TRANSMISSION AND EVOLUTION OF *SALMONELLA*

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yesim Soyer

August 2009

© 2009 Yesim Soyer

TRANSMISSION AND EVOLUTION OF *SALMONELLA*

Yesim Soyer, Ph.D.

Cornell University 2009

Salmonella, a Gram-negative enterobacterium, is a human and animal pathogen. The species *Salmonella enterica* includes serotypes with broad and restricted host ranges. The objectives of this study were to utilize genotypic and phenotypic methods to (i) understand *Salmonella* diversity among human and animal hosts at the population level, and (ii) understand *Salmonella* genome evolution including the mechanisms of diversity and their contributions to new serotypes.

To understand genotypic and phenotypic diversity in *Salmonella* among human and animal hosts, we used serotyping, multilocus sequence typing (MLST) and pulsed field gel electrophoresis (PFGE) subtyping methods. We found that while some subtypes might have represented host adapted strains within the same serotype, some subtypes were widely distributed among human and cattle isolates. We concluded that isolation of common PFGE types among humans and foods or farm animals must be interpreted carefully and that the establishment of causal relationships will require strong epidemiological linkages and/or the use of additional, more sensitive, subtyping methods.

To understand which mechanisms are responsible for diversity in the *Salmonella* genome, we used 5 whole genome sequences representing four serotypes, Typhi, Paratyphi A, Choleraesuis, and Typhimurium, to identify genes with evidence of positive selection and recombination. Positive selection was detected using PAML 3.15. Intragenic recombination was assessed by four different approaches: GENCONV, Max-$\chi 2$, NSS, and PHI. We found that genes having evidence of

recombination may be more likely to be under positive selection. Positive selection may contribute to fixation of new allelic variants generated by recombination.

To understand how new *Salmonella* serotypes emerge, we characterized an emerging serotype, *Salmonella* 4,5,12:i:-, which is closely related to Typhimurium but lacks the expression of second phase flagellar antigen. We characterized *Salmonella* 4,5,12:i:- and *Salmonella* Typhimurium isolates from various sources using MLST, PFGE and PCR screening for differences in presence or absence of genes or intergenic regions. We found that while the majority of 4,5,12:i:- and Typhimurium isolates represented a single MLST type, all 4,5,12:i:- lacked *fljA* and *fljB*, which were present in all Typhimurium isolates. PCR screens further showed differences in deletion genotypes among 4,5,12:i:- strains which suggests that *Salmonella* 4,5,12:i:- appears to represent two different emergence events, both from a serotype Typhimurium ancestor.

Overall, my research gives us insight into the evolution of *Salmonella* that allows us to better understand the transmission and the evolution of *Salmonella* and its ability to cause disease in different host species. Consequently, we will be able to track contamination sources using much more specific subtyping methods in order to eliminate *Salmonella* from food products.

BIOGRAPHICAL SKETCH

Yesim Soyer was born in July 31, 1977 in Ankara, Turkey. Yesim received a Food Engineering degree from Ankara University Food Engineering Department, Turkey in June, 1999. Yesim decided to pursue a Master of Science degree in Food Engineering under Dr. Feryal Karadeniz. Yesim completed her Masters degree with a thesis entitled "Determination of Organic Acid Content of White Grapes and Grape juices" in June, 2001 and began a Ph.D. program in Food Engineering under the guidance of Dr. Feryal Karadeniz. In January, 2004, Yesim received a scholarship from the Higher Education System in Turkey to pursue a Ph.D. in the US. In August, 2004, Yesim began a Ph.D. program in Food Science with minors in Microbiology and Epidemiology under Drs. Martin Wiedmann, Kathryn J. Boor and Yrjö T. Gröhn. Yesim will start a position as a postdoctoral research associate in Dr. Wiedmann's laboratory in May, 2009.

To Mom and Dad, Asiye and Bilgin Soyer

ACKNOWLEDGMENTS

I would like to thank my major advisor, Dr. Martin Wiedmann, for accepting me in into his laboratory as a graduate student. Though I did not have a microbiology background, he taught me a lot about the field. Thank you for being the best advisor during my Ph.D. pursuit. I have learned a lot from him during my Ph.D. studies. Martin has always challenged me during every step of my Ph.D. I truly appreciate his guidance in my education, as well as in my professional development. I was very lucky to have him as an adviser in my Ph.D. studies, as well as in my long term plans. I would also thank Dr. Kathryn J. Boor for being my minor adviser and being a very good role model for women scientist. I am very happy to be a member of Food Safety Laboratory (FSL). I will still have a lot to learn from both of them

I also would like to thank my minor adviser, Dr. Yrjö Gröhn for being a great minor adviser, as well as his family for opening their house doors to me in my first years in Ithaca, when I was very homesick. I have also enjoyed working with Dr. Gröhn, as well as Dr. Lorin Warnick in Zoonotic Research Unit project. While Dr. Lorin Warnick was not my official adviser, he has always supported my research and has found time to answer my questions. I really appreciate his supports in my studies.

For the last five years, members of the FSL have become my family in Ithaca. I will be happy to revisit my memories from FSL in future. I have been very happy to work with every member of this laboratory both in past and in present. I would like to start by thanking Haley Oliver, Esther Fortes, and Barbara Bowen for their technical as well as emotional supports. I can not imagine FSL without them. It was very fun to work with Renato Orsi, Andrea Moreno, Sam Alcaine, Sherry Roof, Emily Wright, Lorraine Rodriguez and Beth Licitra, thanks all. I thank present and also previous members of FSL, Yuwei Hu, Soraya Chaturongakul, Liz Palmer, Shelley Laue, Maria

Teresa da Silva Felicio, Sarita Raengpradub, Kendra Nightingale, Yvonne Chan, Teresa Bergholz, Henk den Bakker, Maureen Gunderson, Sana Mujahid, Kitiya Vongkamjan, Vania Ferreira, and Nancy Carey for their friendship throughout my studies at Cornell University.

Special thanks go to my family. I could not have managed to finish my Ph.D. without their love and encouragement. I thank Mom and Dad, who always believe that I can manage to do everything and support me gratuitously at every level of my life. I thank Canan, Aslı and Dilek for being the best sisters and best friends. Life would be very miserable without you guys. The last, but not the least, I also would like to thank my nephew, Arda, for being my life spark

TABLE OF CONTENTS

Biographical Sketch		iii
Dedication		iv
Acknowledgements		v
List of Figures		viii
List of Tables		x
Chapter One	Introduction	1
Chapter Two	Pulsed Field Gel Electrophoresis Can Be Used to Distinguish Host Adapted and Un-restricted Subtypes among *Salmonella enterica* subsp. *enterica* Serotypes	18
Chapter Three	Genome wide evolutionary analyses reveal serotype specific patterns of positive selection in selected *Salmonella* serotypes	52
Chapter Four	*Salmonella* 4,5,12:i:-: an emerging *Salmonella* serotype that represents multiple distinct clones	99
Chapter Five	Conclusion	138

LIST OF FIGURES

Figure 2.1	PFGE type 94 is an example of PFGE types shared by different serotypes. PFGE type 94 was determined in one bovine *Salmonella* Typhimurium (4,5,12:i:1,2) isolate and two bovine *Salmonella* 4,5,12:i:- isolates	28
Figure 2.2	PFGE type 157 was differentiated in ST43 and ST62 in two serotypes, serotype 1,7:-:1,5 and serotype Thompson	29
Figure 2.3	PFGE types, 60, 89, and 126 obtained from both bovine and human clinical cases	31
Figure 2.4	Spatial analysis of bovine *Salmonella* isolates	38
Figure 3.1	Example of neighbor joining tree used for positive selection Analysis	59
Figure 3.2	Proportions of genes with evidence of recombination among individual JCVI role categories	68
Figure 4.1	Deduced genome structure for the genomic region between STM2691 and STM2774 for (A) four *Salmonella* 4,5,12:i:- isolates from Spain (based on genomic microarray data reported by Garaizar et al., 2002) and (B) the US serotype 4,5,12:i:- isolate CVM 23701 (based on an unfinished genome sequence reported by Rosovitz et al [http://msc.jcvi.org/salmonella/salmonella_enterica_subsp__enterica_serovar_4__5__12_i___str__cvm23701/index.shtml])	110
Figure 4.2	Phylogenetic tree for all 3-gene (*fimA, manB,* and *mdh*) STs identified among 4,5,12:i:- and Typhimurium isolates as well as selected isolates representing other *Salmonella* serotypes	114

Figure 4.3 Representative XbaI PFGE patterns for *Salmonella* 4,5,12:i:- and Typhimurium isolates as well as four isolates with inconsistent serotype data 119

LIST OF TABLES

Table 2.1	*Salmonella* isolate information for clinical bovine isolates collected from farms at multiple visits	24
Table 2.2	Distribution of subtypes found using serotyping, MLST, PFGE typing as well as combinations among clinical bovine and human *Salmonella* isolates and corresponding Simpson's Index of diversity scores for each subtyping method as well as combined methods	27
Table 2.3	Distribution of sequence types and PFGE types among common *Salmonella* serotypes	30
Table 2.4	PFGE patterns shared by *Salmonella* isolates within closely related serotypes	30
Table 2.5	Distribution of PFGE patters that were determined more than four times among clinical human and bovine *Salmonella* isolates and *P*-values representing the association between host species and PFGE patterns	34
Table 2.6	*Salmonella* human isolates that might have been collected from human clinical clusters within a county in two month duration	36
Table 3.1	*Salmonella* genomes used	58
Table 3.2	Genes used to confirm positive selection and recombination patterns identified in genome wide analyses	63
Table 3.3	Genes that show evidence of recombination in all four tests	66
Table 3.4	Genes with evidence for positive selection	71

Table 3.5	*Salmonella* pathogenicity island (SPI) genes with evidence of positive selection and recombination	76
Table 3.6	Evidence of recombination among genes with evidence for positive selection	77
Table 4.1	*Salmonella* 4,5,12:i:- and Typhimurium isolates used	103
Table 4.2	PCR conditions and primers for six genes and one region that show variable presence among serotype 4,5,12:i:- isolates from the US and Spain	108
Table 4.3	Distribution of 4-Gene Sequence Types (STs) among *Salmonella* isolates	116
Table 4.4	Presence/absence of selected genes in isolates Representing the Spanish and US *Salmonella* 4,5,12:i:- clones as well as other *Salmonella* isolates	127
Table A1 [S2.1]	A total of 335 spatially and temporally matched *Salmonella enterica* spp. e*nterica* nontyphoidal isolates used and their characteristics	142
Table A2 [S4.1]	A total of 190*Salmonella* 4,5,12:i:- and Typhimurium isolates used and their characteristics	149
Table A2 [S4.2]	Primers and PCR conditions	156

CHAPTER 1

INTRODUCTION

Salmonella, a Gram-negative enterobacterium, is a ubiquitous human and animal pathogen that can cause a wide range of diseases from self-limiting gastroenteritis to severe systemic infection in humans and animals. The genus *Salmonella* is divided into two species: *Salmonella bongori* and *Salmonella enterica*. *Salmonella enterica* consists of six subspecies: *S. enterica* subsp. *enterica, S. enterica* subsp. *salamae*, *S. enterica* subsp. *arizonae, S. enterica* subsp. *diarizonae, S. enterica* subsp. *houtenae,* and *S. enterica* subsp. *indica* (10). Among *Salmonella enterica* subspecies, *Salmonella enterica* subsp. *enterica* is the most important one, since serotypes of this subspecies are associated with warm blooded vertebrates and are responsible for the vast majority of salmonellosis infections (10, 51).

Salmonella enterica subsp. *enterica* serotypes are mainly transmitted by ingestion of food, feed, or water contaminated with feces from infected humans or animals (41), but can also be transmitted by direct contact (25, 34). The symptoms of nontyphoid *Salmonella* infection (salmonellosis) include fever, abdominal pain, diarrhea, nausea, and sometimes vomiting, and appear 12-72 hours after infection. The salmonellosis lasts 4-7 days and healthy adults can recover without any treatment. However, in infants, and the elderly, as well as in immunocopromised adults, the bacteria can enter the bloodstream and cause septicemia in severe salmonellosis cases (51) Another disease caused by certain *Salmonella* serotypes in certain host species is typhoid fever. For example, in humans, typhoid fever is caused by *Salmonella* Typhi. The symptoms of typhoid fever are high fever, malaise, headache, and diarrhea, rose-colored spots on the chest, and enlarged spleen and liver after 1-3 weeks exposure to *Salmonella* Typhi (51). Although typhoid fever is not common in developed countries such as the US, it is still a major problem in developing countries (56).

Salmonellae are among the most common reported bacterial foodborne pathogens worldwide (46). In the United States, nontyphoidal *Salmonella* serotypes have been estimated to cause >1.4 million human salmonellosis cases with >16,000 hospitalizations and almost 600 deaths annually (31). 95 % of salmonellosis is associated with consumption of contaminated food and water in the US (31). *Salmonella* Typhimurium, which causes predominantly self-limiting gastroenteritis in a large number of animal species, is the most common *Salmonella* serotype in the US (12, 51). According to data from the Foodborne Diseases Active Surveillance Network (FoodNet), the estimated incidence of several foodborne infections, including *Campylobacter, Cryptosporidium,* Shiga-toxin producing *Escherichia coli* O157 (STEC O157), *Listeria* and *Yersinia,* decreased 29-49 % in the United States between 1996-98 and 2005, while *Salmonella* infections declined only by 9 % (46). According to the surveillance research conducted by FoodNet, a division of the Centers for Disease Control and Prevention (CDC)'s Emerging Infections Program, in 10 states in the US in 2007, the incidence of *Salmonella* did not significantly change and the incidence of human cases due to *Salmonella* was the highest among known foodborne pathogens. Among 17,883 reported laboratory-confirmed cases in FoodNet surveillance population (45.5 million people) in 2007, 6,790 cases were due to salmonellosis. It was also reported that the incidence of salmonellosis (i.e., 14.92 per 100,000 population) was the furthest from its national health target, which is 6.80 per 100,000 population, among known foodborne bacterial pathogens (i.e., *Camplyobacter, Listeria,* STEC O157, *Shigella, Vibrio* and *Yersinia*) (11) (http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5714a2.htm).

Similarly in the European Union (EU), *Salmonella* is among the top three commonly known source of foodborne bacterial infections in humans. In 2006, *Salmonella* caused 34.6 human salmonellosis cases per 100,000 population (160,649

confirmed cases) in EU (19), available at http://www.efsa.europa.eu/cs/BlobServer/DocumentSet/Zoon_report_2006_en,0.pdf?ssbinary=true).

Generally, the sources of salmonellosis are meat and poultry related products, as well as fresh vegetables contaminated with feces. In the EU, the major source of foodborne salmonellosis is poultry related food products, followed by pork, while in the US, besides meat, eggs, and their products, fecal contaminated fresh vegetables, as well as nuts, are also the common vehicles of *Salmonella* for humans (12, 40).

Within different countries, the distribution of *Salmonella* serotypes causing human salmonellosis is different. For example, *Salmonella* Typhimurium is the most common serotype causing human salmonellosis in the US (12), while it is the second most common serotype, following *Salmonella* Enteritidis in EU (19). In Asia, although *Salmonella* Typhimurium and Enteritidis are common serotypes among human salmonellosis cases, *Salmonella* Weltevreden, a rare serotype in the US, is one of the most frequently isolated serotypes from human salmonellosis cases (7, 50).

In the US, the estimated economic burden for *Salmonella* human cases was reported as $ 2.8 billion (95% CI: $1.6 to $5.3 billion) annually (1). This amount represents the estimated cost of hospitalization due to *Salmonella* infection, and does not include any economic loss due to recalls in food industry (1).

The characterization of *Salmonellae* beyond species and subspecies is performed by subtyping methods. Serotyping, a phenotype-based subtyping method, with the Kaufmann-White scheme, has been commonly used as a first step to differentiate *Salmonella* isolates since 1968 (10, 53). This technique relies upon the immunoreactivity of *Salmonella*'s lipopolysaccaride moieties on cell surface (O antigens) and the flagellar proteins (H antigens), as well as capsular protein antigen (Vi-antigen), which is only found in a few *Salmonella* serotypes (e.g., Typhi).

According to the Kaufmann-White scheme, *Salmonella* includes over 2,500 recognized serotypes (23). Many *Salmonella* are motile due to peritrichous flagella (32), which include a basal body, a propeller and a hook. The motility of *Salmonella* depends on the rotation of the flagellar propeller (i.e., the filament), which includes either FliC (phase-1 antigen) or FljB (phase-2 antigen) flagellin (15). Most of *Salmonella* serotypes, including *Salmonella* Typhimurium, are bi-phasic, meaning that they can express two distinct flagellar antigens (i.e., phase-1 and phase-2 antigens). Regulation of phase 1 and 2 antigen expression is under control of the recombinase Hin (4).

Since serotyping of *Salmonella* cannot provide sufficient discrimination to track and find the source of outbreaks, modern, rapid and standardized molecular techniques have been developed to subdivide the *Salmonella* isolates within a serotype, such as phage typing (52, 54), DNA sequencing-based subtypes (i.e., random amplification of polymorphic DNA [RAPID] [24, 28], multilocus sequence typing [MLST] [3, 44], multiple-locus variable-number tandem repeat analysis [MLVA] [16]), ribotyping (20), pulsed field gel electrophoresis (PFGE) (52, 55), and microarray (22, 42). For example, phage typing is another traditional commonly used phenotypic subtypic method for *Salmonella*, in which a standardized set of phages is used to differentiate *Salmonella* isolates based on their susceptibility to be lysed by different bacteriophages (53, 54). While a standardized set of phages is used between laboratories to phage type *Salmonella*, the results from different laboratories, even within a laboratory, might show high biological and experimental variation (53). MLST is one of the genetic subtyping methods that uses DNA sequences of mainly multiple housekeeping genes to differentiate the isolates (29). Since the *Salmonella* genome is highly clonal, virulence genes are also used in the MLST scheme (3, 44).

Ribotyping and PFGE are also DNA-subtyping methods that include whole genome digestion by restriction enzymes. Ribotyping uses small digested fragments of DNA (approximately 1-30 kb) are separated on an agarose gel by electrophoresis and then the patterns of DNA fragments of ribosomal RNA (rRNA) genes, which are determined by Southern blot, are compared to differentiate the isolates (53). Alternatively in PFGE typing, the whole genome is digested into 8 to 25 large DNA bands, and then all the DNA fragments are separated on an agarose gel through alternating electric fields (53).

Since PFGE has become the gold standard for bacterial pathogens, including *Salmonella,* subtyping during the past ten years due to its broad applicability, high discriminatory power and epidemiological concordance (8), PFGE is used by CDC and state health departments in the United States (45), as well as in Canada, Latin America, Asia and Europe (EU) (47). The Centers for Disease Control and Prevention and state health departments in the United States developed a national network (PulseNet) to rapidly exchange standardized PFGE subtype data for isolates of foodborne pathogens (45, 47). PulseNet uses the molecular subtyping of isolates that has been commonly used in epidemiological investigations of food borne disease outbreaks and has become an essential component of epidemiologic investigations of infectious diseases (45).

Serotypes of *Salmonella enterica* subsp. *enterica* can be divided in three subdivisions (i.e., host restricted, host adapted and unrestricted serotypes) according to the abilities of serotypes to cause disease in different host species. Host restricted *Salmonella* serotypes are exclusively associated with one particular host (e.g. *Salmonella* Typhi, and *Salmonella* Paratyphi A), while host adapted *Salmonella* serotypes are prevalent in one particular host species, but are able to cause disease in other host species (e.g. *Salmonella* Choleraesuis). Unrestricted *Salmonella* serotypes,

although capable of causing systemic disease in a wide range of host species, usually induce self-limiting gastroenteritis, e.g. *Salmonella* Typhimurium (51). Molecular subtyping data suggest that some unrestricted serotypes might have a narrower host species range (40). For example, *Salmonella* Typhimurium phage type DT2 and DT99 were reported as poultry associated strains, while *Salmonella* Typhimurium phage type DT104 is a broad range variant. Therefore, *Salmonella* unrestricted serotypes might be collections of strains that vary significantly in their host range and their degree of host adaptation. These strains can be determined by molecular subtyping methods (40).

The databases including *Salmonella* isolates, especially with unrestricted serotypes, from different sources (i.e., cattle, human, poultry, food, etc.) with easily comparable phenotypic and genotypic subtypes, such as serotypes, DNA-sequencing based types, ribotypes, and PFGE types, as well as isolate information (i.e., isolate source, isolation date, isolated from), around the world are essential for the faster detection of the outbreaks and the faster tracking of the outbreak sources.

In the evolution of bacteria, gene acquisition and deletion events clearly play an important role. The importance of acquiring of novel (non-homologous) genes by lateral gene transfer has been clearly demonstrated in a number of bacteria, including a number of bacterial pathogens (21, 27, 36, 39). Acquisition of pathogenicity islands has played a critical role in the evolution of *Salmonella* (39) and other Gram-negative and Gram-positive pathogens (43). Gene degradation and gene deletions also have been shown to play a critical role in bacterial evolution, particularly when organisms with a broad niche specificity adapt to narrow and specific ecological niches (30, 49). For example, it has been suggested that gene degradation and gene deletion contribute to host adaptation in both *Salmonella* Typhi and *Salmonella* Paratyphi A (30). Phylogenetic analyses suggest that *Salmonella* and *Escherichia coli* shared a common

ancestor and diverged from that common ancestor about 100 million years ago (6). After the divergence from *E. coli*, *Salmonella* gained virulence associated gene complexes, commonly called "Salmonella Pathogenicity Island (SPI)" (39). One study (13) evaluated 410 genes present in both *S. enterica* and *E. coli*, and reported that 50% of amino acid substitutions in these genes appear to have been fixed by positive selection in one of these species. Microarray technologies have allowed for rapid and large scale studies on gene presence/absence in large numbers of isolates, including in *Salmonella* (38). In addition to gene acquisition and deletion, positive selection and homologous recombination, play important roles in the evolution of bacteria and bacterial pathogens (14, 26, 35).

Genome wide studies on positive selection and recombination in bacterial pathogens, including *Streptococcus* spp. (26), *Listeria monocytogenes* (35), *E. coli* (14, 37), and *Shigella* (37) have contributed to a better understanding of the evolution of these important pathogens.

Recently emerged *Salmonella* serotype 4,5,12:i:-, which appears to be closely related to *Salmonella* Typhimurium (which has the serotype 4,5,12:i:1,2), but lacks the expression of second phase 1, 2 flagellar antigen. *Salmonella* 4,5,12:i:- was the 6th most common *Salmonella* serotype among human cases in the US in 2006 (12) and the 4th most common serotype among human isolates in Spain in 1998 (22). Overall, the prevalence of *Salmonella* 4,5,12:i:- among human cases has increased considerably in many countries in the world over the last 10 years (9, 12, 22, 33, 48). This *Salmonella* serotype has also been responsible for a number of human salmonellosis outbreaks over the last decades, including in Spain (1998), the US (2004 and 2007), and in Luxemburg (2006). *Salmonella* 4,5,12:i:- has also been isolated, particularly over the last decade, from a number of different foods and animals (2, 5, 17, 33, 57). A number of separate studies, using molecular subtyping and characterization tools (e.g.,

genomic microarrays, PCR assays to test for gene presence/absence), have shown that serotype 4,5,12:i:- isolates from Spain (18, 22) and the US (2, 3, 57) are closely related to *Salmonella* Typhimurium.

Salmonella causes one of the most widely known bacterial food borne diseases worldwide, and is still a public health concern, as well as a problem causing huge economic loss in the food industry. Research is required to better understand the transmission and evolution of *Salmonella* and its ability to cause disease in different host species, in order for us to track the contamination sources using highly specific subtyping methods, and eliminate *Salmonella* from food products.

REFERENCES

1. **Adhikari, B., F. Angulo, and M. Meltzer.** 2004. Economic Burden of *Salmonella* Infections in the United States. 2004 Annual meeting from American Agricultural Economics Association (New Name 2008: Agricultural and Applied Economics Association). Denver, CO

2. **Agasan, A., J. Kornblum, G. Williams, C. C. Pratt, P. Fleckenstein, M. Wong, and A. Ramon.** 2002. Profile of *Salmonella enterica* subsp. *enterica* (subspecies I) serotype 4,5,12:i:- strains causing food-borne infections in New York City. J. Clin. Microbiol. **40:**1924-1929.

3. **Alcaine, S. D., Y. Soyer, L. D. Warnick, W. L. Su, S. Sukhnanand, J. Richards, E. D. Fortes, P. McDonough, T. P. Root, N. B. Dumas, Y. Grohn, and M. Wiedmann.** 2006. Multilocus Sequence Typing Supports the Hypothesis that Cow- and Human-Associated Salmonella Isolates Represent Distinct and Overlapping Populations. Appl. Environ. Microbiol. **72:**7575-7585.

4. **Aldridge, P. D., C. Wu, J. Gnerer, J. E. Karlinsey, K. T. Hughes, and M. S. Sachs.** 2006. Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. Proc. Natl. Acad. Sci. U.S.A. **103:**11340-11345.

5. **Amavisit, P., W. Boonyawiwat, and A. Bangtrakulnont.** 2005. Characterization of *Salmonella enterica* serovar Typhimurium and monophasic *Salmonella* serovar 1,4,[5],12:i:- isolates in Thailand. J. Clin. Microbiol. **43:**2736-2740.

6. **Baker, S., and G. Dougan.** 2007. The Genome of *Salmonella enterica* Serovar Typhi. Clin. Infect. Dis. **45:**S29-S33.

