; 192(2):560-67.

(241) Mohammed M, Cormican M. Whole genome sequencing provides possible explanations for the difference in phage susceptibility among two *Salmonella* Typhimurium phage types (DT8 and DT30) associated with a single foodborne outbreak. *BMC Res Notes*. 2015; 8:728. doi: 10.1186/s13104-015-1687-6.

(242) Asmar BI, Abdel-Haq N. Nontyphoidal *Salmonella* infection in children: relation to bacteraemia, age, and infecting serotype. *Infect Dis*. 2016; 48(2):147-51.

(244) Søndberg E, Jelsbak L. *Salmonella* Typhimurium undergoes distinct genetic adaption during chronic infections of mice. *BMC Microbiol*. 2016; 16:30. https://doi.org/10.1186/s12866-016-0646-2.

(245) Zhu X, Lei H, Wu J, Li JV, Tang H, Wang Y. Systemic Responses of BALB/c Mice to Infection. *J Proteome Res*. 2014; 13(10):4436-45.

(247) Burns-Guydish SM, Olomu IN, Zhao H, Wong RJ, Stevenson DK, Contag CH. Monitoring age-related susceptibility of young mice to oral *Salmonella* enterica serovar Typhimurium infection using an in vivo murine model. *Pediatr Res*. 2005; 58(1):153-8.

(248) Lal A, Hales S, Kirk M, Baker MG, French NP. Spatial and temporal variation in the association between temperature and salmonellosis in NZ. *Aust N Z J Public Health*. 2016; 40(2):165-9.

(249) Buer J, Balling R. Mice, microbes and models of infection. *Nat Rev Genet*. 2003; 4(3):195-205.

(250) Coers J, Starnbach MN, Howard JC. Modelling infectious disease in mice: co-adaptation and the role of host-specific IFNgamma responses. *PLoS Pathog*. 2009; 5:e1000333. doi: 10.1371/journal.ppat.1000333. Epub 2009 May 29.

(251) Dietrich WF. Using mouse genetics to understand infectious disease pathogenesis. *Genome Res*. 2001; 11:325-31.

(252) Endt K, Maier L, Käppeli R *et al*. Perioral ciprofloxacin therapy impairs the generation of a protective immune response in a mouse model for *Salmonella* enterica serovar Typhimurium diarrhoea, while parenteral ceftriaxone therapy does not. *Antimicrob Agents Chemother*. 2012; 56(5):2295-2304.

(253) Legrand N, Ploss A, Balling R *et al*. Humanized mice for modelling human infectious disease: challenges, progress, and outlook. *Cell Host Microbe*. 2009; 6(1):5-9.

(254) Magiorakos AP, Srinivasan A, Carey RB *et al*. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect*. 2012; 18(3):268-81.

(255) Mathur R, Oh H, Zhang D *et al*. A mouse model of *Salmonella* Typhi infection. *Cell*. 2012; 151(3):590-602.

(256) Shukla G, Verma I, Sharma L. Effect of *Salmonella* enteric Serovar Typhimurium in Pregnant Mice: A Biochemical and Histopathological Study. *Gastroenterology Res*. 2012; 5(3):103-11.

(257) Strowig T, Flavell RA. Humanized mouse models of infectious diseases. *Drug Discov Today Dis Models*. 2012; 9:e11-e16.

(258) Popovic A, Wu B, Arrowsmith CH, Edwards AM, Davidson AR, Maxwell KL. Structural and biochemical characterization of phage λ FI protein (gpFI) reveals a novel mechanism of DNA packaging chaperone activity. *J Biol Chem*. 2012; 287(38):32085-95.

(259) Panthel K, Meinel KM, Domènech VE *et al*. *Salmonella* pathogenicity island 2-mediated overexpression of chimeric SspH2 proteins for simultaneous induction of antigen-specific CD4 and CD8 T cells. *Infect Immun*. 2005; 73:334-41.

(260) Barua H, Lindblom IL, Bisgaard M, Christensen JP, Olsen RH, Christensen H. In vitro and in vivo investigation on genomic stability of *Salmonella* enterica Typhimurium DT41 obtained from broiler breeders in Denmark. *Vet Microbiol*. 2013; 166(3-4):607-16.

(261) Marvasi M, Cox CE, Xu Y, Noel JT, Giovannoni JJ, Teplitski M. Differential regulation of *Salmonella* Typhimurium genes involved in O-antigen capsule production and their role in persistence within tomato fruit. *Mol Plant Microbe Interact*. 2013; 26(7):793-800.

(262) Murray GL, Attridge SR, Morona R. Altering the length of the lipopolysaccharide O antigen has an impact on the interaction of *Salmonella* enterica serovar Typhimurium with macrophages and complement. *J Bacteriol.* 2006; 188(7):2735-39.