7. **Bangtrakulnonth, A., S. Pornreongwong, C. Pulsrikarn, P. Sawanpanyalert, R. S. Hendriksen, D. M. Lo Fo Wong, and F. M. Aarestrup.** 2004. *Salmonella* serovars from humans and other sources in Thailand, 1993-2002. Emerg. Infect. Dis. **10:**131-136.

8. **Barrett, T. J., P. Gerner-Smidt, and B. Swaminathan.** 2006. Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. Foodborne Pathog. Dis. **3:**20-31.

9. **Boonmar, S., A. Bangtrakulnonth, S. Pornruangwong, S. Samosornsuk, K. Kaneko, and M. Ogawa.** 1998. Significant increase in antibiotic resistance of *Salmonella* isolates from human beings and chicken meat in Thailand. Vet. Microbiol. **62:**73-80.

10. **Brenner, F. W., R. G. Villar, F. J. Angulo, R. Tauxe, and B. Swaminathan.** 2000. *Salmonella* nomenclature - Guest commentary. J. Clin. Microbiol. **38:**2465-2467.

11. **CDC.** 2008. Preliminary FoodNet Data on the Incidence of Infection with Pathogens Transmitted Commonly Through Food --- 10 States, 2007. MMWR **57:**366-370.

12. **CDC.** 2008. *Salmonella* Surveillance: Annual Summary, 2006, vol. Atlanta, Georgia: US Department of Health and Human Services, CDC.

13. **Charlesworth, J., and A. Eyre-Walker.** 2006. The rate of adaptive evolution in enteric bacteria. Mol. Biol. Evol. **23:**1348-56.

14. **Chen, S. L., C. S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D. Blasiar, T. Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J. P. Latreille, J. Spieth, T. M. Hooton, E. R. Mardis, S. J. Hultgren, and J. I. Gordon.** 2006. Identification of genes subject to positive selection in

uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U.S.A. **103:**5977-5982.

15. **Chevance, F. F., and K. T. Hughes.** 2008. Coordinating assembly of a bacterial macromolecular machine. Nat. Rev. Microbiol. **6:**455-465.

16. **Cho, S., T. S. Whittam, D. J. Boxrud, J. M. Bartkus, and A. M. Saeed.** 2008. Allele distribution and genetic diversity of VNTR loci in *Salmonella enterica* serotype Enteritidis isolates from different sources. BMC Microbiol. **8:**146-157.

17. **de laTorre, E., D. Zapata, M. Tello, W. Mejia, N. Frias, F. J. Garcia Pena, E. M. Mateu, and E. Torre.** 2003. Several *Salmonella enterica* subsp. *enterica* serotype 4,5,12:i:- phage types isolated from swine samples originate from serotype typhimurium DT U302. J. Clin. Microbiol. **41:**2395-2400.

18. **Echeita, M. A., A. Aladuena, S. Cruchaga, and M. A. Usera.** 1999. Emergence and spread of an atypical *Salmonella enterica* subsp *enterica* serotype 4,5,12 : i :- Strain in Spain. J. Clin. Microbiol. **37:**3425-3425.

19. **EFSA.** 2007. The Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents, Antimicrobial Resistance and Foodborne Outbreaks in the European Union in 2006. EFSA J. **130:**11-352.

20. **Fontana, J., A. Stout, B. Bolstorff, and R. Timperi.** 2003. Automated ribotyping and pulsed-field gel electrophoresis for rapid identification of multidrug-resistant *Salmonella* serotype newport. Emerg. Infect. Dis. **9:**496-9.

21. **Gal-Mor, O., Y. Valdez, and B. B. Finlay.** 2006. The temperature-sensing protein TlpA is repressed by PhoP and dispensable for virulence of *Salmonella enterica* serovar Typhimurium in mice. Microb. Infect. **8:**2154-2162.

22. **Garaizar, J., S. Porwollik, A. Echeita, A. Rementeria, S. Herrera, R. M. Y. Wong, J. Frye, M. A. Usera, and M. McClelland.** 2002. DNA microarray-

based typing of an atypical monophasic *Salmonella enterica* serovar. J. Clin. Microbiol. **40:**2074-2078.

23. **Grimont, P., and F. Weil.** 2007. Antigenic Formulae of the *Salmonella* Servovars, 9th ed, vol. Pasteur Institute, Paris: World Health Organization Centre for Reference and Research on *Salmonella*.

24. **Hudson, C. R., C. Quist, M. D. Lee, K. Keyes, S. V. Dodson, C. Morales, S. Sanchez, D. G. White, and J. J. Maurer.** 2000. Genetic relatedness of *Salmonella* isolates from nondomestic birds in Southeastern United States. J. Clin. Microbiol. **38:**1860-1865.

25. **Langvad, B., M. N. Skov, E. Rattenborg, J. E. Olsen, and D. L. Baggesen.** 2006. Transmission routes of *Salmonella* Typhimurium DT 104 between 14 cattle and pig herds in Denmark demonstrated by molecular fingerprinting. J. Appl. Microbiol. **101:**883-890.

26. **Lefebure, T., and M. J. Stanhope.** 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol. **8:**R71-88.

27. **Lercher, M. J., and C. Pal.** 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol. Biol. Evol. **25:**559-567.

28. **Lim, H., K. H. Lee, C. H. Hong, G. J. Bahk, and W. S. Choi.** 2005. Comparison of four molecular typing methods for the differentiation of *Salmonella* spp. Int. J. Food Microbiol. **105:**411-418.

29. **Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: A portable

approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. **95:**3140-3145.

30. **McClelland, M., K. E. Sanderson, S. W. Clifton, P. Latreille, S. Porwollik, A. Sabo, R. Meyer, T. Bieri, P. Ozersky, M. McLellan, C. R. Harkins, C. Y. Wang, C. Nguyen, A. Berghoff, G. Elliott, S. Kohlberg, C. Strong, F. Y. Du, J. Carter, C. Kremizki, D. Layman, S. Leonard, H. Sun, L. Fulton, W. Nash, T. Miner, P. Minx, K. Delehaunty, C. Fronick, V. Magrini, M. Nhan, W. Warren, L. Florea, J. Spieth, and R. K. Wilson.** 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica t*hat cause typhoid. Nat. Genet. **36:**1268-1274.

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

32. **Moreno Switt, A. I., Y. Soyer, L. D. Warnick, and M. Wiedmann.** 2009. Emergence, distribution and molecular and phenotypic characteristics of *Salmonella enterica* serotype 4,5,12:i:-. Foodborne Path. Dis. **in press**

33. **Mossong, J., P. Marques, C. Ragimbeau, P. Huberty-Krau, S. Losch, G. Meyer, G. Moris, C. Strottner, W. Rabsch, and F. Schneider.** 2007. Outbreaks of monophasic *Salmonella enterica* serovar 4,[5],12:i:- in Luxembourg, 2006. Euro. Surveill **12:**E11-E12.

34. **Nagano, N., S. Oana, Y. Nagano, and Y. Arakawa.** 2006. A severe *Salmonella enterica* serotype Paratyphi B infection in a child related to a pet turtle, *Trachemys scripta elegans*. Jpn. J. Infect. Dis. **59:**132-134.

35. **Orsi, R. H., Q. Sun, and M. Wiedmann.** 2008. Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. BMC Evol. Biol. **8:**233-254.

36. **Pal, C., B. Papp, and M. J. Lercher.** 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. **37:**1372-1375.

37. **Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen.** 2007. Genes under positive selection in *Escherichia coli*. Genome Res. **17:**1336-1343.

38. **Porwollik, S., E. F. Boyd, C. Choy, P. Cheng, L. Florea, E. Proctor, and M. McClelland.** 2004. Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. J. Bacteriol. **186:**5883-5898.

39. **Porwollik, S., and M. McClelland.** 2003. Lateral gene transfer in *Salmonella*. Microbes. Infect. **5:**977-989.

40. **Rabsch, W., H. L. Andrews, R. A. Kingsley, R. Prager, H. Tschape, L. G. Adams, and A. J. Baumler.** 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. Infect. Immun. **70:**2249-2255.

41. **Ross, I. L., and M. W. Heuzenroeder.** 2005. Discrimination within phenotypically closely related definitive types of *Salmonella enterica* serovar typhimurium by the multiple amplification of phage locus typing technique. J. Clin. Microbiol. **43:**1604-1611.

42. **Scaria, J., R. U. Palaniappan, D. Chiu, J. A. Phan, L. Ponnala, P. McDonough, Y. T. Grohn, S. Porwollik, M. McClelland, C. S. Chiou, C. Chu, and Y. F. Chang.** 2008. Microarray for molecular typing of *Salmonella enterica* serovars. Mol. Cell. Probes **22:**238-243.

43. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17:**14-56.

44. **Sukhnanand, S., S. Alcaine, L. D. Warnick, W. L. Su, J. Hof, M. P. Craver, P. McDonough, K. J. Boor, and M. Wiedmann.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes. J. Clin. Microbiol. **43:**3688-3698.

45. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382-389.

46. **Swaminathan, B., P. Gerner-Smidt, and T. Barrett.** 2006. Focus on *Salmonella*. Foodborne Pathog. Dis. **3:**154-156.

47. **Swaminathan, B., P. Gerner-Smidt, L. K. Ng, S. Lukinmaa, K. M. Kam, S. Rolando, E. P. Gutierrez, and N. Binsztein.** 2006. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog. Dis. **3:**36-50.

48. **Tavechio, A. T., A. C. Ghilardi, and S. A. Fernandes.** 2004. "Multiplex PCR" identification of the atypical and monophasic *Salmonella enterica* subsp. *enterica* serotype 1,4,[5],12:i:- in Sao Paulo State, Brazil: frequency and antibiotic resistance patterns. Rev. Inst. Med. Trop. Sao Paulo **46:**115-117.

49. **Thomson, N. R., D. J. Clayton, D. Windhorst, G. Vernikos, S. Davidson, C. Churcher, M. A. Quail, M. Stevens, M. A. Jones, M. Watson, A. Barron, A. Layton, D. Pickard, R. A. Kingsley, A. Bignell, L. Clark, B. Harris, D. Ormond, Z. Abdellah, K. Brooks, I. Cherevach, T. Chillingworth, J. Woodward, H. Norberczak, A. Lord, C. Arrowsmith, K. Jagels, S. Moule, K. Mungall, M. Sanders, S. Whitehead, J. A. Chabalgoity, D. Maskell, T. Humphrey, M. Roberts, P. A. Barrow, G. Dougan, and J. Parkhill.** 2008. Comparative genome analysis of *Salmonella*

Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. **18:**1624-1637.

50. **Thong, K. L., Y. L. Goh, S. Radu, S. Noorzaleha, R. Yasin, Y. T. Koh, V. K. Lim, G. Rusul, and S. D. Puthucheary.** 2002. Genetic diversity of clinical and environmental strains of *Salmonella enterica* serotype Weltevreden isolated in Malaysia. J. Clin. Microbiol. **40:**2498-2503.

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

52. **Valdezate, S., M. Arroyo, R. Gonzalez-Sanz, R. Ramiro, S. Herrera-Leon, M. A. Usera, M. De la Fuente, and A. Echeita.** 2007. Antimicrobial resistance and phage and molecular typing of *Salmonella* strains isolated from food for human consumption in Spain. J. Food Prot. **70:**2741-2748.

53. **Wiedmann, M.** 2002. Molecular subtyping methods for *Listeria monocytogenes*. J. AOAC Int. **85:**524-531.

54. **Winokur, P. L.** 2003. Molecular epidemiological techniques for *Salmonella* strain discrimination. Front. Biosci. **8:**c14-24.

55. **Wonderling, L., R. Pearce, F. M. Wallace, J. E. Call, I. Feder, M. Tamplin, and J. B. Luchansky.** 2003. Use of pulsed-field gel electrophoresis to characterize the heterogeneity and clonality of *Salmonella* isolates obtained from the carcasses and feces of swine at slaughter. Appl. Environ. Microbiol. **69:**4177-4182.

56. **Yasin, R. M., M. M. Jegathesan, and C. C. Tiew.** 1996. *Salmonella* serotypes isolated in Malaysia over the ten-year period 1983-1992. Asia Pac. J. Public Health **9:**1-5.

57. **Zamperini, K., V. Soni, D. Waltman, S. Sanchez, E. C. Theriault, J. Bray, and J. J. Maurer.** 2007. Molecular characterization reveals *Salmonella enterica* serovar 4,[5],12:i:- from poultry is a variant Typhimurium serovar. Avian Dis. **51:**958-964.

CHAPTER 2

PULSED FIELD GEL ELECTROPHORESIS CAN BE USED TO DISTINGUISH HOST ADAPTED AND UN-RESTRICTED SUBTYPES AMONG *SALMONELLA ENTERICA* SUBSP. *ENTERICA* SEROTYPES

ABSTRACT

We investigated the genetic relationships of temporally and spatially matched clinical, bovine and human *Salmonella enterica* spp. e*nterica* isolates using serotyping, multilocus sequence typing (MLST), and pulsed field gel electrophoresis (PFGE). Among clinical human and bovine *Salmonella* isolates, we determined 51 serotypes, 73 sequence types (ST) and 167 XbaI PFGE patterns. Among PFGE patterns, 116 and 44, were obtained only in human and bovine salmonellosis cases, respectively, while 7 PFGE patterns were shared by human and bovine *Salmonella* isolates. Overall, PFGE provided very high discrimination among human (Simpson's index, D= 0.9907) and cattle (D= 0.9683) isolates that was considerably higher than that achieved by serotyping (D=0.9134) and MLST (D=0.9201). We determined that two PFGE types, associated with bovine clinical cases, might represent *Salmonella* bovine adapted subtypes. On the other hand we identified that three PFGE types, showed no association at causing disease within specific host species, might represent un-restricted *Salmonella* subtypes. In our study, we identified 11 persistent *Salmonella* strains within 16 farms collected in multiple visits. In addition, spatial analysis revealed that among bovine isolates, five PFGE types representing spatial and temporal clusters. Similarly, we also found in 6 instances the same subtypes (i.e. same serotype, ST, and PFGE) were observed for two or three human isolates collected in the same county in the same or consecutive months, possibly indicating small

temporal and geographical human case clusters. We conclude that PFGE can be used to determine host adapted and un-restricted subtypes as well as spatial and temporal subtype clusters, and the further development of large subtype databases for *Salmonella* isolates from different sources will provide a better understanding of *Salmonella* transmission and therefore, facilitate better tracking of outbreaks sources.

INTRODUCTION

Salmonella, a Gram-negative enterobacterium, is a human and animal pathogen. The genus *Salmonella* is divided into two species, *Salmonella bongori* and *Salmonella enterica* (3). While *Salmonella enterica* consists of six subspecies, *Salmonella enterica* subsp. *enterica* is considered the most important, since serotypes of this subspecies are associated with warm blooded vertebrates and are responsible for the vast majority of salmonellosis cases (3, 42). *Salmonella enterica* subsp. *enterica* serotypes usually are transmitted by ingestion of food or water contaminated with feces from infected animals or humans (34), but may also be transmitted by direct contact (19, 27).

Salmonellae are among the most common reported bacterial foodborne pathogens worldwide (40). In the United States nontyphoidal *Salmonella* serotypes cause an estimated 1.4 million human salmonellosis cases with approximately 16,000 hospitalizations and 500 deaths annually (25, 45). 95 % of salmonellosis is associated with consumption of contaminated food and water in the US (25). Similarly in the European Union (EU), *Salmonella* is among the top three common known source of foodborne bacterial infection in humans. In 2006, *Salmonella* caused 34.6 human salmonellosis cases per 100,000 population (160,649 confirmed cases) in the EU (7, available at http://www.efsa.europa.eu/cs/BlobServer/DocumentSet/Zoon_report_2006_en,0.pdf?s

sbinary=true). While the major source of *Salmonella* contamination is poultry related food products, followed by pork, in the EU, in the US, in addition to meat, eggs, and their products, fecal contaminated fresh vegetables, including nuts, are also common vehicles of *Salmonella* (4, 30).

The characterization of *Salmonellae* beyond species and subspecies is performed by subtyping methods. Serotyping is a phenotype-based subtyping method that is traditionally applied to *Salmonella* (3, 46). This technique relies upon the immunoreactivity of *Salmonella*'s O (lipopolysaccaride) and H (flagellar) antigens. *Salmonella* has over 2,500 recognized serotypes according to the Kauffmann-White Scheme (13). Since serotyping has low discriminatory power to differentiate subtypes, phenotypic and genotypic approaches have been commonly used to differentiate the subtypes of *Salmonella* within serotypes, such as phage typing (20, 22), DNA sequencing-based subtyping methods (1, 37, 38, 49), ribotyping (10, 33), pulsed field gel electrophoresis (PFGE) (9, 32, 35, 37, 50), multilocus variable number of tandem repeat analysis (MLVA) (5, 24) and microarray (11, 36). PFGE has become the gold standard for subtyping bacterial pathogens, including *Salmonella,* during the past ten years due to its broad applicability, high discriminatory power and epidemiological concordance (2). Therefore, PFGE has been used by Centers for Disease Control and Prevention (CDC) and state health departments in the United States (39), as well as in Canada, Latin America, Asia and Europe (EU) (41). The Centers for Disease Control and Prevention and state health departments in the United States developed a national network (PulseNet) to rapidly exchange standardized PFGE subtype data for isolates of foodborne pathogens (39, 41). PulseNet uses the molecular subtyping of isolates that has been commonly used in epidemiological investigations of food borne disease outbreaks and has become an essential component of epidemiologic investigations of infectious diseases (39).

Serotypes of *Salmonella enterica* subsp. *enterica* can be divided in three subdivisions (i.e., host restricted, host adapted and unrestricted serotypes) according to the abilities of serotypes to cause disease in different host species. Host-restricted *Salmonella* serotypes are exclusively associated with one particular host (e.g. *Salmonella* Typhi), while host-adapted *Salmonella* serotypes are prevalent in one particular host species, but are able to cause disease in other host species (e.g. *S.* Dublin). Unrestricted *Salmonella* serotypes, although capable of causing systemic disease in a wide range of host species, usually induce a self-limiting gastroenteritis, e.g. *Salmonella* Typhimurium (42). Molecular subtyping data suggest that some un-restricted serotypes might include subtypes having a narrower host species range (30). For example, *Salmonella* Typhimurium phage type DT2 and DT99 were reported as poultry associated strains, while *Salmonella* Typhimurium phage type DT104 is a broad range variant. Therefore, *Salmonella* un-restricted serotypes might be collections of strains that vary significantly in their host range and their degree of host adaptation. These strains can be determined by molecular subtyping methods (30).

It is essential to compile databases including easily comparable phenotypic and genotypic subtypes, such as serotypes, DNA-sequencing based types, ribotypes, and PFGE types, as well as isolate information (i.e., isolate source, isolation date, etc.), for *Salmonella* isolates, especially isolates with unrestricted serotypes, world wide in order to facilitate rapid detection and tracking of the outbreak sources. To provide a better understanding of *Salmonella* transmission and facilitate a better tracking of outbreak sources, we characterized 335 temporally and geographically matched human and bovine *Salmonella* isolates collected from clinical human and bovine cases in New York State and a neighboring state, Vermont, by using serotyping and PFGE typing, as well as MLST data previously reported by our group (1). Our goals in this study were to (i) determine PFGE diversity and discriminatory ability

among clinical human and bovine *Salmonella* isolates, (ii) identify PFGE types associated with certain host species and geographical regions, (iii) evaluate the potential subtypes transmitted from cattle to humans, and (iv) provide a publicly accessible database for *Salmonella* isolates.

MATERIALS AND METHODS

***Salmonella* isolates**: A total of 335 spatially and temporally matched *Salmonella enterica* spp. e*nterica* nontyphoidal isolates were used in this study. A set of 178 human *Salmonella* isolates, collected in 2004, were obtained from the Wadsworth Center, New York State Department of Health. An additional set of 157 animal *Salmonella* isolates, collected from 64 different farms located in New York State and neighboring state Vermont in 2004, were obtained from the New York State Animal Health Diagnostic Laboratory (Table A1 [S2.1]). The procedure used to collect *Salmonella* bovine isolates is detailed in Alcaine et al. (1). While the majority of *Salmonella* isolates used in this study were previously characterized using multilocus sequence type (MLST) by our group (1), there are 10 exceptions: five *Salmonella* isolates (i.e., 1 bovine and 4 human isolates), used in our previous study, were not included in this study, while five human isolates in this study were not included in our previous study (1) . The isolate designations are identical in both studies (i.e., FSL S5-430 isolate reported here is the identical isolate with designation of FSL S5-430 previously reported in [1]).

Serotyping and MLST. Bovine isolates were serotyped at the National Veterinary Services Laboratories (USDA-APHIS-VS, Ames, IA) (8). Human isolates and serotype information for isolates were delivered at the Wadsworth Center, New York State Department of Health.

Human and bovine isolates were previously typed by MLST (Multilocus

sequence typing) based on PCR amplification and sequencing of three genes (*manB, fimA* and *mdh*) (1, 38), and were previously reported (1). The sequence types are identical for the corresponding isolates in both studies (i.e, ST8, reported here was determined by Alcaine et al. [1] and is identical to ST8 previously reported). Sequence types (STs), assigned previously by Alcaine et al. (1), were used in this study to (i) determine the combined subtypes for *Salmonella* clinical human and bovine isolates using three subtyping methods (serotyping, MLST, and PFGE) and (ii) determine which subtyping method among three subtyping methods (serotyping, MLST, and PFGE) has the highest discriminative power for typing *Salmonella* isolates.

Among bovine *Salmonella* isolates, the same serotype, MLST and PFGE subtypes were obtained from multiple isolates that were collected from the same farm during different visits. This indicates the re-isolation of a persistent subtype on a given farm. Therefore, only one isolate representing each unique serotype/MLST/PFGE combined subtype for a given farm was included in the statistical analysis and Simpson's index of diversity calculations to avoid over-representation of a subtype due to re-sampling in multiple visits of a given farm. For example, from farm 510, a combined subtype of serotype Newport, ST 11 and PFGE type 121 was obtained in 20 different visits (Table 2.1). However, only one of the isolates with this combined subtype from farm 510 was used for statistical analysis. Therefore, out of 157 bovine *Salmonella* isolates, a total of 91 cattle isolates were used in statistical analysis.

PFGE anlaysis. PFGE was performed according to the Centers for Disease Control and Prevention PulseNet protocol (http://www.cdc.gov/pulsenet/protocols/ecoli_salmonella_shigella_protocols.pdf) (32). PFGE was performed with the CHEF-Mapper (Bio-Rad Laboratories, Hercules, CA). Electrophoresis conditions were an initial switch time of 2.16 seconds, a final

Table 2. 1. *Salmonella* isolate information for clinical bovine isolates collected from farms at multiple visits

Farm ID	No. of farm visits with positive *Salmonella* samples	Sequence type/serotype (No. of isolates)	PFGE Pattern (No. of isolates)
510	20	ST11/Newport (21)	121 (20), 122 (1)
		ST6/4,5,12:i:- (1)	94 (1)
261	22	ST6/ 4,5,12:i:- (18)	89 (5), 90 (1), 91(1), 94 (10), 95 (1)
		ST17/Kentucky (5)	96 (5)
		ST6/Typhimurium (1)	94 (1)
223	15	ST60/Infantis (15)	107 (13), 108 (1), 109 (1)
329	5	ST9/Montevideo (1)	119 (1)
		ST44/Muenster (3)	7 (3)
		ST62/Thompson (1)	157 (1)
186	4	ST75/Adelaide (1)	44 (1)
		ST8/Typhimurium (3)[a]	104 (3)
524	5	ST6/4,5,12:i: (1)	90 (1)
		ST11/Newport (4)	126 (1), 127 (3)
152	4	ST11/Newport (4)	126 (4)
490	4	ST11/Newport (4)	126 (1), 127 (2), 129 (1)
163	3	ST60/Infantis (1)	114 (1)
		ST11/Newport (2)	126 (2)
259	3	ST44/Muenster (3)	2 (1), 4 (1), 6 (1)
488	3	ST11/Bardo (1)	126 (1)
		ST11/Newport (3)	126 (3)
584	3	ST2/Agona (2)	165 (1), 166 (1)
		ST6/Typhimurium (1)	64 (1)
97	2	ST8/Typhimurium[b] (2)	104 (2)
105	2	ST11/Newport (1)	121 (1)
		ST8/Typhimurium (1)	102 (1)
125	2	ST6/Typhimurium (2)	79 (2)
208	2	ST6/Typhimurium (2)	66 (2)
303	2	ST11/Newport (2)	126 (2)
320	2	ST11/Newport (2)	126 (2)
415	2	ST9/Montevideo (1)	119 (1)
		ST6/Typhimurium (1)	70 (1)
764	2	ST6/Typhimurium (2)	60 (2)

[a]Among 3 Typhimurium isolates, there is a Typhimurium var. 5, previously known as Typhimurium Copenhagen, isolate [b]Among 2 Typhimurium isolates, there is a Typhimurium var. 5, previously known as Typhimurium Copenhagen, isolate

switch time of 63.8 seconds and a run time of 21 hours. CDC *Salmonella* serovar Branderup H19812 was used as the reference strain (17). Pictures of PFGE gels were taken by Gel/ChemiDoc (BIO-RAD Laboratories). Among 335 *Salmonella* isolates, 10 isolates with serotypes Kentucky (n=6), representing all Kentucky isolates, Infantis (n=2), Oranienburg (n=1), and Havana (n=1), could not be typed by the routine CDC PulseNet protocol. The addition of 50 μM thiourea to the running buffer yielded clear, interpretable PFGE patterns for these isolates (26).

Comparison analysis was performed using the BioNumerics Software package (Applied Maths 1998-2004, Austin, TX). Similarity analysis was performed using the Dice coefficient and clustering was created using the unweighted pair group method by arithmetic mean. All the isolates were coded according to CDC PulseNet codes for naming PFGE patterns with initials of New York Cornell University (NYCU), e.g.: NYCU.JP6X01.0001.

Statistical analysis. PFGE types were grouped such that there were at least 5 isolates for each PFGE type; PFGE types including less than five isolates were classified as "rare PFGE types". The frequency distributions of PFGE types for isolates from human and bovine clinical cases were compared using the chi-square test of independence or Fisher's exact test. For comparisons where one or more of the expected values was <5, Fisher's exact test was conducted. P-values lower than 0.05 were considered statistically significant. All statistical analyses were conducted with Statistical Analysis Systems (SAS) 9.1 (SAS Institute Inc., Cary, NC).

Simpson's index of diversity (D) was calculated as described (16) to assess the differentiation of *Salmonella* isolates by using serotyping, PFGE, MLST or the combination of the two or three subtyping methods. Simpson's index of diversity values reported here cannot be directly compared to those reported by Alcaine et al. (1) as the isolate sets used in these two sets were similar but not identical.

Spatial analysis. A New York State and Vermont map from MapViewer software (MapViewer package version 6.0, Golden Software) was used to identify the geographical clusters of cattle *Salmonella* isolates.

Access to detailed isolate information. All isolate information, including isolate source, gene sequence data, allele assignments, antibiotic resistance profiles, and PFGE images, are publicly available in the Pathogen Tracker website (www.pathogentracker.net).

RESULTS

PFGE is more discriminatory than serotyping and MLST. A total of 335 *Salmonella* isolates with 51 serotypes were characterized into 167 PFGE types. Previously, this isolate set was differentiated into 73 sequence subtypes (STs) (1) (Table 2.2). PFGE typing showed the highest discriminatory power among *Salmonella* clinical isolates as determined by Simpson's Index of Discrimination (D=0.9910). PFGE was followed by MLST (D=0.9201) and serotyping (D=0.9188), respectively. For example, 16 *Salmonella* serotype 4,5,12:i:- isolates with ST6 were differentiated into 9 different PFGE types. Overall, subtype diversity was always higher among human *Salmonella* isolates as compared to bovine isolates, regardless of subtyping methods; Simpson's index of diversity values for serotype, MLST, and PFGE diversity among human isolates were 0.9327, 0.9414, and 0.9913, respectively, as compared to Simpson's index of diversity values of 0.8415, 0.8046, and 0.9683 among cattle isolates, respectively (Table 2.2).

Among 51 serotypes, 14 serotypes that occurred more than 3 times among the *Salmonella* isolate set used in this study were assigned as "common serotypes" (Table 2.3). *Salmonella* isolates within these common serotypes were further subdivided by MLST and PFGE into one to six different STs, and two to 32 different PFGE types,

Table 2. 2. Distribution of subtypes found using serotyping, MLST, PFGE typing as well as combinations among clinical bovine and human *Salmonella* isolates and corresponding Simpson's Index of diversity scores for each subtyping method as well as combined methods

	Number of subtypes found in				Simpson's Index of Diversity scores among		
	Only human cases	Only bovine cases	Both human and bovine cases	Total	Human isolates	Bovine isolates	Total
Serotype (SrT)	35	5	11	51	0.9327	0.8229	0.9134
MLST	57	6	10	73	0.9414	0.8046	0.9201
PFGE	116	44	7	167	0.9913	0.9683	0.9910
SrT+PFGE	117	48	7	172	0.9914	0.9736	0.9919
MLST+PFGE	120	44	6	170	0.9916	0.9683	09912
SrT+MLST+PFGE	119	49	6	174	0.9916	0.9736	0.9920

respectively (Table 2.3). For example, 47 human and bovine *Salmonella* isolates with serotype Newport, the second most common serotype found in our isolate set, were differentiated into 6 STs and 22 PFGE types (Table 2.3), while 26 human *Salmonella* isolates with the third common serotype in our isolate set, Enteritidis, were differentiated into two STs and 8 PFGE types.

While, PFGE typing showed higher discriminatory power than serotyping and MLST, there were some exceptions. In five incidences of typing isolates with closely related serotypes, serotyping is more discriminatory than PFGE (Table 2.4). For example, PFGE type 94 represented one bovine Typhimurium (4,5,12:i:1,2) isolate and two bovine 4,5,12:i:- isolates, which lacks the second phase flagellar antigen (Figure 2.1). Similarly, PFGE types 66, 157, and 158 were differentiated into two different STs. PFGE types 66 and 158 were differentiated in two STs (i.e., ST6 and ST7, and ST43 and ST62, respectively) within a serotype (i.e., Typhimurium and Thompson, respectively), while PFGE type 157 was differentiated in two STs (i.e., ST43 and ST62) in two serotypes (i.e., 1,7:-:1,5 and Thompson, respectively) (Figure 2.2) (Table A1 [S2.1]).

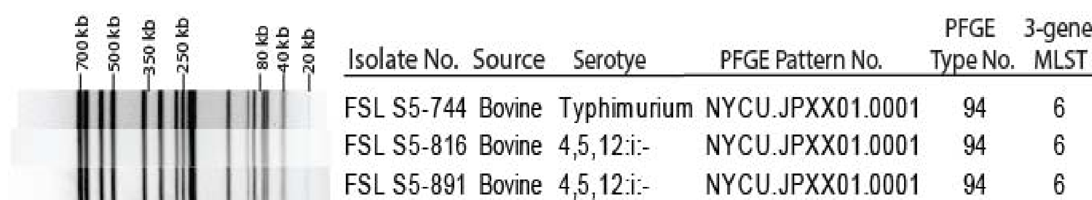
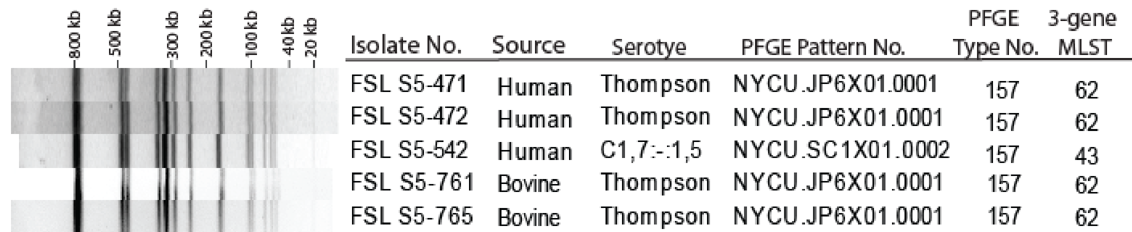


Figure 2. 1. PFGE type 94 is an example of PFGE types shared by different serotypes. PFGE type 94 was determined in one bovine *Salmonella* Typhimurium (4,5,12:i:1,2) isolate and two bovine *Salmonella* 4,5,12:i:- isolates.

To be more discriminatory, we combined three subtyping methods and assigned a total of 174 different "combined" subtypes (D=0.9920) in this set of

Salmonella isolates (Table 2.2). Although the combination of subtyping methods yielded more subtypes than that obtained from PFGE typing, the discriminatory power of three methods combined (D=0.9921) is similar to that of PFGE typing (D=0.9910). Among 174 combined subtypes, only 4 combined subtypes were found among both human and bovine isolates.



Isolate No.	Source	Serotye	PFGE Pattern No.	PFGE Type No.	3-gene MLST
FSL S5-471	Human	Thompson	NYCU.JP6X01.0001	157	62
FSL S5-472	Human	Thompson	NYCU.JP6X01.0001	157	62
FSL S5-542	Human	C1,7:-:1,5	NYCU.SC1X01.0002	157	43
FSL S5-761	Bovine	Thompson	NYCU.JP6X01.0001	157	62
FSL S5-765	Bovine	Thompson	NYCU.JP6X01.0001	157	62

Figure 2. 2. PFGE type 157 was differentiated in ST43 and ST62 in two serotypes, serotype 1,7:-:1,5 and serotype Thompson.

Distribution of PFGE types among human and bovine *Salmonella* isolates. Categorical analysis of the distribution of PFGE types from an overall 9 by 2 table (8 PFGE types that included at least 5 isolates plus one category for rare types [<5 isolates] by two host species, human and bovine) relieved that PFGE types were not independently distributed among human and bovine cases (*P*-value <0.0001; Monte Carlo estimation of exact test).

Among 167 PFGE types, only four PFGE types, 60, 89, 126, and 157, were obtained from both bovine and human clinical cases (Table 2.6) (Figure 2.3, Figure 2.2). For example PFGE pattern 60 was shared by five human and three bovine *Salmonella* Typhimurium isolates, while PFGE type 89 was shared by 6 human and one bovine *Salmonella* 4,5,12:i:- isolates, as well as one bovine *Salmonella* Typhimurium isolate (Figure 2.3).

Table 2. 3. Distribution of sequence types and PFGE types among common *Salmonella* serotypes[a]

Serotype	No. of isolates	Among isolates no of different	
		STs found	PFGE types found
Typhimurium[b]	52	5	32
Newport	47	6	22
Enteritidis	26	2	8
4,5,12:i:-	17	2	9
Heidelberg	10	3	4
Montevideo	9	5	6
Thompson	9	2	3
Agona	8	2	7
Muenster	8	1	6
Infantis	7	1	7
Mbandaka	5	3	5
Saintpaul	5	2	4
Javiana	4	2	4
Urbana	4	1	3

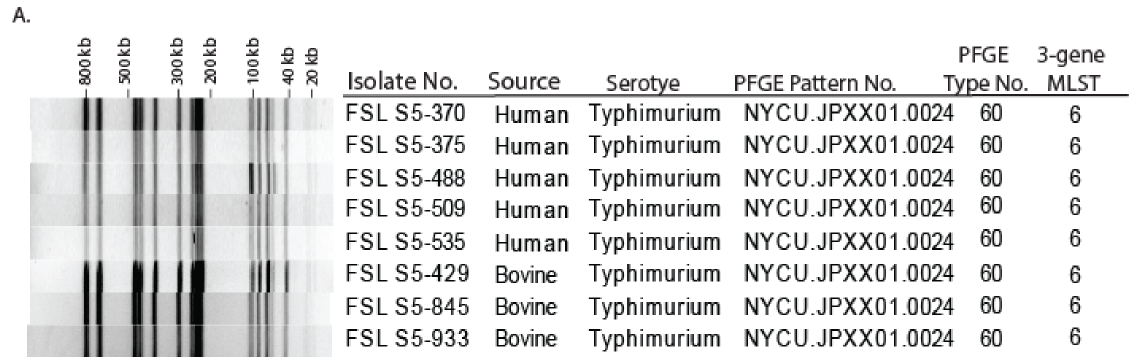
[a] *Salmonella* serotypes, which were found ≥ 4 times among *Salmonella* isolate set used in our study, were called common *Salmonella* serotypes
[b] Among 52 Typhimurium isolates, there are four Typhimurium var. 5, previously known as Typhimurium Copenhagen, isolates

Table 2. 4.. PFGE patterns shared by *Salmonella* isolates within closely related serotypes

PFGE Pattern	Serotype	Antigens	Source (No. of isolates)
89	4,5,12:i:-	4,5,12:i:-	Human (6), Bovine (1)
	Typhimurium	4,5,12:i:1,2	Bovine (1)
94	4,5,12:i:-	4,5,12:i:-	Bovine (2)
	Typhimurium	4,5,12:i:1,2	Bovine (1)
121	Newport	6,8:e,h:1,2	Bovine (9)
	Bardo	8:e,h:1,2	Bovine (1)
126	Newport	6,8:e,h:1,2	Human (2), Bovine (10)
	Bardo	8:e,h:1,2	Bovine (1)
157	Thompson	6,7:k:1,5	Human (2), Bovine (2)
	C 1,7:-:1,5	1,7:-:1,5	Human (1)

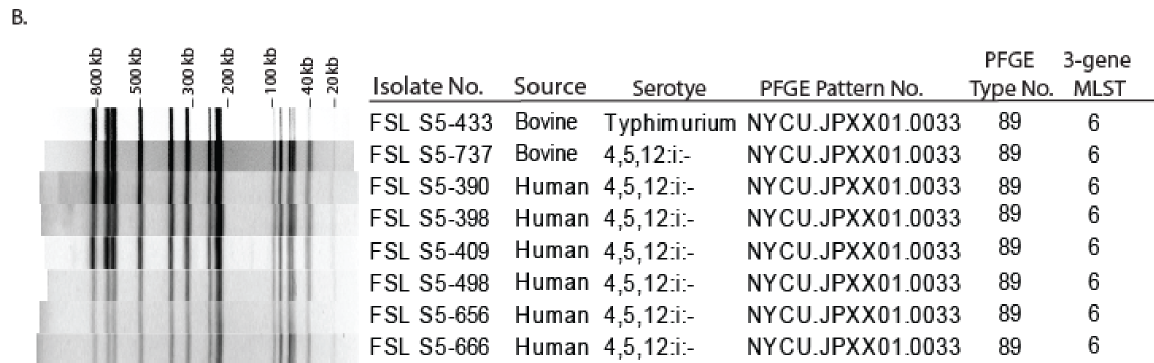
Figure 2. 3. Among 167 PFGE types, only four PFGE types, 60, 89, 126, and 157 (Figure 2.2), were obtained from both bovine and human clinical cases. For example PFGE pattern 60 (A) shared by five human and three bovine *Salmonella* Typhimurium isolates, PFGE type 89 (B) shared by 6 human and one bovine *Salmonella* 4,5,12:i:- isolates, as well as one bovine *Salmonella* Typhimurium isolates, and PFGE type 126 (C) obtained from two and 10 human and bovine *Salmonella* Newport isolates, respectively, as well as from one bovine *Salmonella* Bardo isolate.

A.



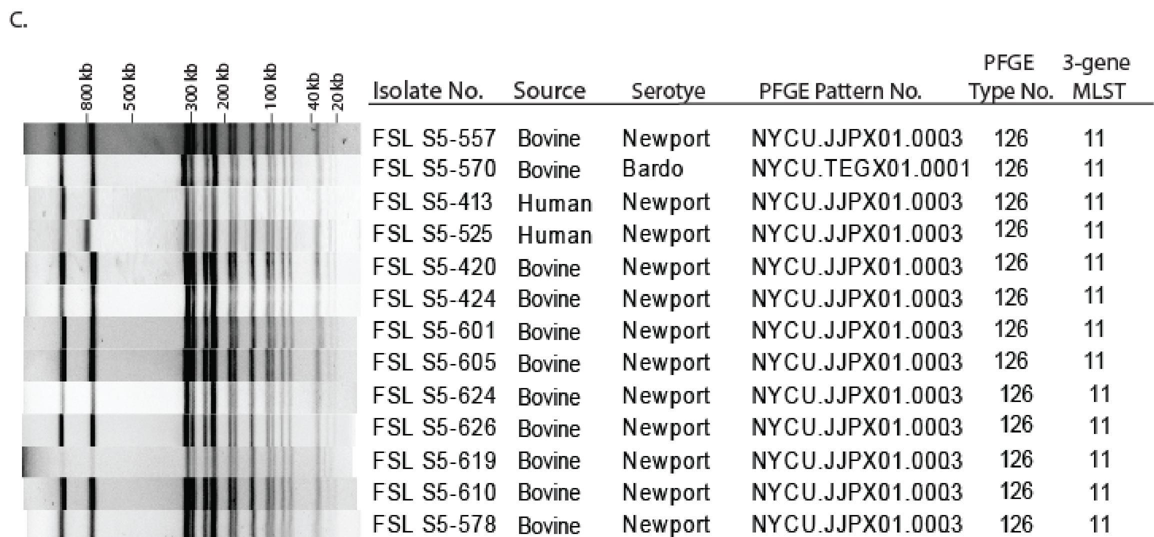
Isolate No.	Source	Serotye	PFGE Pattern No.	PFGE Type No.	3-gene MLST
FSL S5-370	Human	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-375	Human	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-488	Human	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-509	Human	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-535	Human	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-429	Bovine	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-845	Bovine	Typhimurium	NYCU.JPXX01.0024	60	6
FSL S5-933	Bovine	Typhimurium	NYCU.JPXX01.0024	60	6

B.



Isolate No.	Source	Serotye	PFGE Pattern No.	PFGE Type No.	3-gene MLST
FSL S5-433	Bovine	Typhimurium	NYCU.JPXX01.0033	89	6
FSL S5-737	Bovine	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-390	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-398	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-409	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-498	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-656	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6
FSL S5-666	Human	4,5,12:i:-	NYCU.JPXX01.0033	89	6

C.



Isolate No.	Source	Serotye	PFGE Pattern No.	PFGE Type No.	3-gene MLST
FSL S5-557	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-570	Bovine	Bardo	NYCU.TEGX01.0001	126	11
FSL S5-413	Human	Newport	NYCU.JJPX01.0003	126	11
FSL S5-525	Human	Newport	NYCU.JJPX01.0003	126	11
FSL S5-420	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-424	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-601	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-605	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-624	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-626	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-619	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-610	Bovine	Newport	NYCU.JJPX01.0003	126	11
FSL S5-578	Bovine	Newport	NYCU.JJPX01.0003	126	11

Categorical analysis of PFGE type distributions using chi-square or Fischer's exact tests showed that PFGE types 121, and 126 were significantly associated with cattle cases (*P*-value <0.001), while PFGE type 27 was significantly associated with human cases (*P*-value *<0.05)* (Table 5). PFGE type 126 was obtained from both human and bovine cases (i.e, PFGE type 126, which was found 14 *S.* Newport and 1 *S.* Bardo isolates, was collected widely from 10 different farms in 7 counties among New York State and Vermont [Table A1 [S2.1]). The rest of the PFGE types, including the rare PFGE type category, were independently distributed among human and bovine cases.

Evidence of persistent subtypes on farms. As previously described, 64 farms were involved in this study. While 43 farms were only visited once, 20 farms were visited more than once due to extended salmonellosis infection. In multiple visits to given farms, some persistent combined subtypes (i.e., same serotype, MLST and PFGE pattern) were obtained from 16 farms (Table 1). Although our previous study showed the persistence of a given ST on most of the farms (1), this study shows that *Salmonella* isolates with certain persistent STs included multiple PFGE types, since PFGE provided more discrimination than MLST. For example, 18 *Salmonella* 4,5,12:i:- isolates, sharing the same sequence type, ST6, on Farm 261, were differentiated into 5 different PFGE types. Similarly, on farm 223, 15 *Salmonella* Infantis isolates, having the same sequence type, ST60, were differentiated into 3 different PFGE types: 107, 108, 109. In this study we used combined subtypes (serotype, ST, and PFGE type) to identify 11 persistent subtypes (i.e., combined subtype found more than once) within a given farm (Table 2.1). For example, a combined subtype of Newport, ST11 and PFGE type 121 was determined in 20 isolates from farm number 510.

Table2.5. Distribution of PFGE patters that were determined more than four times among clinical human and bovine *Salmonella* isolates and *P*-values representing the association between host species and PFGE patterns

	No. of isolates from		
	Human	Cattle	
27 (Enteritidis)	9	0	0.0291*
32 (Enteritidis)	6	0	0.0765
57 (Heidelberg)	7	0	0.0553
60 (Typhimurium)	5	3	0.8237
89 (4,5,12:i:-, Typhimurium)	6	2	0.2871
121 (Newport)	0	10	<0.0001**
126 (Newport)	2	11	<0.0001**
157 (C 1,7:-:1,5, Thompson)	3	2	0.7685
Rare PFGE types	140	63	0.0893
Total	178	91	

[a] *P* values refer to comparisons of the frequency of a given PFGE type among human and bovine isolates, as determined by chi-square test or Fisher's exact test.
* indicates significance between PFGE pattern within a serotype and host species $P<0.05$
** indicates significance between PFGE pattern within a serotype and host species $P<0.001$

Evidence of temporal and spatial clusters of *Salmonella* subtypes. Among the human isolates, we observed the same subtypes (i.e. same serotype, MLST, and PFGE) in 6 instances for two or three isolates collected in the same county in the same or consecutive months (Table 2.6), possibly indicating small temporal and geographical case clusters. For example, the isolates with serotype Anatum, ST 25 and PFGE type 151, were obtained from 3 different people in the same month and county. Interestingly, PFGE types 14, 47, 86, and 151 were only found in these human clusters (Table 2.6). This may indicate that these isolates are unique to these human cases. While other PFGE types were found in other human and cattle clinical cases in our study (i.e., PFGE type27 was commonly found among human isolates, while PFGE type 157 was also obtained from bovine cases [Table A1 [S2.1.]]).

Spatial analysis of the distribution of PFGE patterns from bovine *Salmonella* isolates, found more than once, showed that (i) nine PFGE types (PFGE types 5, 7, 60, 90, 94, 104, 132, 157, and 166) that were collected from at least two farms in two different adjacent counties might represent widely distributed patterns, (ii) three PFGE types (PFGE type 96, 119, and 127) that were collected from multiple farms in 2 or three adjacent counties might represent spatial clusters, and (iii) two PFGE types (PFGE types 121 and 126) that were collected in bovine cases in 10 farms in 6 and 7 counties with and without boundaries might represent commonly found and widely distributed subtypes (Figure 2.4).

DISCUSSION

A total of 335 human and bovine *Salmonella* clinical isolates, collected from New York and a neighboring state (Vermont), were characterized by serotyping and PFGE, as well as MLST (1), to provide a better understanding of the genetic relationship and epidemiology of human and cattle associated *Salmonella.* Our data

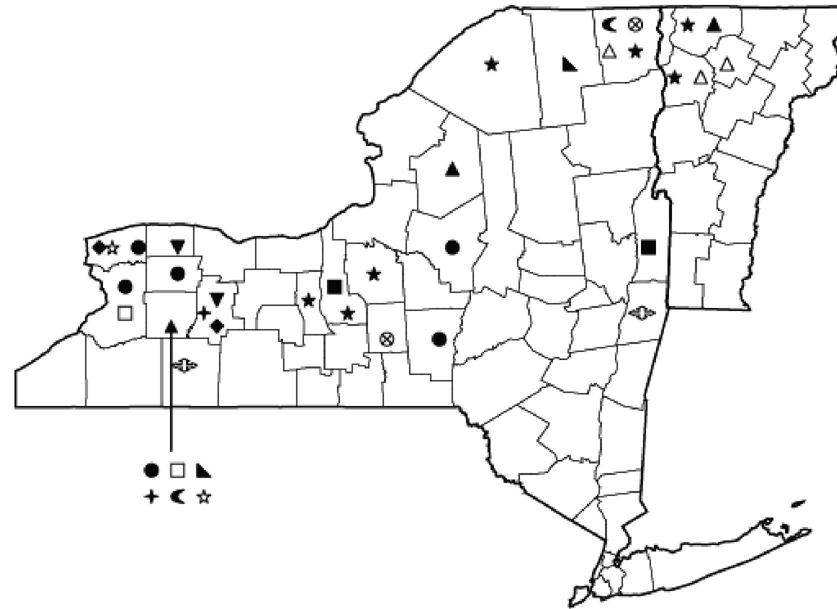
Table 2. 6. *Salmonella* human isolates that might have been collected from human clinical clusters within a county in two month duration

Isolate No.	Date of isolation	County collected from	Serotype	MLST	PFGE Pattern
FSL S5-529	9/15/2004	Erie	Anatum	25	151
FSL S5-530	9/14/2004	Erie	Anatum	25	151
FSL S5-540	9/22/2004	Erie	Anatum	25	151
FSL S5-369	12/22/2003	Monroe	Saintpaul	38	86
FSL S5-405	12/23/2003	Monroe	Saintpaul	38	86
FSL S5-376	12/31/2003	Nassau	Enteritidis	14	27
FSL S5-377	12/31/2003	Nassau	Enteritidis	14	27
FSL S5-471	5/4/2004	Nassau	Thompson	62	157
FSL S5-472	5/3/2004	Nassau	Thompson	62	157
FSL S5-456	4/22/2004	Orleans	Schwarzengrund	4	14
FSL S5-458	4/30/2004	Orleans	Schwarzengrund	4	14
FSL S5-388	1/11/2004	Schenectady	Urbana	52	47
FSL S5-410	2/28/2004	Schenectady	Urbana	52	47

together with the previous MLST data from our group (1) indicate that (i) PFGE provides higher discrimination for human and bovine *Salmonella* isolates than serotyping and MLST; (ii) PFGE can be used to differentiate host specific, un-restricted and widely distributed *Salmonella* isolates, as well as previously unknown human clusters with epidemiological data, and (iii) persistent *Salmonella* isolates might cause infection more than once on the same farm.