(263) Miller KA, Phillips RS, Mrázek J, Hoover TR. *Salmonella* utilizes D-glucosaminate via a mannose family phosphotransferase system permease and associated enzymes. *J Bacteriol.* 2013; 195918):4057-66.

(264) Lam LH, Monack DM. Intraspecies competition for niches in the distal gut dictate transmission during persistent *Salmonella* infection. *PLoS Pathog.* 2014; 10:e1004527.

(265) Borziak K, Fleetwood AD, Zhulin IB. Chemoreceptor gene loss and acquisition via horizontal gene transfer in Escherichia coli. *J Bacteriol*. 2013; 195(16):3596-3602.

(266) Bourgogne A, Hilsenbeck SG, Dunny GM, Murray BE. Comparison of OG1RF and an isogenic fsrB deletion mutant by transcriptional analysis: the Fsr system of Enterococcus faecalis is more than the activator of gelatinase and serine protease. *J Bacteriol.* 2006; 188(8):2875-84.

(267) Godlewska R, Wiśniewska K, Pietras Z, Jagusztyn-Krynicka EK. Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol Lett.* 2009; 298(1):1-11.

(268) Kutsukake K, Nakashima H, Tominaga A, Abo T. Two DNA invertases contribute to flagellar phase variation in *Salmonella* enterica serovar Typhimurium strain LT2. *J Bacteriol.* 2006; 188(3):950-57.

(269) Lamichhane-Khadka R, Frye JG, Porwollik S, McClelland M, Maier RJ. Hydrogen-stimulated carbon acquisition and conservation in *Salmonella* enterica serovar Typhimurium. *J Bacteriol*. 2011; 193(20):5824-32.

(270) Huehn S, La Ragione RM, *et al*. Virulotyping and antimicrobial resistance typing of *Salmonella* enterica serovars relevant to human health in Europe. *Foodborne Pathog Dis* 2010; 7(5):523-35.

(271) Boers SA, van der Reijden WA, Jansen R. High-throughput multilocus sequence typing: bringing molecular typing to the next level. *PLoS One*. 2012; 7:e39630.

(272) Cox R, Su T, Clough H, Woodward MJ, Sherlock C. Spatial and temporal patterns in antimicrobial resistance of *Salmonella* Typhimurium in cattle in England and Wales. *Epidemiol Infect.* 2012; 140(11):2062-73.

(273) Yasuda T, Nagata T, Ohmori H. Multicopy suppressors of the cold-sensitive phenotype of the pcsA68 (dinD68) mutation in Escherichia coli. *J Bacteriol*. 1996; 178(13):3854-9.

(274) Lan R, Reeves PR, Octavia S. Population structure, origins and evolution of major *Salmonella* enterica clades. *Infect Genet Evol*. 2009; 9(5):996-1005.

(275) Löfström C, Hintzmann AS, Sørensen G, Baggesen DL. Outbreak of *Salmonella* enterica serovar Typhimurium phage type DT41 in Danish poultry production. *Vet Microbiol*. 2015; 178(1-2):167-72.

(276) Liu LG, Zhou XY, Lan Z *et al*. *Salmonella* Typhimurium outbreak associated with a contaminated food container in a school in Sichuan Province, China. *Epidemiol Infect.* 2016; 144(2):285-90.

(277) Heilbronn C, Munnoch S, Butler MT, Merritt TD, Durrheim DN. Timeliness of *Salmonella* Typhimurium notifications after the introduction of routine MLVA typing in NSW. *N S W Public Health Bull*. 2014; 24(4):159-63.

(278) Graziani C, Mughini-Gras L, Owczarek S, Dionisi AM, Luzzi I, Busani L. Distribution of *Salmonella* enterica isolates from human cases in Italy, 1980 to 2011. *Eurosurveillance.* 2013; 18(27):20519. https://doi.org/10.2807/1560-7917.ES2013.18.27.20519.

(279) Falay D, Kuijpers LM, Phoba MF *et al*. Microbiological, clinical and molecular findings of non-typhoidal *Salmonella* bloodstream infections associated with malaria, Oriental Province, Democratic Republic of the Congo. *BMC Infect Dis.* 2016; 16:271. doi: 10.1186/s12879-016-1604-1.

(280) Guo C, Hoekstra RM, Schroeder CM *et al*. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: adaptation of a Danish model. *Foodborne Pathog Dis*. 2011; 8(4):509-16.

(281) Bonifield HR, Hughes KT. Flagellar Phase Variation in *Salmonella* enterica Is Mediated by a Posttranscriptional Control Mechanism. *J Bacteriol*. 2003; 185(12):3567-74.

(282) Wong KK, McClelland M. Dissection of the *Salmonella* Typhimurium genome by use of a Tn5 derivative carrying rare restriction sites. *J Bacteriol*. 1992; 174(11):3807-11.