PFGE more discriminatory for human and bovine *Salmonella* isolates than serotyping and MLST. Serotyping, which is a phenotype-based subtyping method, has been used for subtyping *Salmonella* isolates with the Kauffmann-White scheme since 1968 (4). Since serotyping of *Salmonella* often cannot provide enough discriminatory ability to detect outbreaks and track the outbreak sources, modern rapid and standardized molecular techniques have been commonly used to subdivide the *Salmonella* isolates within a serotype, such as phage typing (43, 47), random amplification of polymorphic DNA, RAPD (15, 21), DNA sequeuncing-based subtypes (i.e., multilocus sequence typing, MLST [(1, 23, 38)]), ribotyping, (10), pulsed field gel electrophoresis, PFGE (43, 48), and multiple-locus variable-number tandem repeat analysis (MLVA) (5). While a standardized set of phages is used between laboratories to phage type *Salmonella*, the results from different laboratories, even within a laboratory, might be varied due to high biological and experimental variation (46). Since the *Salmonella* genome is a highly clonal, neither ribotyping nor MLST schemes provide enough discrimination to differentiate *Salmonella* subtypes (1, 38). In PFGE typing, the whole genome is digested into 8 to 25 large DNA bands (46), therefore, PFGE provides better discrimination. PFGE is considered as the gold standard for *Salmonella* subtyping (32),as well as other foodborne bacteria, such as *Listeria monocytogenes* (12) (Graves et al. 2001, by the CDC (39). Other researchers reported that PFGE provided more discriminatory power of differentiation of

Figure 2. 4. Spatial analysis of bovine *Salmonella* isolates showed that PFGE types 5, 60, 90, 104, 132, 157, and 166 are widely distributed and PFGE types 7, 94, 96, and 127 geographically clustered, while PGFE types 121 and 126 were widely distributed among New York State and Vermont



Symbol	PFGE Pattern #	Serotype	No. of County (No. of farms)
◣	5	Muenster	2 (2)
▼	7	Muenster	2 (2)
⊗	60	Typhimurium	2 (3)
☾	90	4,5,12:i:-	2 (2)
☆	94	4,5,12:i:-, Typhimurium	2 (2)
□	96	Kentucky	2 (2)
■	104	Typhimurium, T.Copenhagen	2 (3)
+	119	Montevideo	2 (2)
●	121	Bardo, Newport	6 (10)
★	126	Bardo, Newport	7 (10)
△	127	Newport	3 (4)
▲	132	Newport	2 (2)
◆	157	Thompson	2 (2)
✣	166	Agona	2 (2)

Salmonella isolates than ribotyping (10, 33), RAPD (15, 21), and MLST (Cooke et al. 2008; Fakhr et al. 2001; Harbottle et al. 2006).

In our study, 355 human and bovine *Salmonella* isolates with 51 serotypes (D= 0.9134) were differentiated in 73 sequence types (D= 0.9201) and 167 PFGE types (D= 0.9910). Our data that showed PFGE had more discriminatory power than serotyping and MLST, is consistent with previous studies (6, 9, 14). Harbottle et al. (14) reported that 81 *Salmonella* Newport isolates, which were isolated from humans, feed and foods, were differentiated in 12 sequence types, assigned using 7-gene MLST (i.e., *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, sucA, and thrA), while 43 XbaI PFGE types were found. In another study, 85 cattle clinical *Salmonella* Typhimurium isolates showed no genetic diversity among 4 genes (i.e., *manB*, *pduF*, *glnA*, and *spaM*), but were differentiated into 50 XbaI profiles (9). Similarly, Cooke et al. (6) reported that 13 human clinical *Salmonella* Typhimurium definitive phage type 104 (DT104) isolates having 13 different XbaI PFGE types represented a single sequence type, assigned by using 7-gene MLST (i.e., *aroC*, *dnaN*, *hemD*, *hisD*, *purE*, sucA, and thrA) and these 13 DT104 isolates had also conserved gene contents, determined by microarray. Cooke et al. (6) proposed that the differences among 13 human clinical *Salmonella* Typhimurium DT104 isolates might be due to the variation in prophage and plasmid contents, which can also be determined by PFGE.

PFGE might be used to differentiate host specific, un-restricted and widely distributed *Salmonella* isolates, as well as previously unknown human clusters with epidemiological data. We found that a total of three PFGE types are associated with clinical cases from certain host species (i.e., PFGE type 27 was associated with human clinical cases, while PFGE types 121 and 126 were associated with bovine clinical cases [*P*- values= 0.0291, <0.0001 and <0.0001, respectively]). Since PFGE type 27 was found in *Salmonella* Enteritidis isolates, which is commonly

transmitted form poultry products to humans, we expected that PFGE type 27 would be associated with human clinical cases, and not found in bovine clinical cases. However, PFGE types 121 and 126 were obtained from Newport, un-restricted serotype. PFGE type 121 was obtained from only bovine isolates, while PFGE type 126 was obtained from both human and cattle cases, however it was commonly found on 10 different farms in 7 counties, representing 14 *S.* Newport and 1 *S.* Bardo (Table A1 [S2.1]). These data show that some PFGE types can be responsible for salmonellosis among herds and might be associated with bovine clinical cases (i.e., host adapted isolates). These subtypes are also able to cause disease in humans. The finding of these host adapted subtypes in human cases, especially outbreaks, with the strong epidemiological linkage means that the source of disease or outbreak is more likely be transmitted to humans from cattle associated products. This data is supported by previous studies that showed *Salmonella* Newport had distinct lineages that can cause disease in different host species (1, 14). In addition, the same trend was seen in other un-restricted serotypes, such as Typhimurium DT40, which represents the avian-adapted *Salmonella* Typhimurium phage type (30).

Using a combination of three subtyping methods as well as epidemiological data, we determined previously unknown human clusters in 6 instances for two or three isolates collected in the same county in the same or consecutive months (Table 2.6), possibly indicating small temporal and geographical case clusters. This data suggests that the number of human salmonellosis clusters occurs more often than reported.

Although we used a subtyping method with higher discriminatory power, PFGE, to differentiate isolates among un-restricted serotypes, we identified four PFGE types that were not associated with a specific host species (Table 2.5); 60 (i.e., PFGE type 60 was obtained from five human and three bovine *Salmonella* Typhimurium

isolates), 89 (i.e., PFGE type 89 was obtained from 6 human and one bovine *Salmonella* 4,512:i:- isolates, as well as one Typhimurium isolate), and 157 (i.e., PFGE type 157 was obtained from two human and two bovine *Salmonella* Thompson isolates, as well as one human C1,7:-:1,5 isolate), that were obtained from both bovine and human clinical cases. This data suggests that some PFGE types may be widely distributed among cattle and human cases. Therefore, finding common PFGE types in humans and cattle or food does not mean that they are associated, so the isolation of common PFGE types among humans and foods or farm animals must be interpreted carefully. Establishment of casual relationships will require strong epidemiological linkages and/or the use of additional, more sensitive methods such as multiple-locus variable-number of tandem repeat analysis (MLVA) (i.e., MLVA involves the amplification and fragment size analysis of polymorphic regions of DNA containing variable numbers of tandem repeat sequences) (5, 18, 24), and microarray (11, 31, 36), or whole genome sequencing (6).

Persistent *Salmonella* isolates might cause infection multiple times within a farm. A number of studies have shown that *Salmonella* can persistent on farms and flocks. Previously, Ogilive (28) reported that in a dairy farm, one cow shed *Salmonella* Typhimurium in her milk and contaminated milk products in a 36 day period, while she did not show any symptoms of salmonellosis (28). Similarly, Vanselow et al. (44) reported that *Salmonella* Typhimurium persisted in a dairy herd over approximately 2 years. However in this case, *Salmonella* Typhimurium caused severe salmonellosis among dairy cows and calves, resulting in the death of a cow and a calf. In addition, it was suspected that *Salmonella* Typhimurium was transmitted to a 9 month-old child living on the farm and caused salmonellosis (44). Another epidemiological study, investigating the survival of *Salmonella* Senftenberg in broiler parent stock and broiler farms, showed that *Salmonella* Senftenberg was persistent for

two years in a broiler parent stock, despite management practices, such as cleaning, disinfection, desiccation and also depopulation (29).These studies support our data that showed persistent *Salmonella* isolates with the same serotype, ST, and PFGE type, found in 16 farms. Besides *Salmonella* serotype Typhimurium, we found four more *Salmonella* serotypes, 4,5,12:i:-, Infantis, and Newport that were persistent in the farms included in our study (Table 2.1). These data indicate that some *Salmonella* subtypes might be persistent within a farm and cause salmonellosis among cattle.

Overall, we can conclude that some *Salmonella* serotypes might survive in the environment and might resist farm/broiler management practices, as well as might persist due to re-contamination by rodents, wild animals, birds and/or flies. Therefore, farm management practices should be conducted thoroughly and necessary modifications should be made to eliminate persistent *Salmonella* serotypes, even though there is no sign of salmonellosis found in the herd or flock. The persistent *Salmonella* isolates might not cause any disease in animal host species, but the animals may be the reservoir of these persistent *Salmonella* isolates that can cause severe salmonellosis in humans.

CONCLUSIONS

In this study, we used a phenotypic subtyping method, serotyping, as well as genotypic subtyping methods, MLST and PFGE, to build a database, which will be used to determine host specific, un-restricted, geographically clustered and spatially persistent *Salmonella* subtypes. The combination of these methods provided a better understanding of ecology and transmission of *Salmonella*, one of the commonly known foodborne bacterial pathogen around the world. The data we generated in this study, as well as the isolate information (i.e., isolation year, source, antibiotic resistance, county, etc.) are publicly available in Pathogen Tracker

(www.pathogentracker.net) and can be used by other researchers for their academic purposes and for their epidemiological investigations. We conclude that the development of a larger database including *Salmonella* isolates from various sources from different regions of the US, as well as different countries and continents, with combination of more sensitive molecular biology techniques will enhance our abilities to detect the outbreaks and link the outbreak to the specific food faster, therefore we will improve abilities to assure the public health.

ACKNOWLEDGEMENTS

This work was supported by Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. NO1-A1-30054. We thank Eric B. Fugett for help with laboratory techniques used in this study.

REFERENCES

1. **Alcaine, S. D., Y. Soyer, L. D. Warnick, W. L. Su, S. Sukhnanand, J. Richards, E. D. Fortes, P. McDonough, T. P. Root, N. B. Dumas, Y. Grohn, and M. Wiedmann.** 2006. Multilocus Sequence Typing Supports the Hypothesis that Cow- and Human-Associated *Salmonella* Isolates Represent Distinct and Overlapping Populations. Appl. Environ. Microbiol. **72:**7575-7585.
2. **Barrett, T. J., P. Gerner-Smidt, and B. Swaminathan.** 2006. Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. Foodborne Pathog. Dis. **3:**20-31.
3. **Brenner, F. W., R. G. Villar, F. J. Angulo, R. Tauxe, and B. Swaminathan.** 2000. *Salmonella* nomenclature - Guest commentary. J. Clin. Microbiol. **38:**2465-2467.
4. **CDC.** 2008. *Salmonella* Surveillance: Annual Summary, 2006, vol. Atlanta, Georgia: US Department of Health and Human Services, CDC.
5. **Cho, S., T. S. Whittam, D. J. Boxrud, J. M. Bartkus, and A. M. Saeed.** 2008. Allele distribution and genetic diversity of VNTR loci in *Salmonella enterica* serotype Enteritidis isolates from different sources. BMC Microbiol. **8:**146.
6. **Cooke, F. J., D. J. Brown, M. Fookes, D. Pickard, A. Ivens, J. Wain, M. Roberts, R. A. Kingsley, N. R. Thomson, and G. Dougan.** 2008. Characterization of the genomes of a diverse collection of *Salmonella enterica* serovar Typhimurium definitive phage type 104. J. Bacteriol. **190:**8155-62.
7. **EFSA.** 2007. The Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents, Antimicrobial Resistance and Foodborne Outbreaks in the European Union in 2006. The EFSA Journal **130:**11-352.

8. **Ewing, W. H.** 1972. The genus *Salmonella* p. 146-256. *In* R. E. Edwards and W. H. Ewing (ed.), In Identification of *Enterobacteriaceae* vol. 3rd. Burgess Publishing Co., Minneapolis.

9. **Fakhr, M. K., L. K. Nolan, and C. M. Logue.** 2005. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. J. Clin. Microbiol. **43:**2215-2219.

10. **Fontana, J., A. Stout, B. Bolstorff, and R. Timperi.** 2003. Automated ribotyping and pulsed-field gel electrophoresis for rapid identification of multidrug-resistant *Salmonella* serotype newport. Emerg. Infect. Dis. **9:**496-9.

11. **Garaizar, J., S. Porwollik, A. Echeita, A. Rementeria, S. Herrera, R. M. Y. Wong, J. Frye, M. A. Usera, and M. McClelland.** 2002. DNA microarray-based typing of an atypical monophasic *Salmonella enterica* serovar. J. Clin. Microbiol. **40:**2074-2078.

12. **Graves, L. M., and B. Swaminathan.** 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. Int. J. Food Microbiol. **65:**55-62.

13. **Grimont, P., Weil, F.** 2007. Antigenic Formulae of the *Salmonella* Servovars, 9th ed, vol. Pasteur Institute, Paris: World Health Organization Centre for Reference and Research on *Salmonella*.

14. **Harbottle, H., D. G. White, P. F. McDermott, R. D. Walker, and S. Zhao.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J. Clin. Microbiol. **44:**2449-2457.

15. **Hudson, C. R., C. Quist, M. D. Lee, K. Keyes, S. V. Dodson, C. Morales, S.**

Sanchez, D. G. White, and J. J. Maurer. 2000. Genetic relatedness of *Salmonella* isolates from nondomestic birds in Southeastern United States. J. Clin. Microbiol. **38:**1860-1865.

16. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465-6.

17. **Hunter, S. B., P. Vauterin, M. A. Lambert-Fair, M. S. Van Duyne, K. Kubota, L. Graves, D. Wrigley, T. Barrett, and E. Ribot.** 2005. Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard. J. Clin. Microbiol. **43:**1045-1050.

18. **Hyytia-Trees, E., S. C. Smole, P. A. Fields, B. Swaminathan, and E. M. Ribot.** 2006. Second generation subtyping: A proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). Foodborne Pathog. Dis. **3:**118-131.

19. **Langvad, B., M. N. Skov, E. Rattenborg, J. E. Olsen, and D. L. Baggesen.** 2006. Transmission routes of *Salmonella* Typhimurium DT 104 between 14 cattle and pig herds in Denmark demonstrated by molecular fingerprinting. J. Appl. Microbiol. **101:**883-890.

20. **Liesegang, A., D. Davos, J. C. Balzer, W. Rabsch, R. Prager, D. Lightfoot, A. Siitonen, H. Claus, and H. Tschape.** 2002. Phage typing and PFGE pattern analysis as tools for epidemiological surveillance of *Salmonella enterica* serovar Bovismorbificans infections. Epi. Infec. **128:**119-130.

21. **Lim, H., K. H. Lee, C. H. Hong, G. J. Bahk, and W. S. Choi.** 2005. Comparison of four molecular typing methods for the differentiation of

Salmonella spp. Int. J. Food Microbiol. **105:**411-8.

22. **Lukinmaa, S., U. M. Nakari, A. Liimatainen, and A. Siitonen.** 2006. Genomic diversity within phage types of *Salmonella enterica* ssp. enterica serotypes Enteritidis and Typhimurium. Foodborne Pathog. Dis. **3:**97-105.

23. **Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. **95:**3140-3145.

24. **Malorny, B., E. Junker, and R. Helmuth.** 2008. Multi-locus variable-number tandem repeat analysis for outbreak studies of *Salmonella enterica* serotype Enteritidis. BMC Microbiol. **8:**84.

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

26. **Murase, T., M. Nagato, K. Shirota, H. Katoh, and K. Otsuki.** 2004. Pulsed-field gel electrophoresis-based subtyping of DNA degradation-sensitive *Salmonella enterica* subsp *enterica* serovar Livingstone and serovar Cerro isolates obtained from a chicken layer farm. Vet. Microbiol. **99:**139-143.

27. **Nagano, N., S. Oana, Y. Nagano, and Y. Arakawa.** 2006. A severe *Salmonella enterica* serotype Paratyphi B infection in a child related to a pet turtle, Trachemys scripta elegans. Jpn. J. Infect. Dis. **59:**132-134.

28. **Ogilvie, T. H.** 1986. The Persistent Isolation of *Salmonella* typhimurium from the Mammary Gland of a Dairy Cow. Can. Vet. J. **27:**329-331.

29. **Pedersen, T. B., J. E. Olsen, and M. Bisgaard.** 2008. Persistence of *Salmonella* Senftenberg in poultry production environments and investigation

of its resistance to desiccation. Avian Pathol. **37:**421-7.

30. **Rabsch, W., H. L. Andrews, R. A. Kingsley, R. Prager, H. Tschape, L. G. Adams, and A. J. Baumler.** 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. Infect. Immun. **70:**2249-2255.

31. **Reen, F. J., E. F. Boyd, S. Porwollik, B. P. Murphy, D. Gilroy, S. Fanning, and M. McClelland.** 2005. Genomic comparisons of *Salmonella enterica* serovar Dublin, Agona, and Typhimurium strains recently isolated from milk filters and bovine samples from Ireland, using a *Salmonella* microarray. Appl. Environ. Microbiol. **71:**1616-1625.

32. **Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J. Barrett.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of Escherichia coli O157:H7, *Salmonella*, and Shigella for PulseNet. Foodborne Pathog. Dis. **3:**59-67.

33. **Ridley, A. M., E. J. Threlfall, and B. Rowe.** 1998. Genotypic characterization of *Salmonella* enteritidis phage types by plasmid analysis, ribotyping, and pulsed-field gel electrophoresis. J. Clin. Microbiol. **36:**2314-2321.

34. **Ross, I. L., and M. W. Heuzenroeder.** 2005. Discrimination within phenotypically closely related definitive types of *Salmonella enterica* serovar typhimurium by the multiple amplification of phage locus typing technique. J. Clin. Microbiol. **43:**1604-1611.

35. **Sahlstrom, L., J. B. de, and A. Aspan.** 2006. *Salmonella* isolated in sewage sludge traced back to human cases of salmonellosis. Lett. Appl. Microbiol. **43:**46-52.

36. **Scaria, J., R. U. Palaniappan, D. Chiu, J. A. Phan, L. Ponnala, P. McDonough, Y. T. Grohn, S. Porwollik, M. McClelland, C. S. Chiou, C.**

Chu, and Y. F. Chang. 2008. Microarray for molecular typing of *Salmonella enterica* serovars. Mol. Cell Probes **22:**238-43.

37. **Seo, Y. S., S. H. Lee, E. K. Shin, S. J. Kim, R. Jung, and T. W. Hahn.** 2006. Pulsed-field gel electrophoresis genotyping of *Salmonella* gallinarum and comparison with random amplified polymorphic DNA. Vet. Microbiol. **115:**349-57.

38. **Sukhnanand, S., S. Alcaine, L. D. Warnick, W. L. Su, J. Hof, M. P. Craver, P. McDonough, K. J. Boor, and M. Wiedmann.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes. J. Clin. Microbiol. **43:**3688-3698.

39. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infec. Dis. **7:**382-389.

40. **Swaminathan, B., P. Gerner-Smidt, and T. Barrett.** 2006. Focus on *Salmonella*. Foodborne Pathog. Dis. **3:**154-156.

41. **Swaminathan, B., P. Gerner-Smidt, L. K. Ng, S. Lukinmaa, K. M. Kam, S. Rolando, E. P. Gutierrez, and N. Binsztein.** 2006. Building PulseNet International: an interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. Foodborne Pathog. Dis. **3:**36-50.

42. **Uzzau, S., D. J. Brown, T. Wallis, S. Rubino, G. Leori, S. Bernard, J. Casadesus, D. J. Platt, and J. E. Olsen.** 2000. Host adapted serotypes of *Salmonella enterica*. Epidemiol. Infec. **125:**229-255.

43. **Valdezate, S., M. Arroyo, R. Gonzalez-Sanz, R. Ramiro, S. Herrera-Leon, M. A. Usera, M. De la Fuente, and A. Echeita.** 2007. Antimicrobial resistance and phage and molecular typing of *Salmonella* strains isolated from

food for human consumption in Spain. J. Food Prot. **70:**2741-8.

44. **Vanselow, B. A., S. Hum, M. A. Hornitzky, G. J. Eamens, and K. Quinn.** 2007. *Salmonella* Typhimurium persistence in a Hunter Valley dairy herd. Aust. Vet. J. **85:**446-50.

45. **Voetsch, A. C., T. J. Van Gilder, F. J. Angulo, M. M. Farley, S. Shallow, R. Marcus, P. R. Cieslak, V. C. Deneen, and R. V. Tauxe.** 2004. FoodNet estimate of the burden of illness caused by nontyphoidal *Salmonella* infections in the United States. Clin. Infect. Dis. **38 Suppl 3:**S127-S134.

46. **Wiedmann, M.** 2002. Molecular subtyping methods for Listeria monocytogenes. J. AOAC Inter. **85:**524-531.

47. **Winokur, P. L.** 2003. Molecular epidemiological techniques for *Salmonella* strain discrimination. Front. Biosci. **8:**c14-24.

48. **Wonderling, L., R. Pearce, F. M. Wallace, J. E. Call, I. Feder, M. Tamplin, and J. B. Luchansky.** 2003. Use of pulsed-field gel electrophoresis to characterize the heterogeneity and clonality of *Salmonella* isolates obtained from the carcasses and feces of swine at slaughter. Appl. Environ. Microbiol. **69:**4177-82.

49. **Woo, Y. K., and S. H. Lee.** 2006. Genetic diversity of multi-resistant *Salmonella enterica* serotype Typhimurium isolates from animals and humans. J. Microbiol. **44:**106-112.

50. **Zamperini, K., V. Soni, D. Waltman, S. Sanchez, E. C. Theriault, J. Bray, and J. J. Maurer.** 2007. Molecular characterization reveals *Salmonella enterica* serovar 4,[5],12:i:- from poultry is a variant Typhimurium serovar. Avian Dis. **51:**958-64.

CHAPTER 3

GENOME WIDE EVOLUTIONARY ANALYSES REVEAL SEROTYPE SPECIFIC PATTERNS OF POSITIVE SELECTION IN SELECTED *SALMONELLA* SEROTYPES

ABSTRACT

The bacterium *Salmonella enterica* includes a diversity of serotypes that cause disease in humans and different animal species. Some *Salmonella* serotypes show a broad host range, some are host restricted and exclusively associated with one particular host, and some are associated with one particular host species, but able to cause disease in other host species and are thus considered “host adapted”. Five available *Salmonella* genome sequences, representing a broad host range serotype (Typhimurium), two host restricted serotypes (Typhi [two genomes], Paratyphi) and one host adapted serotype (Choleraesuis) were used to identify core genome genes that show evidence for recombination and positive selection. Overall, 3323 orthologous genes were identified in all 5 *Salmonella* genomes analyzed. Use of four different methods to assess homologous recombination identified 270 genes that showed evidence for recombination with at least one of these methods (false discovery rate [FDR] <10%). Site and branch specific models identified 41 genes as showing evidence for positive selection (FDR <20%), including a number of genes with confirmed or likely roles in virulence and genes encoding outer membrane proteins, which have also been found to be under positive selection in other bacteria. A total of 8, 16, 7, and 5 genes with no evidence of recombination were found to be under positive selection in Choleraesuis, Typhi, Typhimurium, and Paratyphi branch analyses, respectively. Sequencing and evolutionary analyses of four genes in an additional 42 isolates representing 23 serotypes confirmed branch specific positive

selection and recombination patterns. Our data show that (i) less than 10% of *Salmonella* genes in the core genome show evidence for homologous recombination, (ii) a number of *Salmonella* genes are under positive selection, including genes that appear to contribute to virulence, and (iii) branch specific positive selection contributes to the evolution of different *Salmonella* serotypes, including host restricted serotypes.

INTRODUCTION

Salmonella is a ubiquitous human and animal pathogen and the genus *Salmonella* is divided into two species, *Salmonella bongori* and *Salmonella enterica*. *S. enterica* consists of six subspecies (i.e., *enterica, salamae*, *arizonae, diarizonae, houtenae,* and *indica*) (7). Phylogenetic analyses suggest that *Salmonella* and *Escherichia coli* diverged from a common ancestor about 100 million years ago (5). After the divergence from *E. coli*, *Salmonella* gained virulence associated gene complexes, commonly called "Salmonella Pathogenicity Island (SPI)" (54).

The genus *Salmonella* contains >2,500 recognized serotypes. *Salmonella enterica* subsp. *enterica* serotypes can also be divided into subdivisions according to their host adaptation (71). For example, Uzzau et al. (2000) proposed that *Salmonella* serotypes can be divided into (i) host-restricted *Salmonella* serotypes (i.e., serotypes exclusively associated with one particular host, e.g. *Salmonella* Typhi and Paratyphi A); (ii) host-adapted *Salmonella* serotypes (i.e., serotypes prevalent in one particular host species, but able to cause disease in other host species, e.g. *Salmonella* Choleraesuis); and (iii) unrestricted *Salmonella* serotypes (i.e., serotypes capable of causing self-limiting gastroenteritis and, less commonly, systemic disease in a wide range of host species, e.g. *Salmonella* Typhimurium).

Multi-locus sequence typing (MLST) data indicate that the last common ancestor of the human host-adapted *Salmonella* Typhi existed 15,000-150,000 years

ago (27). The evolution of *Salmonella* Typhi towards a lifestyle characterized by systemic infection and transmission by excretion through the gall bladder rather than luminal gut colonization (5) involved a combination of acquisition events (e.g., acquisition of Vi capsule related genes), and deletion events (e.g., loss of virulence-associated genes, such as several genes in SPI-1, SPI-2, SPI-3, SPI-4 and SPI-5). *Salmonella* Paratyphi A also causes typhoid fever, although disease is typically milder than that caused by *Salmonella* Typhi. While *Salmonella* Paratyphi A also appears to have evolved recently, *Salmonella* Typhi and Paratyphi A clearly show distinct differences in their genome evolution, including a number of unique gene inactivation events in these two serotypes (39).

Nontyphoidal *Salmonella* serotypes are responsible for gastroenteritis in humans and other animals. These serotypes are mainly transmitted by ingestion of food, feed, or water contaminated with infected feces (58), but can also be transmitted by direct contact (33, 45). Disease caused by non-typhoidal *Salmonella* is one of the most common bacterial foodborne diseases worldwide (65). In the United States nontyphoidal *Salmonella* serotypes have been estimated to cause >1.4 million human salmonellosis cases with >16,000 hospitalizations and almost 600 deaths annually (41). *Salmonella* Typhimurium is one of the most common *Salmonella* serotypes, is found worldwide, and can cause disease, predominantly self limiting gastroenteritis, in a large number of animal species (71). The host adapted *Salmonella* Choleraesuis can cause severe disease, characterized by septicemia and enterocolitis, in swine. While relatively uncommon, this serotype can also infect humans where it typically causes severe invasive infections, e.g., infective aneurysm (15).

Gene acquisition and deletion events clearly play an important role in evolution. The importance of acquisition of novel (non-homologous) genes by lateral gene transfer has been clearly demonstrated in a number of bacteria, including a number of bacterial pathogens (21, 35, 50, 54). Acquisition of pathogenicity islands

has played a critical role in the evolution of *Salmonella* (54) and other Gram-negative and Gram-positive pathogens (61). Gene degradation and gene deletions also have been shown to play a critical role in bacterial evolution, particularly when organisms with a broad niche specificity adapt to narrow and specific ecological niches (39, 67). For example, it has been suggested that gene degradation and gene deletion contribute to host adaptation in both *Salmonella* Typhi and *Salmonella* Paratyphi A (39). Microarray technologies have also allowed for rapid and large scale studies on gene presence/absence in large numbers of isolates, including in *Salmonella* (53). In addition to gene acquisition and deletion, positive selection and homologous recombination, play important roles in the evolution of bacteria and bacterial pathogens (14, 34, 48).

Genome wide studies on positive selection and recombination in bacterial pathogens, including *Streptococcus* spp. (34), *Listeria monocytogenes* (48), *E. coli* (13, 52), and *Shigella* (52) have contributed to a better understanding of the evolution of these important pathogens. So far, no genome wide analyses of positive selection in *Salmonella* have been reported. One study (12) evaluated 410 genes present in both *S. enterica* and *E. coli*, though, and reported that 50% of amino acid substitutions in these genes appear to have been fixed by positive selection in one of these species. In order to further improve our understanding of the evolution of *Salmonella*, we thus performed full genome analyses for recombination and positive selection using the completed and published genome sequences for five *Salmonella*, including the host restricted *Salmonella* Typhi (two strains) and Paratyphi A, the host adapted *Salmonella* Choleraesuis, and the broad-host range *Salmonella* Typhimurium. We specifically hypothesized that analysis of these *Salmonella* serotypes would provide an improved understanding in the roles of positive selection and recombination in the evolution of host-adapted pathogen strains. A particular focus of our study was thus to evaluate lineage specific positive selection in different *Salmonella* serotypes.

METHODS AND MATERIALS

Genome sequences. Five available annotated *Salmonella enterica* subsp. *enterica* genome sequences were used in this study (Table 3.1). Genome sequences were downloaded from the Comprehensive Microbial Resource at The Institute for Genomic Research (TIGR; now J. Craig Venter Institute, JCVI) on November 25, 2005. Updated role category information for all genes was obtained from JCVI on October 14, 2008; the *Salmonella* Typhi CT18 genome was used as reference for role categories.

Identification of orthologous genes presents in all five *Salmonella* genomes analyzed. OrthoMCL (36) was used to identify orthologous genes in the five *Salmonella* genomes. Orthologs present in all five genomes were aligned using ClustalW (66). Multiple sequence alignments were carried out on amino acid sequences from each orthologous group, followed by conversion to nucleotide sequence alignments using the PAL2NAL software (64). Alignments containing variable sequence lengths or with low alignment scores were manually evaluated and edited, using BioEdit software (23), as previously described (48).

Detection of genes under positive selection. Positive selection can be detected by comparing the rate of non-synonymous substitutions (d_N) to the rate of synonymous substitutions (d_S). Among the programs that can be used to identify positive selection from DNA sequences, PAML (Phylogenetic Analysis by Maximum Likelihood) has been widely used to detect positive selection in bacteria (3, 13, 34, 48, 52, 70), viruses (69), and eukaryotes (11, 47). We thus used two types of tests implemented in PAML v3.15 to identify genes with evidence for positive selection (75), as previously detailed (48). Briefly, an overall test for positive selection (Test

Overall; TO) was carried out to identify genes under positive selection in any or all of the branches of a given phylogeny; this test compares the null model M1a (Nearly-neutral) to the alternative model M2a (positive selection) (73). To identify genes that are under positive selection in specific branches of the *Salmonella* phylogeny, the branch-site test2 (77) was used. The branch-site test was specifically used to identify genes under positive selection in the ancestral branches of (i) the human restricted serotypes Typhi (Ty#) and (ii) Paratyphi A (Pty#), (iii) the porcine adapted serotype Choleraesuis (Ch#), and (iv) the unrestricted serotype Typhimurium (Tym#) (Figure 3.1). Overall, 18 different phylogenetic trees represented the phylogeny of the 3316 *Salmonella* orthologous genes, including one tree that represented the phylogeny of 1198 genes. Both the overall test and the branch site tests were performed using the gene specific phylogenetic tree for each gene.

For each test, nested models (one null model that does not allow for positive selection and one alternative model that allows for positive selection) were compared using a Likelihood Ratio Test (LRT) (76). For each model, three replicates were generated and the maximum likelihood values for each model were used in the LRT in order to eliminate the runs that could not reach the global maximum likelihood score. Tests that yielded LRT values < -0.1were re-run 10 times and the maximum values for each model were used to calculate the LRT. Negative LRT values (i.e., some tests yielded values $\geq - 0.1$) were rounded to zero (*P*-value=1). For all branch-specific tests, one degree of freedom was used to calculate *p*-values, while for the overall test, two degrees of freedom were used to calculate *p*-values.

Detection of genes with evidence of recombination. GENECONV version 1.81 (60), Maximum χ^2 (62), pairwise homoplasy index (PHI) (8) and neighbor similarity score (NSS) (25) were used to determine which of the 3316 orthologous genes

Table 3. 1. *Salmonella* genomes used in this study

Serotype	No. of ORFs	Accession No.	Sequencing Center	Reference
Choleraesuis	4801	NC_006905	Chang Gung Univ.	Chiu et al. 2005
Paratyphi A	4093	NC_006511	Washington Univ.	McClelland et al. 2004
Typhi CT18	4395	NC_003198	Sanger Centre/Imperial College	Parkhill et al. 2001
Typhi Ty2	4323	NC_004631	Univ. of Wisconsin	Deng et al. 2003
Typhimurium	4553	NC_003197	Washington University Consort.	McClelland et al. 2001

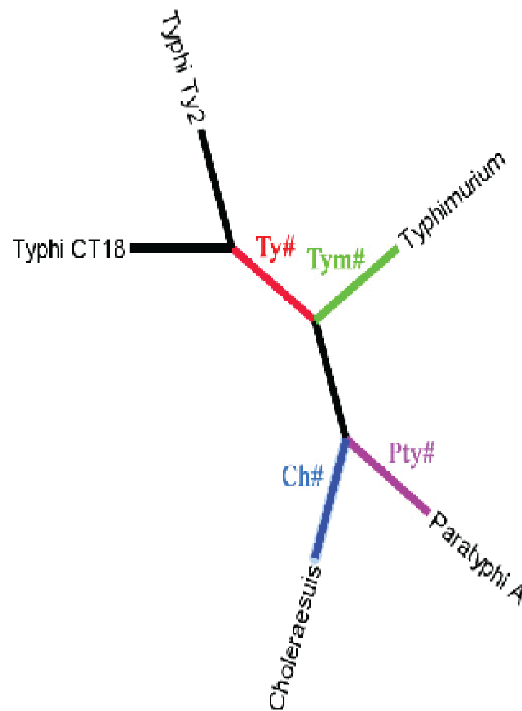


Figure 3.1. Example of neighbor joining tree used for positive selection analysis. Gene specific trees were used for all positive selection analysis. The tree shown here represented the phylogeny of 849 genes. Branches used for branch specific analyses are indicated; Ch#= Choleraesuis branch specific test; Ty# = Typhi branch specific test; Tym#= Typhimurium branch specific test; Pty#= Paratyphi A branch specific test.

showed evidence of recombination. The last three approaches are implemented in PhiPack (8). For the GENECONV analyses, the parameter g-scale was set to 1 and inner *p*-values were used to identify genes with evidence for recombination (60). For Maximum χ^2, a fixed window size of 2/3 the number of polymorphic sites was used, while for PHI, a window size of 50 nucleotides was used. *P*-values were estimated using 10,000 permutations of the alignment for GENECONV and 1,000 permutations for NSS, Maximum χ^2 and PHI.

Assessment of codon bias. To assess the codon bias, we identified the effective number of codons used in a gene (N_C) using the program "chips" in the EMBOSS package (55). Nucleotide diversity and number of informative sites were obtained from PhiPack outputs.

Statistical analyses. Correction for multiple testing was performed using the procedure reported by Benjamini and Hochberg (6) as implemented in the program Q-Value (63). As previously detailed by our group (48), for each *p*-value, the *q*-value was calculated; the q-value represents the false discovery rate [FDR], i.e., the expected proportion of false positives among the significant tests. Corrections were performed separately for each test to account for testing of multiple genes . In an initial analysis of positive selection, all 3316 genes were used for FDR correction. A second correction was then performed including only those 3046 genes that showed no evidence for recombination. Genes with evidence for positive selection in the first analysis and evidence for recombination are treated as possible false positives. As the tests used for positive selection are already conservative (77), a false discovery rate (FDR) of 20% was used for the positives selection analyses (48). For recombination analyses, an FDR of 10% was used to compensate the fact that no correction for

multiple tests (GENECONV, NSS, Maximum χ^2 and PHI) was carried out due to the high correlation among the tests (48).

Associations between JCVI role categories and number of genes with (i) evidence of positive selection and (ii) evidence of recombination were tested using chi-square tests (or Fisher's exact tests where appropriate). Mann-Whitney U-tests (Wilcoxon tests) were used to determine whether selected continuous variables (i.e., gene length, codon bias, and nucleotide diversity) differed between a given role categories and all other role categories. In addition, Mann-Whitney U-tests were used to test whether the *p*-values of the positive selection tests for genes in a given role category were significantly lower than the *p*-values among the genes in the other role categories. All Mann-Whitney U-tests were performed as one-sided tests. Bonferroni corrections for all tests were performed based on the number of tests performed. All tests were performed in SAS. The cut off value for significance was set at 0.05; actual Bonferroni corrected *p*-values are reported unless otherwise stated. Actual *p*-values are reported unless *p*-values were <0.001 or <0.0001.

Verification of positive selection and recombination patterns in selected genes in a larger *Salmonella* set .For four genes that were identified as being under positive selection in the initial genome wide analyses, gene sequences were determined for an additional 42 *Salmonella* isolates to verify positive selection and recombination patterns in these genes. The 42 *Salmonella* isolates were selected to reflect a diversity of human and animal associated serotypes; specifically, the isolates were selected to represent the 15 most common human and animal associated serotypes in the US [as detailed in the 2003 *Salmonella* Annual Report from the US Centers of Disease Control and Prevention (10)] as well as two additional *Salmonella* Typhi isolates. Human and cattle isolates representing the common human and animal associated serotypes were selected from the strain collection available at Cornell. For

common serotypes (e.g., Typhimurium) more isolates were included in this set as compared to less common serotypes (e.g., Dublin) (see Supplemental Table S1 for a listing of all isolates used). Multiple isolates with the same serotype were selected to represent the most common distinct Pulse Field Gel Electrophoresis (PFGE) and Multilocus sequence typing (MLST) types within a given serotype.

The four genes that were used to confirm positive selection and recombination patterns identified in the full genome analyses included *folK-2*, *sseC*, *purE*, and STM3258 (Table 3.2). PCR conditions and primers are described in Supplemental Table S2. PCR products were purified using Exonuclease I (USB) and shrimp alkaline phosphatase (USB). Purified PCR products were sequenced using the Applied Biosystems Automated 3730 DNA Analyzer at the Cornell University Life Sciences Core Laboratories Center. Big Dye Terminator chemistry and AmpliTaq-FS DNA Polymerase were used for sequencing. Alignments for positive selection and recombination analyses, which were performed as detailed above, were constructed using the gene sequences for the five genomes analyzed and the gene sequences for the additional isolates sequenced.

RESULTS

Initial identification and characterization of orthologous genes present in the five *Salmonella* genomes representing serotypes Typhi, Typhimurium, Choleraesuis, and Paratyphi A. Using OrthoMCL, a total of 3323 orthologous genes present in all 5 *Salmonella* genomes were identified. Since seven orthologous genes had low quality alignments, we excluded these genes and used 3316 orthologous genes for the analyses described below. Genes that were not found in all of the five strains were excluded from our analyses. The 3316 core genes represented 69, 81, 73 and 75 %, respectively, of the Salmonella Choleraesuis, Paratyphi A, Typhimurium, and Typhi genes annotated in the genomes analyzed.

Table 3. 2. Genes used to confirm positive selection and recombination patterns identified in genome wide analyses

Gene Name	Protein name	JCVI Role Category	Gene length (bp)	Genome analyses results for		Sequence analyses results (based on 47 sequences) for[c]	
				Positive Selection[a]	Recombination[b]	Positive Selection[d]	Recombination[e]
folK-2	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	Biosynthesis of cofactors, prosthetic groups, and carriers	480	TO, Ty#	GEN, MAX	Ty#	GEN, MAX, NSS
STM3258	Putative PTS system IIA component	Transport and binding proteins	465	Ty#	-	Ty#	-
sseC	Probable pathogenicity island effector protein	Unclassified	1455	Ch#	GEN, MAX	TO, Ch#	GEN, MAX, NSS, PHI
purE	Phosphoribosylamino-imidazole carboxylase, catalytic subunit	Purines, pyrimidines, nucleosides, and nucleotides	510	Ty#	-	Ty#	NSS

[a]positive selection tests that were significant (q<0.2) are listed; TO = overall test; Ch#= Choleraesuis branch specific test; Ty# = Typhi branch specific test

[b]recombination tests that were significant (q<0.1) are listed; GEN = GENECONV; MAX= Maximum χ^2; PHI= pairwise homoplasy; NSS= neighbor similarity

[c]Results of positive selection and recombination analyses were based on gene sequence data for the 5 genomes and 42 additional *Salmonella* isolates (see Supp. Table 1); for *flolK-2* and *sseC* sequences were only obtained for 36 additional isolates; for STM3258 sequences were only obtained for 37 additional isolates.

[d]positive selection tests that were significant ($P< 0.05$)

[e]recombination tests that were significant ($P< 0.05$)

Interestingly, we identified one 2-gene cluster (i.e., STM0947 and STM0948), which was repeated 12 times in the *Salmonella* Choleraesuis genome, present once in Typhimurium genome and absent in the Typhi and Paratyphi A genomes. These two genes encode a putative integrase (STM0947) and a putative cytoplasmic protein (STM0948), which differ by 4 and 1 non-synonymous substitution(s), respectively, between Choleraesuis and Typhimurium LT2. In addition, we identified one other gene (NT03ST2087, encoding a putative Tn10 transposase), which was repeated 7 times in the *Salmonella* Choleraesuis and found once in the *Salmonella* Typhi CT18, while not present in the other genomes analyzed. *Salmonella* Choleraesuis thus appears to contain at least two multicopy mobile genetic elements.

Genes categorized in the JCVI role categories "Hypothetical Proteins", "Protein synthesis", "Unclassified" and "Unknown function" showed a tendency to have shorter alignments ($P<0.001$, $P=0.027$, $P=0.002$, $P=0.017$, respectively; one sided U-test), while genes in the JCVI role categories "Amino Acid Biosynthesis", "DNA Metabolism", "Energy Metabolism", and "Transport and Binding Proteins" showed a tendency to have longer alignments ($P<0.001$, $P=0.001$, $P<0.001$, and $P<0.001$, respectively; one sided U-test).

Genes in the JCVI role categories "Cellular envelope", "Hypothetical proteins", and "Unclassified" showed a tendency to have more non-synonymous substitutions ($P=0.009$, $P<0.001$, and $P<0.001$, respectively, one sided U-test). Genes in the JCVI role categories "Biosynthesis of cofactors, prosthetic groups, and carriers", "Energy Metabolism", and "Transport and Binding Proteins" showed a tendency to have more synonymous substitutions ($P<0.001$, $P<0.001$, and $P=0.001$, respectively, one sided U-test). Genes in the JCVI role categories "Amino acid biosynthesis", "Energy metabolism", "Protein Synthesis", "Purines, pyrimidines, nucleosides, and nucleotides", "Transcription", and "Transport and binding proteins"

showed a tendency to have higher codon bias (i.e., low chip values [N_C]; *P*=0.006, *P*<0.001, *P*<0.001, *P*<0.001, *P*=0.033, and *P*=0.010, respectively; one sided U-test).

Approximately 8% of core genes show significant evidence for recombination. Among the 3316 orthologous genes, 233 genes showed no substitutions; these genes thus were not analyzed for evidence of recombination. While the remaining 3083 genes were analyzed for recombination using GENECONV, only 2849 genes were analyzed using Max χ^2, NSS and Phi (467 orthologs have ≤1 informative site and thus could not be analyzed with these programs in PhiPack).

Overall, 270 genes (8.14 % of all 3,316 core genes) showed evidence for recombination in at least one of the four tests used (FDR < 10%). A total of 192, 155, 69, and 20 orthologs showed evidence of recombination using GENECONV, Max χ^2, NSS and Phi, respectively. Interestingly, only 10 genes showed evidence for recombination with all 4 approaches (Table 3.3). Genes with higher numbers of informative sites (*P* <0.0001; one sided U-test), longer alignments (*P*-value <0.0001; one sided U-test), higher codon bias (*P*-value <0.0001; one sided U-test), and higher nucleotide diversity (*p*-value <0.0001; one sided U-test) were more likely to have evidence for recombination.

An overall chi-square test showed that genes with evidence of recombination were not randomly distributed among the 20 JCVI role categories (*P*<0.001; Fisher's exact test with Monte Carlo simulation). Subsequent individual chi-square tests, determining whether genes with evidence for recombination were associated with individual role categories, showed that genes with evidence of recombination were

Table 3. 3. Genes that show evidence of recombination in all four tests[a]

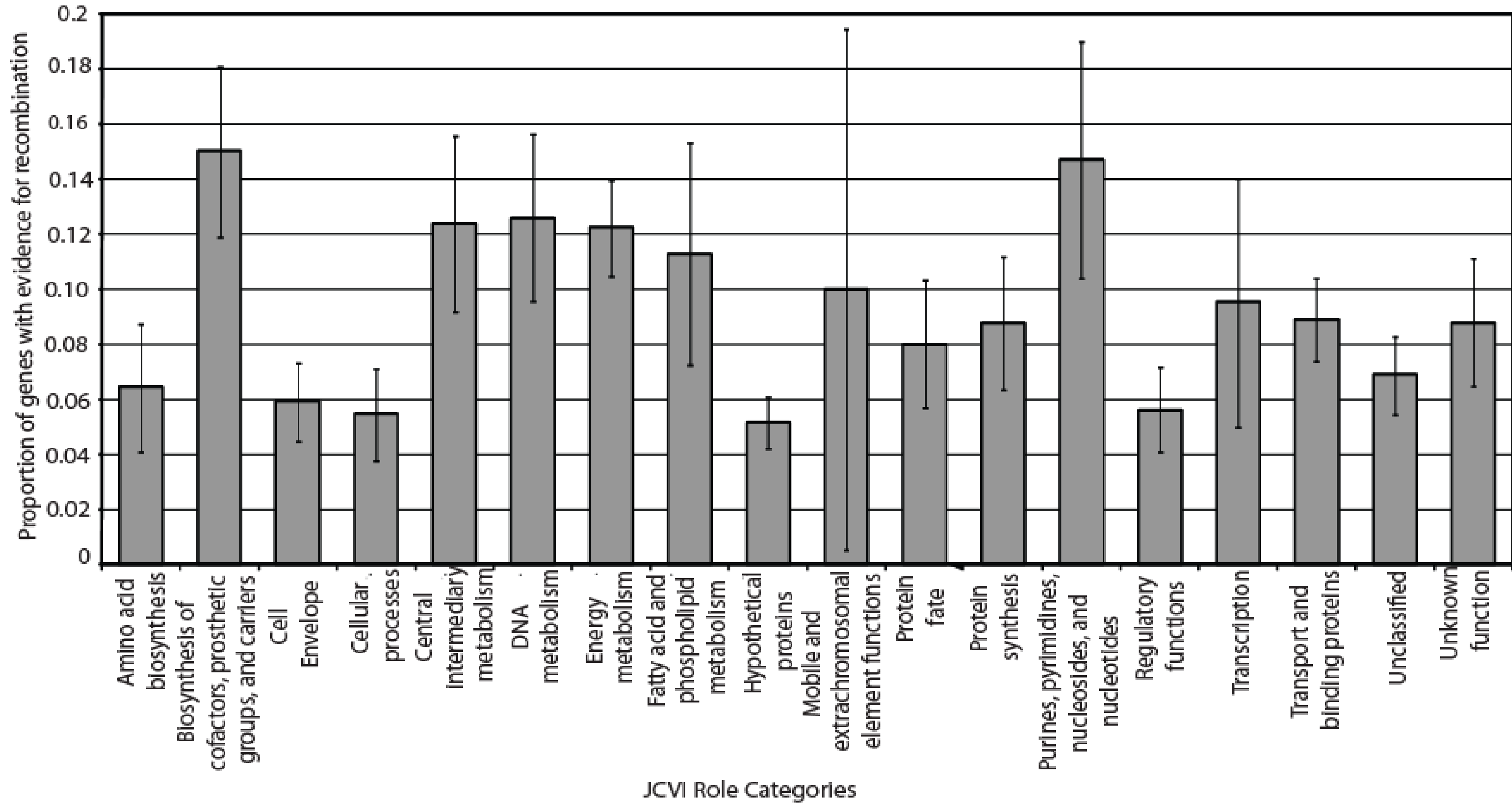
Gene annotation no. for *S*. Typhimurium LT2	Protein name	Gene name	JCVI Role Category
STM0067	Carbamoyl-phosphate synthase, large subunit	*carB*	Purines, pyrimidines, nucleosides, and nucleotides
STM0224	Surface antigen	*b0177*	Unknown function
STM0540	Conserved hypothetical protein	-	Hypothetical proteins
STM0661	Inosine-uridine preferring nucleoside hydrolase	*iunH*	Purines, pyrimidines, nucleosides, and nucleotides
STM2287	Conserved hypothetical protein	-	Hypothetical proteins
STM2660	ATP-dependent protease, Hsp 100, part of novel	*clpB-1*	Protein fate
STM2947	Sulfite reductase (nADPh) hemoprotein beta-component	-	Central intermediary metabolism
STM2948	Sulfite reductase (nADPh) flavoprotein alpha-component	-	Central intermediary metabolism
STM3174	DNA topoisomerase IV, A subunit	*parC*	DNA metabolism
STM4066	Fructokinase	*cscK*	Energy metabolism

[a]These genes showed evidence for recombination (Q<0.1) in four tests (i.e., GENECONV, Maximum χ^2 (MAX), pairwise homoplasy index (PHI), and neighbor similarity score (NSS)

significantly overrepresented in the role categories "Biosynthesis of cofactors, prosthetic groups, and carriers", "Energy metabolism", "Hypothetical proteins" and "Purines, pyrimidines, nucleosides, and nucleotides" (uncorrected *P*= 0.0035, *P*= 0.0037, *P*= 0.0034, and *P*= 0.0493, respectively, chi square test) (Figure 3.2). However, after corrections for multiple comparisons, the associations are not significant (*P*= 0.063, *P*= 0.066, *P*= 0.061, and *P*= 0.887, respectively, Bonferroni correction).

Initial analysis revealed a total of 81 *Salmonella* genes showing evidence for positive selection. Among the 3316 orthologous genes identified, 328 genes did not contain any non-synonymous substitutions. We used the gene specific phylogenetic tree for each gene in our analysis. A total of 21 genes showed evidence for positive selection (FDR <20%) in the overall test (TO) (Supplementary Table 3). Since the two Typhi isolates formed a single branch in only 1261 genes, we only used these 1261 genes to test for positive selection in the Typhi branch. A total of 23, 21, 13, and 14 genes, respectively, showed evidence of positive selection (FDR <20%), using the branch-site test, in the Choleraesuis, Typhi, Typhimurium, and Paratyphi A branch (Supplementary Table 3). Overall, 81 genes showed evidence of positive selection in at least one test. Among these 81 genes with evidence of positive selection, 32 genes also showed evidence of recombination with at least one of the four recombination tests used (Table 3.6) (Supplementary Table 3). Statistical analyses showed that genes with evidence of recombination were more likely to be under positive selection (*P*-value <0.0001; Chi- square test). While this may indicate that positive selection contributes to fixation of new allelic variants that were generated by recombination (48), it may also reflect that the positive selection tests used were affected by intragenic recombination (4). Thus, positive selection analyses were repeated without 270 genes with evidence of recombination. To test for positive selection, 3046 genes with no recombination evidence were used in the overall (TO)

Figure 3. 2. Proportions of genes with evidence of recombination among individual JCVI role categories. Genes that showed evidence for recombination ($q<0.1$) in at least one of the four tests were included. Bars indicate estimated standard error for the proportion of genes with evidence of recombination in each role category; standard errors were calculated as square root of p (1-p)/ n, where p is the frequency of genes with evidence of positive selection in a given role category, and n is the total number of genes in a given role category. Among the 20 JCVI role categories, two did not include genes with evidence of recombination (i.e., "Signal Transduction" and "Viral functions") and are thus not included in this figure.