(283) Wickham ME, Brown NF, Provias J, Finlay BB, Coombes BK. Oral infection of mice with *Salmonella* enterica serovar Typhimurium causes meningitis and infection of the brain. *BMC Infect Dis*. 2007; 7:65.

Appendices

Appendix A. A sample of Roary original data file as output for the successful/non-successful MLVA - STM analysis (full file from Roary software is provided as .xls electronic attachment).

Gene	STM-01_S46 *3-9-7-12-523 (2010)*	STM-3S58 *3-9-7-12-523 (2011)*	STM-5S13 *3-9-7-12-523 (2012)*	STM-8S31 *3-9-7-12-523 (2013)*	STM-11_S50 *3-9-7-12-523 (2014)*	STM-14_S19 *3-9-7-12-523 (2015)*
garK_1	STM-01_S46_00812	STM-3S58_02370	STM-5S13_01334			STM-14_S19_00100
sspH2_2	STM-01_S46_04486	STM-3S58_01290				
phnT_1					STM-11_S50_04068	
group_28				STM-8S31_04704		
group_362					STM-11_S50_03743	
oadB_2						
group_112		STM-3S58_04038		STM-8S31_03246		
group_269			STM-5S13_00094			
iptC			STM-5S13_00092			
ccmH_1		STM-3S58_01975				STM-14_S19_00860
group_170			STM-5S13_02843			
ccmH_2		STM-3S58_03089				
group_121	STM-01_S46_01462					
atpD_2						STM-14_S19_02538
group_137				STM-8S31_04304		
group_139	STM-01_S46_02620					
group_140						STM-14_S19_03666
hsrA_1		STM-3S58_00180				
pntB_1		STM-3S58_01372				
ydiN_1	STM-01_S46_00003					
group_1563	STM-01_S46_00166					
corA_2	STM-01_S46_02890					
group_1590	STM-01_S46_03286					
group_1592	STM-01_S46_03363					
group_1599	STM-01_S46_04256					
mhbT_2	STM-01_S46_04305					
argF_2	STM-01_S46_04478					
group_169					STM-11_S50_03092	
rhaA_1					STM-11_S50_02163	

Appendix B. Identification and characterization of SNPs in the STM isolates outbreak

	Reference nucleotide and genome position in *S*. Typhimurium LT 2 (GenBank Accession No. NC_003197)																																
	740808	1850858	3045820	6136220	7233663	7516737	0417198	3671974	4554773	6140329	6116298	6194081	7897881	8660600	9088597	0133636	0195601	0411404	1622287	3362259	4139080	4917755	5967274	5988383	7913696	9833630	0980816	2733259	5176425	6577045	3325558		
Isolate	T^	G	C	C	C	G	G	C	C	C	C	C	A	C	G	C	G	G	A	G	G	C	T	A	G	G	G	C	C	G	ATTTTAT		**Epi-confirmed source**
5_A4078	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
6_A4103	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
7_A2128	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
8_A2097	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
9_A2096	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
10_A2099	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
11_A2137	A.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
14_A1970	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
15_A1974	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
16_A1954	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
17_A1965	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
18_A2474	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
19_A4521	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
20_A2475a	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
21_A2472	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
23_A2465	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
24_A2471	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
25_A4615	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
26_A4611	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
27_A4619	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
28_A4613	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
30_A4318	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (C)
33_S4804	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (SC)
56_E4809	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (E)
57_E4810	A.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Outbreak A (E)
38_M2261	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
39_M2258	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
40_M2268	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
41_M2107	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
42_M5132	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	–	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
43_M4657	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
44_M4945	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
45_M4703	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
46_M4704	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
48_M4732	.	.	.	.	T	.	.	.	.	.	.	.	*	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
50_M2475b	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
51_M5033	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (C)
58_E4692	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
59_E4696	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
60_E4720	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
61_E4721	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	T	.	C	.	.	–		Outbreak M (E)
62_E4742	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
63_E4743	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
64_E4499	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
65_E2330	.	.	.	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	–		Outbreak M (E)
54_N4956	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	–		Sporadic
1_N4472	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
22_N2054	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
31_N4583	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
32_N5047	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
2_N4246	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
35_N4800	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	A	.	.	.	.	.	C	.	.	.		Sporadic
4_N4210	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		Sporadic
37_N4954	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.		Sporadic
67_N4799	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	–		Sporadic
3_N2463	.	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	A	.	.	.	.	.	C	.	.	.	.	.	A	–		Sporadic
29_N4955	.	.	.	.	.	.	A	T	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	–		Sporadic
34_N4802	.	.	.	.	.	.	A	T	T	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	A	–		Sporadic
36_N5175	.	.	.	.	.	.	A	T	T	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	A	–		Sporadic
52_N5064	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.	.	A	.	.	.		Sporadic
53_N5139a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	C	.	.	A	.	.	.		Sporadic
55_N4714	.	.	A	.	.	T	.	.	.	T	T	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	T	.		Sporadic