Proportion of genes with evidence for recombination
0.2
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
Amino acid biosynthesis
Biosynthesis of cofactors, prosthetic groups, and carriers
Cell Envelope
Cellular processes
Central intermediary metabolism
DNA metabolism
Energy metabolism
Fatty acid and phospholipid metabolism
Hypothetical proteins
Mobile and extrachromosomal element functions
Protein fate
Protein synthesis
Purines, pyrimidines, nucleosides, and nucleotides
Regulatory functions
Transcription
Transport and binding proteins
Unclassified
Unknown function
JCVI Role Categories

test and the branch tests of Choleraesuis, Typhimurium and Paratyphi, while 1108 genes with no evidence of recombination were used in Typhi branch test. All data in the subsequent sections represent the data for genes with no evidence recombination, unless otherwise stated.

A total of 41 *Salmonella* genes with no evidence of recombination showed evidence of positive selection. We found 5 genes with evidence for positive selection (FDR <20%) in the overall test (TO) (Table 3.4). A total of 8, 16, 7, and 5 genes, respectively, showed evidence of positive selection (FDR <20%), using the branch-site test, in the Choleraesuis, Typhi, Typhimurium, and Paratyphi A branches (Table 3.4). None of these genes showed of evidence of positive selection in more than 1 test. No association between the low effective number of codons used by a gene (Nc) and positively selected genes was observed (P= 0.4276; one-sided U-test) suggesting that results of positive selection were not biased by constrains on codon usage, which could result in a low synonymous substitution rate in these genes. Moreover, no association between low dS (the number of synonymous substitutions divided by the number of synonymous sites) and positively selected genes was observed (p-value =0.999 ; one-sided, U-test), supporting that the results were not biased by a low synonymous substitution rate .In order to initially determine whether specific JCVI role categories are more likely to include genes under positive selection, we used a contingency table to test for associations between the 20 JCVI role categories and the 41 genes under positive selection (Figure 3.3). This test did not find any significant association between JCVI role categories and number of genes with evidence of positive selection. A significant association was observed between JCVI role categories "Purines, pyrimidines, nucleosides, and nucleotides" and "Cell

Table 3. 4. Genes with evidence for positive selection

Gene annotation no. for *S.* Typhimurium LT2[a]	Gene name	Protein name[b]	Align-ment length (bp)	JCVI role category[c]	Positive selection[d]	BEB (*P* >95%)[e]
STM1450	-	Pyridoxal kinase	666	Biosynthesis of cofactors, prosthetic groups, and carriers	Ty# (0.0147)	-
STM3057	*ubiH*	2-octaprenyl-6-methoxyphenol hydroxylase, UbiH	1176	Biosynthesis of cofactors, prosthetic groups, and carriers	Tym# (0.1149)	-
STM1441	-	mMembrane protein, putative	1995	Cell Envelope	Ch# (0.0043)	-
STM2267	*ompC*	Outer membrane protein C precursor	1134	Cell Envelope	Ch# (0.0986)	274
STM0743	-	Putative lipoprotein	273	Cell Envelope	Ch# (0.1830)	-
STM2801	*ygaC*	Conserved hypothetical protein	300	Cell envelope	Pty# (0.020)	-
STM0301	*safC*	Outer membrane usher, *Salmonella* atypical fimbria	2508	Cell envelope	TO (0.0104)	85, 111, 405, 692
STM4106	*katG*	Catalase hydroperoxidase HPI(I)	2178	Cellular processes	TO (0.0035)	-
STM1425	*ydhE*	Hypothetical integral membrane protein	1371	Cellular processes	Tym# (0.0145)	-
STM0603	*araT*	Aminotransferase, class I	1158	Central intermediary metabolism	Ty# (0.1976)	-

Table 3. 4. (Continued)

STM0395	-	Exonuclease SbcC, putative	3096	DNA metabolism	Ty# (0.1041)	-
STM4023	-	Putative 3-hydroxyisobutyrate dehydrogenase	840	Energy metabolism	Ch# (0.0138)	-
STM3680	*aldB*	Aldehyde dehydrogenase B	1536	Energy metabolism	Pty# (0.020)	-
STM0698	*pgm*	Phosphoglucomutase, alpha-D-glucose phosphate-specific	1638	Energy metabolism	Ty# (0.0157)	-
STM3515	*malT*	MalT regulatory protein	2703	Energy metabolism	Ty# (0.0198)	801
STM4187	*iclR*	Acetate operon repressor	819	Energy metabolism	Ty# (0.0693)	-
STM0401	*malZ*	Glycosyl hydrolase, family 13	1815	Energy metabolism	Tym# (0.1005)	-
STM3329	-	Conserved hypothetical protein TIGR01212	927	Hypothetical proteins	Ch# (0.1471)	-
STM1854	-	Hypothetical protein	162	Hypothetical proteins	Pty# (0.1973)	32, 40, 44, 45
STM0861	-	Conserved hypothetical protein	471	Hypothetical proteins	Tym# (0.1149)	-
STM1515	-	Conserved hypothetical protein TIGR00156 domain protein	384	Hypothetical proteins	Ty# (0.0167)	-
STM4015	-	Hypothetical protein	846	Hypothetical proteins	Ty# (0.0693)	-
STM4258	-	Conserved hypothetical protein	1386	Hypothetical proteins	Ty# (0.0884)	-
STM1532	-	Hypothetical protein	678	Hypothetical proteins	Ty# (0.1614)	-
STM1280	-	Conserved hypothetical protein	396	Hypothetical proteins	Tym# (0.0145)	-

Table 3. 4. (Continued)

STM3463	-	Conserved hypothetical protein	201	Hypothetical proteins	Tym# (0.1005)	-
STM4522	-	Hypothetical protein	699	Hypothetical proteins	Tym# (0.1649)	-
STM3655	*glyS*	Glycyl-tRNA synthetase, beta subunit	2067	Protein synthesis	Ty# (0.0393)	313
STM0534	*purE*	Phosphoribosylaminoimidazole carboxylase, catalytic subunit	507	Purines, pyrimidines, nucleosides, and nucleotides	Ty# (0.0194)	-
STM2806	*nrdI*	NrdI protein	408	Purines, pyrimidines, nucleosides, and nucleotides	Ty# (0.0693)	-
STM2107	*wcaH*	GDP-mannose mannosyl hydrolase	435	Purines, pyrimidines, nucleosides, and nucleotides	Tym# (0.1081)	-
STM3262	*glpR*	Transcriptional regulator, DeoR family	771	Regulatory functions	Tym# (0.1843)	-
STM1679	*oppA*	Oligopeptide ABC transporter, periplasmic oligopeptide-binding protein	1605	Transport and binding proteins	Ch# (0.0121)	-
STM3685	-	PTS system, mannitol-specific IIC component subfamily, putative	1914	Transport and binding proteins	TO (0.0035)	-
STM3258	-	PTS system IIA component, putative	462	Transport and binding proteins	Ty# (0.0157)	124, 139, 143, 144, 147
STM3626	*oppF*	Oligopeptide ABC transporter, ATP-binding protein	1011	Transport and binding proteins	Ty# (0.0157)	-

Table 3. 4. (Continued)

STM3592	-	Proton/peptide symporter family protein	1470	Unclassified	TO (0.0104)	-
STM1088	*pipB*	Pathogenicity island encoded protein: SPI5, PipB	873	Unclassified	TO (0.0688)	173
STM0248	-	Histidinol phosphatase-related protein	573	Unknown function	Ch# (<0.0001)	175, 184, 185, 191
STM3565	-	Acetyltransferase, GNAT family	381	Unknown function	Pty# (0.0391)	-
STM3955	*rarD*	RarD protein	879	Unknown function	Ty# (0.0194)	-
STM2678	*b2611*	Putative membrane protein, CorE	750	Viral functions	Ch# (0.1411)	-
STM4242	-	99% identical to TraF of plasmid R64	1284	Viral functions	Pty# (0.0329)	-

[a]Role categories were assigned based on annotations for *S*. Typhi CT18; JCVI locus names for Typhi CT18 for these genes are listed in Supp. Table S3

[b]Protein designations were taken from the Typhi CT18 annotation; where limited annotation information was available, additional information was extracted from JCVI primary annotations and Typhimurium LT2 and Paratyphi annotations

[c]some genes are assigned multiple JCVI role categories; all role categories for a given gene are listed here

[d]tests that were significant for positive selection (FDR<20 %) are shown; TO = overall test; Ch#= Choleraesuis branch specific test; Pty#= Paratyphi A branch specific test ; Ty# = Typhi branch specific test; Tym#= Typhimurium branch specific test; numbers in brackets indicate *q*-values

[e]aa sites identified by Bayes Empirical Bayes (BEB) as having probability > 95% of being under positive selection are shown; aa sites are based on site location in the alignment (alignments for genes under positive selection are provided as Supp. Material S4)

Envelope" and genes with evidence of positive selection in branch analyses for *S*. Typhi and *S*. Choleraesuis (uncorrected *P*= 0.034 and *P*=0.023, respectively, Fisher's exact test). However, after corrections for multiple comparisons, these associations are not significant (Bonferroni corrected *P*= 0.301 and *P*=0.114, respectively; Fisher's exact test) (Figure.3. 3). Because of the low number of genes under positive selection, it was not possible to assess the association between positive selection and most role categories. We thus assessed whether the distribution of the *p*-values for each test deviates from the random distribution for any of the role categories using the non-parametric U-test. The JCVI role category "Hypothetical proteins" showed significant trends of having genes with low *p*-values in the Choleraesuis, Typhimurium and Paratyphi A branch specific tests for positive selection (Bonferroni corrected *P*=0.042, *P*= 0.034 and *P*<0.001, respectively; one sided U-test). In addition, this role category had also significant tendencies to have lower *p*-values in the overall (TO) and Typhi branch tests (uncorrected *P*= 0.010 and *P*=0.008, respectively, one sided U-test). However, after Bonferroni correction for multiple comparisons, these associations are not significant (*P*= 0.124 and *P*=0.068, respectively; one sided U-test). In addition, the JCVI role categories "Unclassified" and "Protein synthesis" showed the significant trends of having genes with low *p*-values in Choleraesuis and Typhimurium branch tests for positive selection, respectively (Bonferroni corrected *P*= 0.002 and *P*=0.013, respectively; one sided U-test).

Among the 41 genes with evidence for positive selection, two were located in *Salmonella* pathogenicity islands 1 through 5 (i.e., *pipB,* STM1088 (*siiB)*; see Table 3.5). Overall 78 of the orthologs analyzed were located in five *Salmonella* pathogenicity islands (40), thus, genes in the pathogenicity islands were not significantly overrepresented (*P*= 0.283; Fisher's exact test) among the genes under positive selection. In addition, three SPI-1 genes (i.e., *spaM, iagB,* and *sipD*), and one

Table 3. 5. *Salmonella* pathogenicity island (SPI) genes with evidence of positive selection and recombination

SPIs[a]	Locations in Typhimurium LT2[b]	No. of orthologous genes found in SPI	No. of genes under positive selection	No. genes with evidence of recombination
1	STM2865-2914	33	0	1 (*stpA*)
2	STM1379-1422	30	0	3 (*sseC,sseD,* STM1379)
3	STM3752-3764	5	0	0
	STM3766-3775			
4	STM4257-4262	6	1 (*siiB*)	0
5	STM1087-1094	4	1 (*pipB*)	0

[a]This table only lists genes in the five common *Salmonella* Pathogenicity islands (i.e., SPIs 1 to 5)
[b]genes in a given island are reported as described by McClelland et al. 2001

Table 3. 6. Evidence of recombination among genes with evidence for positive selection

Test for positive selection[a]	No. of genes under positive selection with no evidence of recombination	No. of genes under positive selection that show evidence of recombination with[b]				No. of genes under positive selection with evidence of recombination[c]
		GENE-CONV	Max-χ^2	PHI	NSS	
TO	5	12	7	3	4	21
Ch#	8	11	10	1	3	23
Ty#	16	4	3	0	3	21
Tym#	7	4	4	0	1	13
Pty#	5	8	7	1	2	14

[a]TO = overall test; Ch#= Choleraesuis branch specific test; Ty# = Typhi branch specific test; Tym#= Typhimurium branch specific test; Pty#= Paratyphi A branch specific test

[b]Based on our preliminary analysis, among 3316 orthologous genes, 81 genes showed evidence of positive selection in at least one test. Among 81 genes, 32 genes also showed evidence of recombination with at least one of the four recombination tests used in our study. Statistical analysis showed that genes evidence of recombination were more likely to be under positive selection (*P*<0.0001; chi-square test). Therefore, we excluded 270 genes with evidence of recombination from our positive selection analysis.

[c]This column lists the number of genes with evidence for positive selection in a given test (e.g., TO); since some genes showed evidence of recombination in > 1 recombination test, the total number of genes in this column is typically lower than the sum of the numbers in a given row. While a total of 81 genes showed evidence of positive selection, the sum of the numbers in this column is > 81 as 11 genes showed evidence of positive selection in two tests.

SPI-2 gene (*ssaI*) showed *P*-values <0.05 in the overall positive selection tests (*P*= 0.049, 0.017, 0.003 and 0.047, respectively), but failed to meet the FDR cutoff (<0.2) (*Q*-values=1, 1, 0.925, and 1, respectively). Similarly, one SPI-2 gene (*sseF*) showed very low *P*-value (*P*= 0.001) in the Choleraesuis branch test, but failed to meet the FDR cutoff (<0.2) (*Q*-value= 0.332).

ompC showed evidence for positive selection in our study (Table 3.4) as well as in a previous study of *Shigella* and *E. coli* (52). Our analyses showed that aa residues 228 and 274 show evidence for positive selection, while aa 163, 202, and 203 showed evidence for positive selection in *E. coli* and *Shigella* (52). *Salmonella* OmpC aa site 228, which was found to be under positive selection here, is located in a region that is absent from the *E. coli* and present in *Shigella* OmpC, while *Salmonella* OmpC aa site 274 is located in a region that is absent from the *E. coli* and *Shigella* OmpC.

Verification of positive selection and recombination patterns, identified by genome wide analyses, for four genes among 42 *Salmonella* isolates. In order to confirm positive selection and recombination patterns identified by the full genome analyses, we sequenced and analyzed four genes that showed evidence for positive selection, including two genes that showed evidence for positive selection and recombination (i.e., *folK-2*, *sseC*) and two genes that only showed evidence for positive selection (i.e., STM3258, *purE*). *folK-2*, which encodes an enzyme involved in the synthesis of folic acid, could not be PCR amplified in 6 *Salmonella* isolates, representing serotypes Montevideo, Oranienburg, Javiana, Urbana, Muenster. Analyses of 41 *folK-2* sequences (5 sequences from the genomes and 36 newly determined sequences) confirmed that this gene is under positive selection in the *Salmonella* Typhi branch and shows evidence for recombination, but did not find evidence for positive selection in the overall analysis. The STM3258 gene, which encodes a putative PTS component, could not be PCR amplified in one *Salmonella*

Typhimurium and three serotype 4,5,12:i:- isolates. Results from the analyses of the resulting 43 STM3258 gene sequences was consistent with the genome analyses data and confirmed that this gene shows no evidence for recombination, but is under positive selection in the *Salmonella* Typhi branch. *sseC*, which is located in the *Salmonella* pathogenicity island 2, could not be PCR amplified in 6 *Salmonella* isolates, representing serotypes Agona (n=2), Havana, Kentucky, and Mbandaka (n=2). Analyses of the *sseC* sequences confirmed positive selection in the Choleraesuis branch, but did not find evidence for recombination (even though the genome wide analyses found evidence for recombination using GENECONV; Table 3.2). *purE*, which encodes an enzyme involved in the synthesis of purine ribonucleotide, was successfully amplified and sequenced in all 42 isolates; analyses of the resulting sequences confirmed that this gene is under positive selection in the *Salmonella* Typhi branch. Overall, analyses of additional isolates thus confirmed branch specific positive selection for all four genes.

DISCUSSION

In this study, we used 5 *Salmonella* genomes representing host restricted (i.e., Typhi and Paratyphi A), host adapted (i.e., Choleraesuis), and unrestricted (i.e., Typhimurium) serotypes to study the evolution of core genes in different *Salmonella* serotypes. A total of 3,316 orthologs among these 5 *Salmonella* genomes were used to (i) identify genes under positive selection and (ii) identify genes with evidence of recombination. Positive selection and recombination patterns for four genes of interest were confirmed in a larger set of isolates representing 23 different serotypes. Overall, our data show that (i) less than 10% of *Salmonella* genes in the core genome show evidence for homologous recombination, (ii) a number of core *Salmonella* genes are under positive selection, including genes that appear to contribute to virulence, and (ii)

cell surface protein, *ompC,* that is targeted by positive selection in both *Salmonella* and *E. coli* (52), may contribute to multi drug resistance in *Salmonella*.

Less than 10% of *Salmonella* genes show evidence for intragenic recombination. Since the first bacterial genome was sequenced in 1995, comparative tools have showed that horizontal gene transfer is the major process for the evolution of prokaryotes (31, 35, 50). Horizontal gene transfer has also been proposed to have played an important role in the evolution of the *Salmonella* genome. *Salmonella* Typhimurium LT2 seems to have acquired a number of novel genomic regions after the divergence from *E. coli* around 100 millions years ago (32) and it has been estimated that 25 % of the *Salmonella* Typhimurium genome may have been introduced by horizontal gene transfer (Porwollik and McClelland 2003). Groups of genes introduced by horizontal gene transfer include prophages and *Salmonella* pathogenicity islands (SPI) (54). While the role of horizontal gene transfer in introducing novel genes into the *Salmonella* genome has thus been well established, our analyses show that horizontal transfer (and recombination) of homologous genes also plays an important role in the diversification of *Salmonella*; we found that 270 of the 3316 genes characterized (8.1 %) showed evidence for intragenic recombination. This relatively low level of recombination is consistent with the observation that *Salmonella* is highly clonal (5, 17). By comparison, analysis of four *E. coli* and two *Shigella* found 236 genes with evidence for intragenic recombination, representing approximately 6.3 % of genes analyzed (52). Chen et al. (2006) reported that 12.8% of core genome genes, found in seven *E. coli* genomes, showed evidence for recombination. A study of 410 genes present in six *E. coli* and six *Salmonella enterica* genomes reported that 23% of these genes showed evidence of recombination in *Salmonella;* this estimate may be higher than the one reported here as the 410 genes evaluated do not represent a random sample of the *Salmonella* core genome (12).

Interestingly, even novel genes that were initially introduced into the *Salmonella* genome through horizontal gene transfer and non-homologous recombination, showed evidence for further subsequent diversification through homologous recombination (e.g., one and two genes in SPI-1 and 2, respectively, showed evidence for intergenic recombination). A recent analysis by Didelot et al. (2007) also suggested that convergence of *Salmonella* Typhi and Paratyphi A, two human host-restricted serotypes, through >100 recombination events involving both transfer of novel genes as well as transfer of homologous genes, further supporting the importance of horizontal transfer of homologous gene sequences in the evolution of *Salmonella* (18).

A number of core *Salmonella* genes are under positive selection, including genes that appear to contribute to virulence and systemic infection. A total of 1.2 % of genes found in all five *Salmonella* genomes (i.e., 41 genes) showed evidence for positive selection and no evidence for recombination. While 5 genes showed evidence for positive selection in the overall analyses, 36 genes showed evidence for positive selection only in specific branches, indicating considerable branch specific positive selection in the species *Salmonella enterica.* Previously, Petersen et al. (2007) reported that, among 3,505 *E. coli* and *Shigella* genes that showed no evidence for recombination, a total of 23 genes (0.66%) showed evidence for positive selection. Among Gram-positive pathogens, Orsi et al. (2008) reported that 36 *L. monocytogenes* and *L. innocua* genes (1.6 %) showed evidence of positive selection (among a total of 2267 genes analyzed), while Lefebure and Stanhope (2007) reported that 11 to 34% of the genes in the *Streptococcus* core genome showed evidence for positive selection, although this study did not control for multiple comparisons and thus may have somewhat overestimated the number of genes under positive selection. Overall, these data suggest that typically few genes in a given bacterial species show evidence for positive selection.

Interestingly, three *Salmonella* genes with evidence for positive selection were located in *Salmonella* pathogenicity islands (SPIs). SPIs are chromosomal regions that contain genes contributing to a particular virulence phenotype (22, 38, 72). These regions appear to have been acquired by *Salmonella* after the divergence of *E. coli* and *Salmonella* (54). So far, five common SPIs (i.e., SPI-1 through SPI-5), found among the majority of *Salmonella enterica* strains, as well as a number of additional less common SPIs have been reported. A gene with evidence for positive selection (i.e., STM4258; *siiB*) is located in SPI-4 and encodes a probable membrane protein (putative methyl-accepting chemotaxis protein). Morgan et al. (2004) reported that the SPI-4 genes, *siiD, siiE,* and *siiF* play a role in *Salmonella* Typhimurium intestinal colonization of calves (44). It also has been proposed that SPI-1 and SPI-4 play complementary roles in the infection of host epithelial cells (37). Kiss et al. (28) specifically showed that a *Salmonella* Typhimurium *siiB* mutant shows reduced secretion of SiiE, as compared to the wildtype, suggesting a possible involvement of *siiB* in calf virulence (as an *siiE* mutant showed reduced colonization in a calf model [(44)]). *pipB* (STM1088), located in SPI-5, also showed evidence for positive selection. SPI-5 encodes T3SS-1 and T3SS-2 effector proteins (74). PipB localizes to the *Salmonella* Containing Vacuole (SCV) in mammalian host cells (30). In addition, Wood et al. (1998) reported that a *pipB* null mutant showed reduced intestinal secretory and inflammatory responses in ligated bovine ileal loops, suggesting that this, as well as other genes in SPI-5, may contribute to bovine enteric infections (74). PipB also appears to be required for colonization of the cecum, by *Salmonella* Typhimurium, in chickens (43). *safC* (STM0301)*,* a gene located in SPI-6 (51), a region called *Salmonella enterica* centrisome 7 genomic island (SCI) in *Salmonella* Typhimurium (20), was also found to be under positive selection. *safC* encodes an outer membrane usher protein for *Salmonella* atypical fimbriae (20). While a *Salmonella* Typhimurium strain with a deletion of SCI (SPI-6) showed reduced ability

to invade Hep2 cells (20), we are not aware of any studies characterizing virulence of a *safC* null mutant. Overall, our findings are consistent with a previous study (19) that reported that a number of genes located in *Salmonella* pathogenicity islands show evidence for differential evolution in different *Salmonella* serotypes.

Interestingly, two genes (i.e., *sseC* and *sseF*), which are located in SPI-2, (16) showed evidence for positive selection in the Choleraesuis branch in our initial analysis. In our second analyses (i.e. analysis excluding genes with recombination), *sseF* was removed, since *sseF* showed evidence of recombination and *sseC* could not meet the cut off *Q*-value for 20% FDR. *sseC* showed also evidence for positive selection and recombination in the follow-up analysis using a larger number of *Salmonella* serotypes. SPI-2 encodes a type III secretion system (T3SS), which interferes with phagosome maturation and facilitates formation of a *Salmonella*-containing vacuole (SCV) (16). Genes in SPI-2 encode a number of effector proteins (e.g., *sseC* and *sseF*), which are required for bacterial replication in macrophages (38). *sseC*, encoding the translocon component SseC, has also been shown to be required for systemic *Salmonella* Typhimurium infection in mice. Localization of SseC to bacterial membrane is essential for function of T3SS (29). *sseF* encodes an effector protein that is required for dynein recruitment of intercellular *Salmonella* and a *sseF* null mutant showed reduced intracellular proliferation and reduced formation of intracellular *Salmonella* microcolonies (1). *sseF* and *sseC* also have previously been shown to contain distinct clusters of polymorphic sites that might be unique to the human adapted serotypes Typhi and Paratyphi (68). Moreover, the two genes in SPI-2 identified here as showing evidence for lineage specific evolution (i.e., *sseC, sseF*), were also identified in a previous study (19) as showing evidence for differential evolution.

Overall, three genes with no recombination evidence in the JCVI role category "Purine, pyrimidine, nucleotide and nucleotide biosynthesis" (i.e., *wcaH, purE* and

nrdI) showed evidence for positive selection. *wcaH,* which encodes a GDP-mannose mannosyl hydrolase, is under positive selection in the Typhimurium branch, while *purE* and *nrdI* were found to be under positive selection in the Typhi branch. *purE* encodes a phosphoribosylaminoimidazole carboxylase, while *nrdI*, which is located in an operon with genes that encode a Class 1b ribonucleotide reductase, encodes a small flavoprotein with unknown function in *Streptococcus pyogenes* (56) .

Genes in the JCVI role category "Purine, pyrimidine, nucleotide and nucleotide biosynthesis" also showed a significant trend of having genes with evidence of positive selection in the Typhi branch (uncorrected *P*=0.034, Fisher's exact test). However, after corrections for multiple comparisons, the association is not significant (Bonferroni corrected *P*= 0.309, Fisher's exact test). Positive selection for *purE* in the *Salmonella* Typhi branch was also confirmed in our analyses of 22 human and 20 animal *Salmonella* isolates, which included two additional Typhi strains. This is a striking finding since Samant et al. (2008) recently reported that the novo nucleotide biosynthesis is essential for bacterial growth in blood (59). As *Salmonella* Typhi predominantly causes systemic septicemic infections in humans, these findings suggests that adaptive changes in genes encoding purine, pyrimidine, nucleotide and nucleotide biosynthesis functions may have been critical in the evolution of this host restricted human pathogen. Our findings thus further support that development of novel drugs targeting appropriate purine, pyrimidine, nucleotide and nucleotide biosynthesis pathways may represent an opportunity for therapeutic approaches for bacterial pathogens causing septicemic infections (Samant et al. 2008).

Additional genes with evidence for positive selection and possible roles in host infection include *katG* (STM4106)*,* which encodes a catalase. While antioxidant defenses mechanism appear to contribute to virulence in a number of pathogens, *Salmonella katG* null mutations have shown no affect on *Salmonella*'s ability to survive inside phagocytic cells and in a murine model of infections (9). The

importance of adaptive changes in *Salmonella katG* thus remains to be determined. It seems possible that adaptive changes in genes involved in anaerobic growth may contribute to an improved ability of different strains of this gastrointestinal pathogen to survive under anaerobic conditions encountered in the intestinal tract.

We also identified a number of genes with evidence for positive selection that have no apparent link to infection and virulence, including *malZ* (STM0401), *malT* (STM3515) and *mtlA* (STM3685), which encode, respectively, a maltodextrin glucosidase, a transcriptional activator of *mal* genes, and a mannitol specific PTS system component. While it has been proposed that horizontal transfer of genes encoding proteins involved in acquisition and synthesis of nutrients and genes encoding components of metabolic networks is critical as bacteria adapt to specific environments and ecological niches (50), our findings suggest that positive selection of genes encoding metabolic capabilities also contribute to adaptation to new environments.

Cell surface proteins are targeted by positive selection in both *Salmonella* and *E. coli*. While, in our preliminary analysis, we identified three genes encoding outer membrane proteins (*ompC, ompS1* and *ompS2*) that showed evidence for positive selection, *ompC* was the only gene that remained significant in our positive selection analysis for genes without evidence of recombination, since both *ompS1* and *ompS2* showed evidence of recombination. *ompC,* a highly expressed *omp* gene, encodes a protein that not only appears to play a role in *Salmonella* virulence (46), but also is a receptor for Gifsy-1 and Gifsy-2 phages (24). While *ompS1* and *ompS2* encode two porins that appear to be expressed at low levels, mutants in these two genes showed attenuated virulence in a mouse model (57). An analysis of 6 *E. coli* and *Shigella* genomes also found that three *omp* genes (i.e., *ompF,*

ompC and *ompA*) showed evidence of positive selection (52), while Chen et al. (2006) reported that *ompC* and *ompF* were under positive selection in uropathogenic *E. coli* strains. Furthermore, in *Rickettsia* spp., genes encoding the outer membrane proteins OmpA and OmpB showed evidence for positive selection (26). Overall, these data strongly suggest that adaptive changes in genes encoding outer membrane proteins critically contribute to the evolution of a variety of bacteria, including pathogenic enterobacteriaciae. In particular, *ompC*, which encodes one of the most abundant *E. coli* proteins (52), appears to be under positive selection in a number of pathogenic enterobacteriaciae. As proposed by Petersen et al. (2007), positive selection in *omp* genes may be an important mechanism that facilitates adaptation of bacterial pathogens allowing them to escape recognition by the host immune system and phages. In addition, mutations in porin genes (e.g., those belonging to OmpC and OmpF groups), as well as changes in Omp expression levels, have been linked to increased resistance to beta-lactam antibiotics (2, 42, 49). For example, under strong antibiotic pressure, bacteria can reduce the influx of antibiotic through downregulation of porin expression or expression of modified porins. Positive selection in porin genes, particularly *ompC* thus may also be associated with selection to increase antibiotic resistance. These findings provide potentially interesting avenues for future mutagenesis studies to elucidate the role of *ompC* polymorphisms in various phenotypes, including -lactam resistance.

CONCLUSIONS

Our analyses clearly show that both homologous recombination and positive selection (particularly lineage specific positive selection) contribute critically to the evolution of the *Salmonella* core genome. Genes with evidence of positive selection

identified here may provide promising targets for future mutational studies aimed at further identifying mechanisms that contribute to *Salmonella* diversification, including its adaptation to specific host species. The relevance of the lineage specific positive selection patterns identified is supported by the convergence of the positive selection patterns identified in the *Salmonella* Typhi lineage (i.e., for genes encoding proteins involved in purine, pyrimidine, nucleotide and nucleotide biosynthesis) and experimental evidence that genes involved in de novo nucleotide biosynthesis are essential for bacterial growth in blood (59).

In conjunction with previous genome wide studies on positive selection in uropathogenic *E. coli* (13), *Shigella* and *E. coli* (52), *Listeria* spp. (48), and *Streptococcus* spp. (34), our data clearly indicate the positive selection and homologous recombination among core genome genes play an important role in the evolution of bacterial pathogens, in addition to the well established importance of gene acquisition and deletion. Positive selection and homologous recombination also appear to contribute to further evolution of novel genes initially acquired by lateral gene transfer, such as selected genes in the *Salmonella* pathogenicity islands. As additional pathogen genomes, including additional *Salmonella* genomes, become available, positive selection and recombination analyses on larger numbers of genomes will further improve our understanding of bacterial pathogens.

Acknowledgements

This work was partially supported by USDA Special Research Grants 2005-34459-15625 and 34459-16952-06 (to MW) and the National Institute of Allergy and Infectious Disease (NIAID) National Institute of Health (NIH), US Department of Health and Human Services under contract N01-AI-30054 (to Lorin Warnick). The computer cluster used in the data analysis is partially funded by Microsoft. The authors thank Paige Smith for help with DNA sequencing.

Supplementary materials

All supplementary materials are available at

http://www.foodscience.cornell.edu/cals/foodsci/research/labs/wiedmann/links/soyer2009.cfm

Supplementary Table 1. *Salmonella* isolates (n=42) used to verify genome wide positive selection and recombination patterns in four selected genes

Supplementary Table 2. PCR conditions and primers for the four genes that were used to verify genome wide positive selection and recombination patterns in an additional 42 *Salmonella* isolates.

Supplementary Table 3. Preliminary results from positive selection and recombination analyses

Supplementary Table 4. Results from positive selection for genes without recombination

Supplementary Material 5. Alignments for gene under positive selection

REFERENCES

1. **Abrahams, G. L., P. Muller, and M. Hensel.** 2006. Functional dissection of SseF, a type III effector protein involved in positioning the salmonella-containing vacuole. Traffic **7:**950-65.
2. **Alcaine, S. D., L. D. Warnick, and M. Wiedmann.** 2007. Antimicrobial resistance in nontyphoidal *Salmonella*. Journal of Food Protection **70:**780-790.
3. **Andrews, T. D., and T. Gojobori.** 2004. Strong positive selection and recombination drive the antigenic variation of the PilE protein of the human pathogen Neisseria meningitidis. Genetics **166:**25-32.
4. **Anisimova, M., R. Nielsen, and Z. Yang.** 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics **164:**1229-1236.
5. **Baker, S., and G. Dougan.** 2007. The Genome of *Salmonella enterica* Serovar Typhi. Clin. Infect. Dis. **45:**S29-S33.
6. **Benjamini, Y., and Y. Hochberg.** 1995. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society B **57:**289-300.
7. **Brenner, F. W., R. G. Villar, F. J. Angulo, R. Tauxe, and B. Swaminathan.** 2000. *Salmonella* nomenclature - Guest commentary. J. Clin. Microbiol. **38:**2465-2467.
8. **Bruen, T. C., H. Philippe, and D. Bryant.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172:**2665-2681.
9. **Buchmeier, N. A., S. J. Libby, Y. Xu, P. C. Loewen, J. Switala, D. G. Guiney, and F. C. Fang.** 1995. DNA repair is more important than catalase for Salmonella virulence in mice. J Clin Invest **95:**1047-53.

10. **CDC** 2003, posting date. Center for Disease Control and Prevention. Salmonella surveillance: Annual Summary, 2003. CDC. [Online.]

11. **Chapman, M. A., J. H. Leebens-Mack, and J. M. Burke.** 2008. Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. Mol Biol Evol **25:**1260-73.

12. **Charlesworth, J., and A. Eyre-Walker.** 2006. The rate of adaptive evolution in enteric bacteria. Mol. Biol. Evol. **23:**1348-56.

13. **Chen, S. L., C. S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D. Blasiar, T. Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J. P. Latreille, J. Spieth, T. M. Hooton, E. R. Mardis, S. J. Hultgren, and J. I. Gordon.** 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. Proc. Natl. Acad. Sci. U.S.A. **103:**5977-5982.

14. **Chen, Z. H., and T. D. Schneider.** 2006. Comparative analysis of tandem T7-like promoter containing regions in enterobacterial genomes reveals a novel group of genetic islands. Nucleic Acids Research **34:**1133-1147.

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

16. **Coburn, B., I. Sekirov, and B. B. Finlay.** 2007. Type III secretion systems and disease. Clin Microbiol Rev **20:**535-49.

17. **Deng, W., S. R. Liou, G. Plunkett, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, D. C. Schwartz, and F. R. Blattner.** 2003. Comparative genomics of Salmonella enterica serovar typhi strains Ty2 and CT18. Journal of Bacteriology **185:**2330-2337.

18. **Didelot, X., M. Achtman, J. Parkhill, N. R. Thomson, and D. Falush.** 2007. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: convergence or divergence by homologous recombination? Genome Res **17:**61-8.

19. **Eswarappa, S. M., J. Janice, A. G. Nagarajan, S. V. Balasundaram, G. Karnam, N. M. Dixit, and D. Chakravortty.** 2008. Differentially evolved genes of Salmonella pathogenicity islands: insights into the mechanism of host specificity in Salmonella. PLoS ONE **3:**e3829.

20. **Folkesson, A., S. Lofdahl, and S. Normark.** 2002. The Salmonella enterica subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells. Res Microbiol **153:**537-45.

21. **Gal-Mor, O., Y. Valdez, and B. B. Finlay.** 2006. The temperature-sensing protein TlpA is repressed by PhoP and dispensable for virulence of *Salmonella enterica* serovar Typhimurium in mice. Microb. Infect. **8:**2154-2162.

22. **Hacker, J., and J. B. Kaper.** 2000. Pathogenicity islands and the evolution of microbes. Annual Review of Microbiology **54:**641-679.

23. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
program for Windows 95/98/NT. . Nucl. Acids. Symp. Ser. **41:**95-98.

24. **Ho, T. D., and J. M. Slauch.** 2001. OmpC is the receptor for Gifsy-1 and Gifsy-2 bacteriophages of Salmonella. J Bacteriol **183:**1495-8.

25. **Jakobsen, I. B., and S. Easteal.** 1996. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Computer Applications in the Biosciences **12:**291-295.

26. **Jiggins, F. M.** 2006. Adaptive evolution and recombination of Rickettsia antigens. J Mol Evol **62:**99-110.

27. **Kidgell, C., U. Reichard, J. Wain, B. Linz, M. Torpdahl, G. Dougan, and M. Achtman.** 2002. Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old. Infect Genet Evol **2:**39-45.

28. **Kiss, T., E. Morgan, and G. Nagy.** 2007. Contribution of SPI-4 genes to the virulence of Salmonella enterica. FEMS Microbiol Lett **275:**153-9.

29. **Klein, J. R., and B. D. Jones.** 2001. Salmonella pathogenicity island 2-encoded proteins SseC and SseD are essential for virulence and are substrates of the type III secretion system. Infect Immun **69:**737-43.

30. **Knodler, L. A., J. Celli, W. D. Hardt, B. A. Vallance, C. Yip, and B. B. Finlay.** 2002. Salmonella effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. Molecular Microbiology **43:**1089-1103.

31. **Koonin, E. V., and Y. I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res.

32. **Koski, L. B., R. A. Morton, and G. B. Golding.** 2001. Codon bias and base composition are poor indicators of horizontally transferred genes. Mol Biol Evol **18:**404-12.

33. **Langvad, B., M. N. Skov, E. Rattenborg, J. E. Olsen, and D. L. Baggesen.** 2006. Transmission routes of *Salmonella* Typhimurium DT 104 between 14 cattle and pig herds in Denmark demonstrated by molecular fingerprinting. J. Appl. Microbiol. **101:**883-890.

34. **Lefebure, T., and M. J. Stanhope.** 2007. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition. Genome Biol. **8:**R71-88.

35. **Lercher, M. J., and C. Pal.** 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. Mol. Biol. Evol. **25:**559-567.

36. **Li, L., C. J. Stoeckert, Jr., and D. S. Roos.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178-2189.

37. **Main-Hester, K. L., K. M. Colpitts, G. A. Thomas, F. C. Fang, and S. J. Libby.** 2008. Coordinate regulation of Salmonella pathogenicity island 1 (SPI1) and SPI4 in Salmonella enterica serovar Typhimurium. Infect Immun **76:**1024-35.

38. **Marcus, S. L., J. H. Brumell, C. G. Pfeifer, and B. B. Finlay.** 2000. Salmonella pathogenicity islands: big virulence in small packages. Microbes and Infection **2:**145-156.

39. **McClelland, M., K. E. Sanderson, S. W. Clifton, P. Latreille, S. Porwollik, A. Sabo, R. Meyer, T. Bieri, P. Ozersky, M. McLellan, C. R. Harkins, C. Y. Wang, C. Nguyen, A. Berghoff, G. Elliott, S. Kohlberg, C. Strong, F. Y. Du, J. Carter, C. Kremizki, D. Layman, S. Leonard, H. Sun, L. Fulton, W. Nash, T. Miner, P. Minx, K. Delehaunty, C. Fronick, V. Magrini, M. Nhan, W. Warren, L. Florea, J. Spieth, and R. K. Wilson.** 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica t*hat cause typhoid. Nat. Genet. **36:**1268-1274.

40. **McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Y. Du, S. F. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston,

and R. K. Wilson. 2001. Complete genome sequence of Salmonella enterica serovar typhimurium LT2. Nature **413:**852-856.

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

42. **Medeiros, A. A., T. F. O'Brien, E. Y. Rosenberg, and H. Nikaido.** 1987. Loss of OmpC porin in a strain of Salmonella typhimurium causes increased resistance to cephalosporins during therapy. J Infect Dis **156:**751-7.

43. **Morgan, E.** 2007. *Salmonella* Pathogenecity Islands, vol. Cromwell Press Trowbridge Wiltshire, Norfolk.

44. **Morgan, E., J. D. Campbell, S. C. Rowe, J. Bispham, M. P. Stevens, A. J. Bowen, P. A. Barrow, D. J. Maskell, and T. S. Wallis.** 2004. Identification of host-specific colonization factors of Salmonella enterica serovar Typhimurium. Mol.Microbiol. **54:**994-1010.

45. **Nagano, N., S. Oana, Y. Nagano, and Y. Arakawa.** 2006. A severe *Salmonella enterica* serotype Paratyphi B infection in a child related to a pet turtle, *Trachemys scripta elegans*. Jpn. J. Infect. Dis. **59:**132-134.

46. **Negm, R. S., and T. G. Pistole.** 1999. The porin OmpC of Salmonella typhimurium mediates adherence to macrophages. Can J Microbiol **45:**658-69.

47. **Nielsen, R., C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fledel-Alon, D. M. Tanenbaum, D. Civello, T. J. White, J. Sninsky, M. D. Adams, and M. Cargill.** 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. **3:**e170.

48. **Orsi, R. H., Q. Sun, and M. Wiedmann.** 2008. Genome-wide analyses reveal lineage specific contributions of positive selection and recombination to the evolution of *Listeria monocytogenes*. BMC Evol. Biol. **8:**233-254.

49. **Pages, J. M., C. E. James, and M. Winterhalter.** 2008. The porin and the permeating antibiotic: a selective diffusion barrier in Gram-negative bacteria. Nat Rev Microbiol **6:**893-903.

50. **Pal, C., B. Papp, and M. J. Lercher.** 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. **37:**1372-1375.

51. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. G. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell.** 2001. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. Nature **413:**848-852.

52. **Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen.** 2007. Genes under positive selection in *Escherichia coli*. Genome Res. **17:**1336-1343.

53. **Porwollik, S., E. F. Boyd, C. Choy, P. Cheng, L. Florea, E. Proctor, and M. McClelland.** 2004. Characterization of *Salmonella enterica* subspecies I genovars by use of microarrays. J. Bacteriol. **186:**5883-5898.

54. **Porwollik, S., and M. McClelland.** 2003. Lateral gene transfer in *Salmonella*. Microbes. Infect. **5:**977-989.

55. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **16:**276-7.

56. **Roca, I., E. Torrents, M. Sahlin, I. Gibert, and B. M. Sjoberg.** 2008. NrdI essentiality for class Ib ribonucleotide reduction in Streptococcus pyogenes. J Bacteriol **190:**4849-58.

57. **Rodriguez-Morales, O., M. Fernandez-Mora, I. Hernandez-Lucas, A. Vazquez, J. L. Puente, and E. Calva.** 2006. Salmonella enterica serovar Typhimurium ompS1 and ompS2 mutants are attenuated for virulence in mice. Infect Immun **74:**1398-402.

58. **Ross, I. L., and M. W. Heuzenroeder.** 2005. Discrimination within phenotypically closely related definitive types of *Salmonella enterica* serovar typhimurium by the multiple amplification of phage locus typing technique. J. Clin. Microbiol. **43:**1604-1611.

59. **Samant, S., H. Lee, M. Ghassemi, J. Chen, J. L. Cook, A. S. Mankin, and A. A. Neyfakh.** 2008. Nucleotide biosynthesis is critical for growth of bacteria in human blood. PLoS Pathog. **4:**e37.

60. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol.Biol.Evol. **6:**526-538.

61. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17:**14-56.

62. **Smith, J. M.** 1992. Analyzing the Mosaic Structure of Genes. Journal of Molecular Evolution **34:**126-129.

63. **Storey, J. D., and R. Tibshirani.** 2003. Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America **100:**9440-9445.

64. **Suyama, M., D. Torrents, and P. Bork.** 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res **34:**W609-12.

65. **Swaminathan, B., and P. Gerner-Smidt.** 2006. Foodborne disease trends and reports. Foodborne Pathog. Dis. **3:**220-221.

66. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673-4680.

67. **Thomson, N. R., D. J. Clayton, D. Windhorst, G. Vernikos, S. Davidson, C. Churcher, M. A. Quail, M. Stevens, M. A. Jones, M. Watson, A. Barron, A. Layton, D. Pickard, R. A. Kingsley, A. Bignell, L. Clark, B. Harris, D. Ormond, Z. Abdellah, K. Brooks, I. Cherevach, T. Chillingworth, J. Woodward, H. Norberczak, A. Lord, C. Arrowsmith, K. Jagels, S. Moule, K. Mungall, M. Sanders, S. Whitehead, J. A. Chabalgoity, D. Maskell, T. Humphrey, M. Roberts, P. A. Barrow, G. Dougan, and J. Parkhill.** 2008. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. Genome Res. **18:**1624-1637.

68. **Tracz, D. M., H. Tabor, M. Jerome, L. K. Ng, and M. W. Gilmour.** 2006. Genetic determinants and polymorphisms specific for human-adapted serovars of Salmonella enterica that cause enteric fever. J Clin Microbiol **44:**2007-18.

69. **Twiddy, S. S., C. H. Woelk, and E. C. Holmes.** 2002. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. J.Gen.Virol. **83:**1679-1689.

70. **Urwin, R., E. C. Holmes, A. J. Fox, J. P. Derrick, and M. C. Maiden.** 2002. Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. Mol.Biol.Evol. **19:**1686-1694.

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

72. **van Asten, A. J., and J. E. van Dijk.** 2005. Distribution of "classic" virulence factors among Salmonella spp. FEMS Immunol.Med.Microbiol. **44:**251-259.

73. **Wong, W. S., Z. Yang, N. Goldman, and R. Nielsen.** 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics **168:**1041-1051.

74. **Wood, M. W., M. A. Jones, P. R. Watson, S. Hedges, T. S. Wallis, and E. E. Galyov.** 1998. Identification of a pathogenicity island required for Salmonella enteropathogenicity. Molecular Microbiology **29:**883-891.

75. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput.Appl.Biosci. **13:**555-556.

76. **Yang, Z., R. Nielsen, N. Goldman, and A. M. Pedersen.** 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155:**431-49.

77. **Zhang, J. Z., R. Nielsen, and Z. H. Yang.** 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Molecular Biology and Evolution **22:**2472-2479.

CHAPTER 4

SALMONELLA 4,5,12:I:-: AN EMERGING *SALMONELLA* SEROTYPE THAT REPRESENTS MULTIPLE DISTINCT CLONES

ABSTRACT

The prevalence, among human clinical cases, of *Salmonella* 4,5,12:i:-, a serotype closely related to *Salmonella* Typhimurium but lacking second phase flagellar antigens, has increased considerably over the last 10 years. To probe the evolution and ecology of this emerging serotype, we characterized 190 *Salmonella* isolates initially classified as serotypes 4,5,12:i:- (n= 90) and Typhimurium (n=100) and obtained from various sources in the United States and Spain. These 190 isolates were characterized into six sequence types [determined by multilocus sequence typing (MLST)] and 79 pulsed-field gel electrophoresis (PFGE) types. The majority of 4,5,12:i:- and Typhimurium isolates (85 and 84 isolates, respectively) represented a single MLST type. Existing genome information revealed different genome deletions (which included genes responsible for phase 2 flagella expression) in four Spanish and one US 4,5,12:i:- isolate. Fifty-nine isolates of both serotypes, representing different source and geographical locations, as well as different molecular subtypes were thus screened for the presence of six genes and one specific region¸ which all showed variable presence, based on existing genomic information, among 4,5,12:i:- and Typhimurium strains. All 4,5,12:i:- isolates lacked the phase 2 flagella genes *fljA* and *fljB*, which were present in all Typhimurium isolates. While all Spanish 4,5,12:i:- isolates carried the same deletion surrounding *fljAB,* all but two US isolates showed a different genomic deletion; one of the two atypical US isolates represented the "Spanish" deletion genotype, while the other isolate represented a unique deletion

genotype. *Salmonella* 4,5,12:i:-, thus, appears to represent at least two common clones with different geographical distributions.

INTRODUCTION

Salmonella spp. are one of the most common causes of bacterial foodborne diseases worldwide (34). In the United States nontyphoidal *Salmonella* serotypes cause an estimated 1.4 million human salmonellosis cases, including approximately 550 deaths annually (27). Serotyping with the Kaufmann-White scheme is used commonly as a first step to differentiate *Salmonella* isolates. Serotyping of *Salmonella* isolates is based on lipopolysaccharide moieties on cell surface (O antigens) and the flagellar proteins (H antigens), as well as capsular protein antigens (Vi-antigen), which are only found in a few *Salmonella* serotypes (e.g., Typhi). According to Kaufmann-White scheme, *Salmonella* includes over 2,500 recognized serotypes (20). Many *Salmonella* are motile due to peritrichous flagella (28), which include a basal body, a propeller and a hook. The motility of *Salmonella* depends on the rotation of the flagellar propeller (i.e., the filament), which includes either FliC (phase-1 antigen) or FljB (phase-2 antigen) flagellin (11). Most of *Salmonella* serotypes, including *Salmonella* Typhimurium, are bi-phasic, meaning that they can express two distinct flagellar antigens (i.e., phase-1 and phase-2 antigens). Regulation of phase 1 and 2 antigen expression is under control of the recombinase Hin. This recombinase facilitates inversion of a promoter element so that it either (i) transcribes *fljB* (which encodes the phase 2 antigen FljB) and *fljA*, which encodes a repressor of *fliC*, the gene encoding the phase 1 antigen FliC (4, 37) or (ii) does not transcribe either of these genes. If this promoter is located in an orientation that does not allow for transcription of *fljB* and *fljA*, lack of a repression of *fliC* transcription leads to expression of phase 1 flagellar antigens.

Salmonella 4,5,12:i:- is a serotype that appears to be closely related to *Salmonella* Typhimurium (which has the serotype 4,5,12:i:1,2), but lacks the expression of second phase 1, 2 flagellar antigen. *Salmonella* 4,5,12:i:- was the 6th most common *Salmonella* serotype among human cases in the US in 2006 (10) and the 4th most common serotype among human isolates in Spain in 1998 (18). Overall, the prevalence of *Salmonella* 4,5,12:i:- among human cases has increased considerably in many countries in the world over the last 10 years (9, 10, 18, 29, 36). This *Salmonella* serotype has also been responsible for a number of human salmonellosis outbreaks over the last decades, including in Spain (1998), the US (2004 and 2007), and in Luxemburg (2006). *Salmonella* 4,5,12:i:- has also been isolated, particularly over the last decade, from a number of different foods and animals (1, 6, 13, 29, 38). While a number of separate studies, using molecular subtyping and characterization tools (e.g., genomic microarrays, PCR assays to test for gene presence/absence), have shown that serotype 4,5,12:i:- isolates from Spain (15, 18) and the US (1, 2, 38) are closely related to *Salmonella* Typhimurium, we are not aware of any comparative studies of serotype 4,5,12:i:- isolates from Europe and the US that have been published to date. In order to provide a better understanding of the transmission, ecology, and evolution of *Salmonella.* 4,5,12:i:-,we have assembled a collection of 190 serotype 4,5,12:i:- and Typhimurium isolates from various sources and from two countries, including the US and Spain. These isolates were characterized by different molecular subtyping methods (i.e., multilocus sequence typing [MLST] and pulsed-field gel electrophoresis [PFGE]), followed by characterization of selected isolates for genomic deletions that may be responsible for lack of phase 2 flagella expression. In combination with an analysis of existing genome data for a US serotype 4,5,12:i:- isolate (32) and genomic microarray data for Spanish serotype 4,5,12:i:- isolates (18), our data indicate that serotype 4,5,12:i- represents at least two discrete genotypes with distinct geographical

distributions, supporting the hypothesis that at least two distinct emergence events lead to the evolution of this serotype.

MATERIALS AND METHODS

***Salmonella* isolates.** A total of 190 *Salmonella* isolates initially serotyped as Typhimurium (n=100) and 4,5,12:i:- (n=90) were used in this study (Table 4.1). These isolates were obtained from different states in the US, including New York (69 isolates), Washington (52 isolates) and Georgia (26 isolates), and Spain (43 isolates) as well as different sources, including human clinical isolates, foods, cattle, poultry, and other warm-blooded animals (Table 4.1). Human isolates from New York State and Washington were obtained from the New York State Department of Health and the Washington State Department of Health, respectively. Bovine isolates from New York State and Washington State were obtained from the Animal Health Diagnostic Center (AHDC) at Cornell University and the Washington Animal Disease Diagnostic Laboratory, Pullman, respectively; isolates from foods collected in New York State were obtained from the Food and Drug Administration (FDA) (Table A2 [S4..1]). *Salmonella* Typhimurium and 4,5,12:i:- isolates from Georgia have previously been described (38). *Salmonella* Typhimurium and 4,5,12:i:- isolates from Spain have also been described previously (5, 18) and were provided by Dr. Garaizar, University of the Basque Country, Vitoria-Gasteiz, Spain. Detailed information for all isolates (see Table A2 [S4..1]), including serotype, source, gene sequence data, allelic types and pulsed field gel electrophoresis (PFGE) patterns are available via the Pathogen Tracker website (http://www.pathogentracker.net).

While serotype data were provided for all isolates, isolates that were initially classified as serotype 4,5,12:i:- but contained an intact copy of the phase 2 flagella gene *fljB* were re-submitted for serotyping at the National Veterinary Service

Table 4. 1.. *Salmonella* 4,5,12:i:- and Typhimurium isolates used in this study

State, Country	No. of isolates from[a]						
	Bovine	Human	Poultry	Food	Others	Non-domestic birds	Total
Serotype 4,5,12:i:-							
Georgia, US	2 (2)	0	10 (7)	0	0	1 (1)	13 (10)
New York, US	7 (5)	9 (4)	0	2 (1)	0	0	18 (10)
Washington, US	0	40 (10)	0	0	2 (1)	0	42 (11)
Spain	0	11	0	2	0	0	13 (10)
Total							86 (41)
Serotype Typhimurium							
Georgia, US	6	0	5 (2)	0	0	1 (1)	12 (3)
New York, US	22	29 (4)	0	0	0	0	51 (4)
Washington, US	1	6 (2)	0	0	0	0	7 (2)
Spain[b]	0	5	0	8	17	0	30 (5)
Total							100 (14)
Inconsistent (Typhimurium or 4,5,12:i:-)[c]							
Georgia, US	0	0	0	0	0	1	1 (1)
Washington, US	3 (1)	0	0	0	0	0	3 (3)
Total							4 (4)
Total							190 (59)

[a]numbers in brackets represent numbers of isolates that were used for further PCR screens to determine the presence/absence of selected genes and one specific region.
[b]Spanish *Salmonella* Typhimurium isolates in category "others" were obtained from water and other environment; detailed information for Spanish isolates is in Supplementary Table 1.
[c]these isolates were serotyped as 4,5,12:i:- in one replicate and Typhimurium in another replicate (including one isolates that was classified as 4,5,12:i:- in two replicates and Typhimurium in one replicate) and were thus designated as "inconsistent".

Laboratories (NVSL). Isolates that were serotyped as 4,5,12:i:- in one replicate and Typhimurium in another replicate (including one isolate that was classified as 4,5,12:i:- in two replicates and Typhimurium in one replicate) were designated as "inconsistent serotype" isolates.

Multilocus Sequence Typing (MLST). While traditional MLST schemes target 7 housekeeping genes (24), we initially used a previously reported MLST scheme targeting three genes (i.e., *manB, mdh, fimA*) (2, 33) to characterize all isolates used in this study. While this MLST scheme had previously been shown to provide similar discriminatory power as a 7-gene MLST, the 3-gene MLST only allowed for limited discrimination among the isolates used here. We thus also sequenced a 826 nt fragment of a 4th gene (*aroC)* in all isolates to determine whether use of additional genes would increase discriminatory power. *aroC* was chosen as an additional gene as it was found to represent the greatest number of different allelic types among all isolates in the 7-gene MLST database for *Salmonella* in July 2007 (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica). Allelic types for *fimA, mdh,* and *manB* and 3-gene sequence types (STs) were assigned to be consistent with previous studies published by our group (2, 3, 33). Allelic types for *aroC* were also assigned to be consistent with 7-gene MLST Max Planck Institute database (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica). For example, 3-gene ST6 includes the same allelic type combination for 3 genes as reported in two studies by Alcaine et al. (2, 3), while *aroC* allelic type 18 for is identical to allelic type AROC18 in the 7-gene MLST database (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica). STs were also determined based on allelic types for all four genes; these STs do not correspond to any previously reported STs.

Salmonella DNA used as template for PCR reactions performed for MLST was purified using QIAmp DNA Mini kit (QIAGEN Inc., Chatsworth, CA). PCR primers

for *manB, mdh,* and *fimA* have previously been reported (2, 33); PCR primers for *aroC* amplification were obtained from the *Salmonella enterica* MLST database at Max Plank Institute (http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica); all primers used are summarized in Table A2 [S4.2]. PCR products were purified using Exonuclease I (USB, Cleveland, OH) and shrimp alkaline phosphatase (USB, Cleveland, OH). Purified PCR products were sequenced using Big Dye Terminator chemistry and Ampli Taq-FS DNA Polymerase and sequencing reaction were analyzed using the Applied Biosystems Automated 3730 DNA Analyzer at the Cornell University Life Sciences Core Laboratories Center. Sequences were assembled and proofread using SeqMan and aligned using the Clustal W algorithm implanted in MegAlign (DNAStar Inc., Madison, WI).

Phylogenetic analysis. Phylogenetic analysis was performed using all 3-genes STs found among *Salmonella* Typhimurium and 4,5,12:i:- isolates as well as 3-gene STs available for other serotypes (e.g., STs reported by Alcaine et al. [2]). As *manB* is duplicated in some isolates, thus yielding sequence data not suitable for phylogenetic analyses (2), STs representing sequences with *manB* duplications were not included in the phylogenetic analyses. For each unique ST, the sequences of the three genes were concatenated. Concatenated sequences were aligned using MegAlign (DNAStar Inc., Madison, WI) and MACCLADE version 4.08 (Sunderland, Massachusetts: Sinauer Associates Inc). MODELTEST (30) was used to determine the best fitting model of evolution (i.e., TrN+I+G), which was used for construction of a maximum likelihood (ML) tree. The ML tree was constructed, using the concatenated 3-gene MLST sequences, using PAUP* Portable version 4.0b10 for Unix (35). No phylogenetic analyses were performed on the 4-gene MLST data as insufficient ST data are available for serotypes other than Typhimurium and 4,5,12:i:-.

Pulsed-Field Gel Electrophoresis (PFGE). XbaI PFGE was performed according to the Centers for Disease Control and Prevention PulseNet protocol (http://www.cdc.gov/pulsenet/protocols/ecoli_salmonella_shigella_protocols.pdf) (31). Analysis of PFGE types was performed using the BioNumerics Software package (Applied Maths 1998-2004, Austin, TX). Similarity analysis was performed by using the Dice coefficient and clustering was performed using the unweighted pair group method with arithmetic mean (UPGMA).

Simpson's index of diversity. Simpson's index of diversity (SID) was calculated as previously described (23).

Analysis of microarray and genome sequence data to identify gene deletions in serotype 4,5,12:i:-. In order to identify gene deletions and other genomic differences between serotypes 4,5,12:i:- and Typhimurium, we used (i) comparative genomic microarray data on gene presence/absence patterns in four Spanish serotype 4,5,12:i:- isolates (as compared to *Salmonella* Typhimurium LT2) (as reported by Garaizar et al. [18]), and (ii) full genome sequence data for the US serotype 4,5,12:i:- isolate CVM23701 (GenBank accession no. NZ_ABAO00000000, http://msc.jcvi.org/salmonella/salmonella_enterica_subsp__enterica_serovar_4__5__12_i___str__cvm23701/index.shtml) (32) and *Salmonella* Typhimurium LT2 (AE006468). Genomic microarray data reported by Garaizar et al. (18) revealed one genomic deletion (termed cluster V) in serotype 4,5,12,:i- , which included deletion of *fljB* (encoding phase 2 flagella, thus providing a functional explanation for the absence of phase 2 flagellar expression observed in serotype 4,5,12:i:-) as well as a second deletion (termed cluster IV) located approximately 16 kb 5' of cluster V. BLAST searches were used to determine whether genes in cluster IV (genes STM2694 to STM2740) and cluster V (genes STM2758 to STM2773) as well as genes in the intervening regions and upstream and downstream were present in the serotype

4,5,12:i:- isolate CVM23701 genome. Specifically, BLAST searches were used to determine whether *Salmonella* Typhimurium LT2 genes STM2691 through STM2775 were present in the CVM23701 genome. BLAST searches were performed using the National Center for Biotechnology Information (NCBI) BLAST tools and gene sequences downloaded from the J. Craig Venter Institute Comprehensive Microbial Resource (JCVI CMR). BLAST searches were also used to determine whether genes in three other clusters (I, II, and III), which were previously reported to be present in *Salmonella* Typhimurium LT2, but absent in Spanish 4,5,12:i:- isolates, were present in the genome sequence for the US 4,5,12:i:- isolate CVM23701.

PCR-based characterization of gene deletion patterns in representative serotype 4,5,12:i:-and Typhimurium isolates. Based on our analyses of (i) the genomic microarray data reported by Garaizar et al. (18) and (ii) the serotype 4,5,12:i:- isolate CVM23701 genome, we designed PCR primers to test for the presence of selected genes in clusters IV and V and adjoining regions (Table 4.2). We initially designed 8 primer sets for genes that are at the junctions of clusters IV and V (as reported by Garaizar et al. [18]); these primers target STM2692, STM2694, STM2740, STM2741, STM2757, STM2758, STM2773 (*iroB*) and STM2774 (see Figure 4.1 for primer locations). In addition, we designed primer sets for (i) two genes (*fljA, fljB*) absent from both the 4,5,12:i:- isolates from Spain (based on the genomic microarray data reported by Garaizar et al., 2002) and CVM23701 genome as well as for (ii) one gene (*hin*) present in CVM23701 and absent in the 4,5,12:i:- isolates from Spain. We also designed one set of primers targeting a region found upstream of *hin* in only the CVM23701 genome; this region was designated as "STM1053-1997 region", as primers designed are located in genes with homology to STM1053 (forward primer) and STM1997 (reverse primer) (see Figure 4.1). PCR was performed on DNA purified using the QIAmp DNA Mini Kit (Qiagen Inc, Chartsworth) as detailed below, using

Table 4. 2. PCR conditions and primers for six genes and one region that show variable presence among serotype 4,5,12:i:- isolates from the US and Spain

Gene	Gene function[a]	Amplicon Size	Primers (5' to 3')[b]	Temperature (time) for Denaturation	Annealing	Extension
fljA	Repressor of phase-1 flagellin gene	642 bp	F: TTC ATT AGG TCC CCT CCG G R: ATT CAG CCC CGT GAA TTC GGG	95°C (10 min)	55ºC (45 sec)	72ºC (1 min)
fljB	Phase-2 flagellin structural protein	561 bp	F: TTTACCGTCTACGCCACCC R: GGTACTACACTGGATGTATCGGG	95ºC (10 min)	52ºC (45 sec)	72ºC (1 min)
hin	H inversion: regulation of flagellar gene expression	570 bp	F: TGG CTA CTA TTG GGT ATA TTC GGG R: AAT TCA TTC GTT TTT TTA TGC GGC	95ºC (10 min)	52ºC (45 sec)	72ºC (1 min)
STM1053-1997	-	614 bp	F: CCA TTT TTA TAC TGC CAG TCG CC R: CAG CGA AAT ACT GAT GGC GG	95ºC (10 min)	55ºC (45 sec)	72ºC (1 min)
STM2740	Integrase, phage family	980 bp	F: AAT GTG GAG ATC GCT GGC GCG R: AGT TCG CCG CCG AAC CCC	95ºC (2 min)	55ºC (45 sec)	72ºC (1.5 min)
STM2757	Putative cytoplasmic protein	717 bp	F: ATG ATG ATG GCG TAA TGG CGC R: AAA ACG TTC CGG TGC GGC G	95ºC (10 min)	55ºC (45 sec)	72ºC (1 min)

Table 4. 2. (Continued)

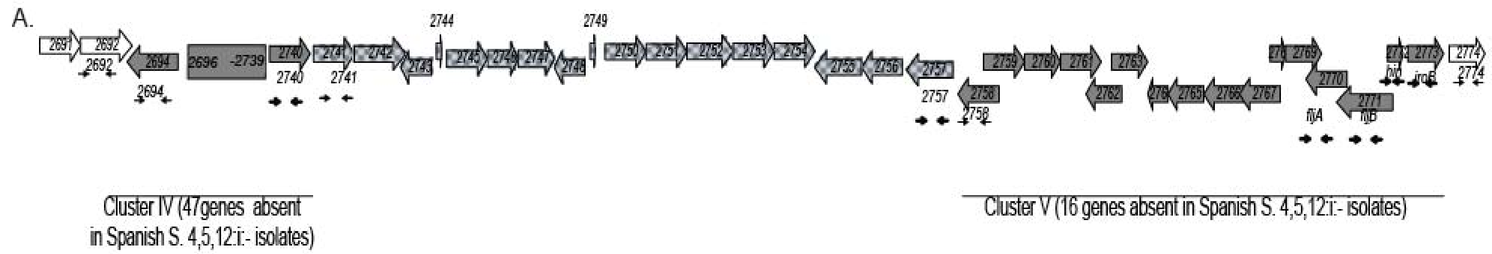
iroB	Glucosyl-transferase homolog	858 bp	F: TTC GAT TCG GAA GCG GGT TAT CGC CG R: CTC GCG AAG CGC GCG	95ºC (2 min)	65-55ºC TD[c] (45 sec)	72ºC (1.5 min)

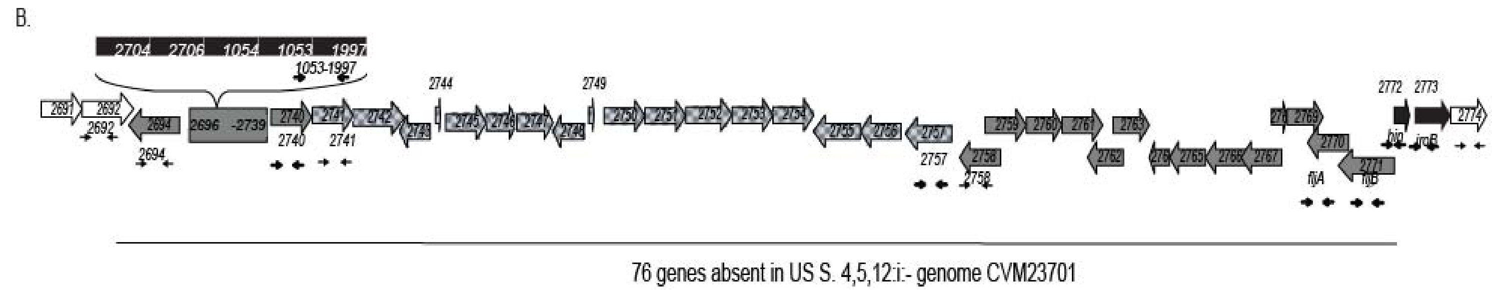
[a]Gene functions for *S.* Typhimurium LT2 were obtained from JVCI CMR (J. Craig Venter Institute The Comprehensive Microbial Resource) website

[b]R: reverse primer; F: forward: primer

[c]TD, Touch down PCR; annealing temperatures decreased 0.5 ºC/cycle during the first 20 cycles, followed by 20 cycles at 55°C

Figure 4. 1. Deduced genome structure for the genomic region between STM2691 and STM2774 for (A) four *Salmonella* 4,5,12:i:- isolates from Spain (based on genomic microarray data reported by Garaizar et al., 2002) and (B) the US serotype 4,5,12:i:- isolate CVM 23701 (based on an unfinished genome sequence reported by Rosovitz et al. [http://msc.jcvi.org/salmonella/salmonella_enterica_subsp__enterica_serovar_4__5__12_i___str__cvm23701/index.shtml]). Genes are represented as open arrows or boxes; gene numbers (e.g., 2691) represent locus numbers based on primary annotation of *Salmonella* Typhimurium LT2 (with the prefix "STM"). White color represents the genes present in *S.* Typhimurium LT2 and all 4,5,12:i:- isolates; grey represents genes present in *S.* Typhimurium LT2 and absent from both US and Spanish 4,5,12:i:- isolates; halftone pattern represents genes present *Salmonella* Typhimurium LT2 and Spanish 4,5,12:i:- isolates, but absent from the US 4,5,12:i:- isolate. Black represents a unique insertion in the US 4,5,12:i:- isolate CVM 23701, which includes genes with genes with full or partial homology with the *Salmonella* Typhimurium LT2 genes STM1054 (94 % homology with LT 2 over 79 % gene length), STM1053 (93 % homology with LT 2 over 85 % gene length), STM1997 (92 % homology with LT 2 over 42 % gene length), STM2704 (87 % homology with LT 2 over 100 % gene length), STM2706 (87 % homology with LT 2 over 18 % gene length); *hin* and *iroB*¸ which are present in LT2 and the US 4,5,12:i:- isolate are also shown in black. Small arrows represent PCR primers, including five primer sets (See Supp. Table S2) used only for an initial screen of 6 isolates (two 4,5,12:i:- isolates from each Spain and the US and one Typhimurium isolate from each Spain and the US), shown as thin black arrows, as well as 7 primer sets (See Table 4.2) used to screen a total of 59 isolates, shown as thick arrows.


A.
Cluster IV (47genes absent
in Spanish S. 4,5,12:i:- isolates)
Cluster V (16 genes absent in Spanish S. 4,5,12:i:- isolates)


B.
2704 2706 1054 1053 1997
1053-1997
76 genes absent in US S. 4,5,12:i:- genome CVM23701

either Ampli Taq Gold (Applied Biosystems, Foster City, CA) or Go Taq (Promega, Madison, WI).

All PCR primers were used initially to screen for gene presence/absence among four serotype 4,5,12:i:- isolates (two each from Spain and the US) as well as two Typhimurium isolates (one each from Spain and the US). Subsequently, primers targeting six genes (i.e., *STM2740, STM2757, fljA, fljB, hin,*and *iroB*) and the STM1053-1997 region (see Table 2 for all primers) were used to screen for presence/absence of the selected genes among 59 representative isolates, representing serotypes 4,5,12:i:- (41 isolates) and Typhimurium (14 isolates) as well as all four isolates with inconsistent serotype data (i.e., serotyped as 4,5,12:i:- and Typhimurium). These isolates were selected to represent all PFGE types and STs found among the Spanish isolates. Isolates obtained in the US were selected to represent the most common PFGE types found among different isolate sources (e.g., human, food, cattle, poultry, non-domestic birds); for serotype 4,5,12:i:-, isolates from the US were selected to assure inclusion of at least one representative of each ST and PFGE.

RESULTS AND DISCUSSION

In order to better understand the evolution and ecology of *Salmonella* 4,5,12:i:-, we characterized 190 *Salmonella* 4,5,12:i:- and Typhimurium isolates from the US and Spain with a variety of molecular methods. Overall, our data indicate that (i) *Salmonella* 4,5,12:i:- and Typhimurium represent a highly clonal group, which can be differentiated by PFGE, (ii) US and Spanish 4,5,12:i:- isolates show different patterns of gene deletion in the regions encoding phase 2 flagella and represent distinct PFGE patterns, and (iii) in addition to two common 4,5,12:i:- genotypes (designated here as the “Spanish” and the “US” 4,5,12:i:- clone), other 4,5,12:i:- genotypes exist. We thus

conclude that *Salmonella* 4,5,12:i:- most likely represents multiple clones that emerged through independent deletion events.

***Salmonella* 4,5,12:i:- and Typhimurium represent a highly clonal group, which can be differentiated by PFGE.** Among the 190 *Salmonella* initially characterized as serotypes Typhimurium (100 isolates) and 4,5,12:i:-, we identified six distinct sequence types (STs) based on a four gene MLST scheme (Table 4.3). A single ST (ST1) represented the vast majority of Typhimurium and 4,5,12:i:- isolates; 84 out of 100 Typhimurium and 85 out of 86 serotype 4,5,12:i:- isolates were classified as ST1. Analyses of the relevant genes in the genomes of *Salmonella* LT2 and the US *Salmonella* 4,5,12:i:- isolate CVM23701 showed that these two strains also represent ST1. One serotype 4,5,12:i:- isolate from Spain represented ST 3; ST 3 also represented 7 US *Salmonella* Typhimurium isolates and one US isolate with inconsistent serotype data (i.e., serotyped as 4,5,12:i:- and Typhimurium in replicate experiments). ST3 differs from ST1 by only one nucleotide difference in *manB*. While serotype 4,5,12:i:- represented only two STs (Simpson Index of Discrimination [SID] = 0.02), Typhimurium isolates represented six STs (SID=0.29), indicating considerably higher ST diversity among the Typhimurium isolates characterized. Guerra et al. (21) previously also proposed that *Salmonella* 4,5,12:i:- represents a lower diversity as compared to *Salmonella* Typhimurium, even though their molecular subtype study only used 16 serotype 4,5,12:i:- and two Typhimurium isolates from Spain.

Phylogenetic analyses of 3-gene MLST data (performed here; Figure 4.2) also supported that serotypes 4,5,12:i:- and Typhimurium are closely related and highly clonal as shown by the fact that all serotype 4,5,12:i:- and Typhimurium STs form a single branch with strong bootstrap support. The observation that serotypes 4,5,12:i:- and Typhimurium represents a single highly clonal group is consistent with a number

Figure 4. 2. Phylogenetic tree for all 3-gene (*fimA, manB,* and *mdh*) STs identified among 4,5,12:i:- and Typhimurium isolates as well as selected isolates representing other *Salmonella* serotypes (these STs were taken from Alcaine et al. [2]). For each unique ST, *fimA, manB,* and *mdh* sequences were concatenated and aligned, followed by construction of a maximum likelihood (ML) tree (100 bootstrap replicates), using the TrN+I+G model of evolution (identified by MODELTEST as the appropriate model for the data set used). While a number of nodes in this tree were supported by high bootstrap values, bootstrap support is only shown for the clade containing serotype 4,5,12:i:- and Typhimurium isolates.

ST81_Saintpaul
ST82_Berta
ST83_Berta
ST77_Hartford
ST40_Hadar
ST84_Stanley
ST39_Stanley
ST27_Blockley
ST17_Kentucky
ST34_Kentucky
ST3_Heidelberg
ST50_Heidelberg
ST18_Senftenberg
ST26_Heidelberg
ST48_Paratyphi B
ST74_Muenchen
ST78_Newport
ST46_Newport
ST76_Newport
ST24_Dublin
ST23_Dublin
ST22_Dublin
ST36
ST14_Enteritidis
ST15_Enteritidis
ST73_Mbandaka
ST2_Agona
ST1_Agona
ST79_Weltvreden
90 %
ST8_T.Copenhagen
ST8_Typhimurium
ST8_Rough
ST47_Typhimurium
ST49_Typhimurium
ST6_Typhimurium
ST7_Typhimurium
ST51_4,5,12:i:-
ST43_Thompson
ST38_Saintpaul
ST80_Paratyphi C
ST86_Anatum
ST75_Adelaide
ST11_Newport
ST33_Newport
ST11_Bardo
ST85_Newport
ST31_Litchfield
ST32_Java
ST16_Kentucky
ST29_Pomona
ST30_Give
ST28_Poona
ST4_Schwarzengrund
ST44_Muenster
ST12_Amsterdam
ST37_Panama
ST9_Montevideo
ST21_Javiana
ST19_Javiana
ST19_Panama

0.0060

TABLE 4. 3. Distribution of 4-Gene Sequence Types (STs) among *Salmonella* isolates

4-Gene Sequence Type[a]	No. isolates among			Total
	S. 4,5,12:i:-	*S.* Typhimurium	Inconsistent[b]	
1	85	84	3	172
2	0	1	0	1
3	1	7	1	9
7	0	5	0	5
8	0	1	0	1
9	0	2	0	2

[a]sequence types were based on allelic types for partial *fimA, mdh, manB* and *aroC* sequences

[b]these isolates were serotyped as 4,5,12:i:- in one replicate and Typhimurium in another replicate (including one isolates that was classified as 4,5,12:i:- in two replicates and Typhimurium in one replicate) and were thus designated as "inconsistent"

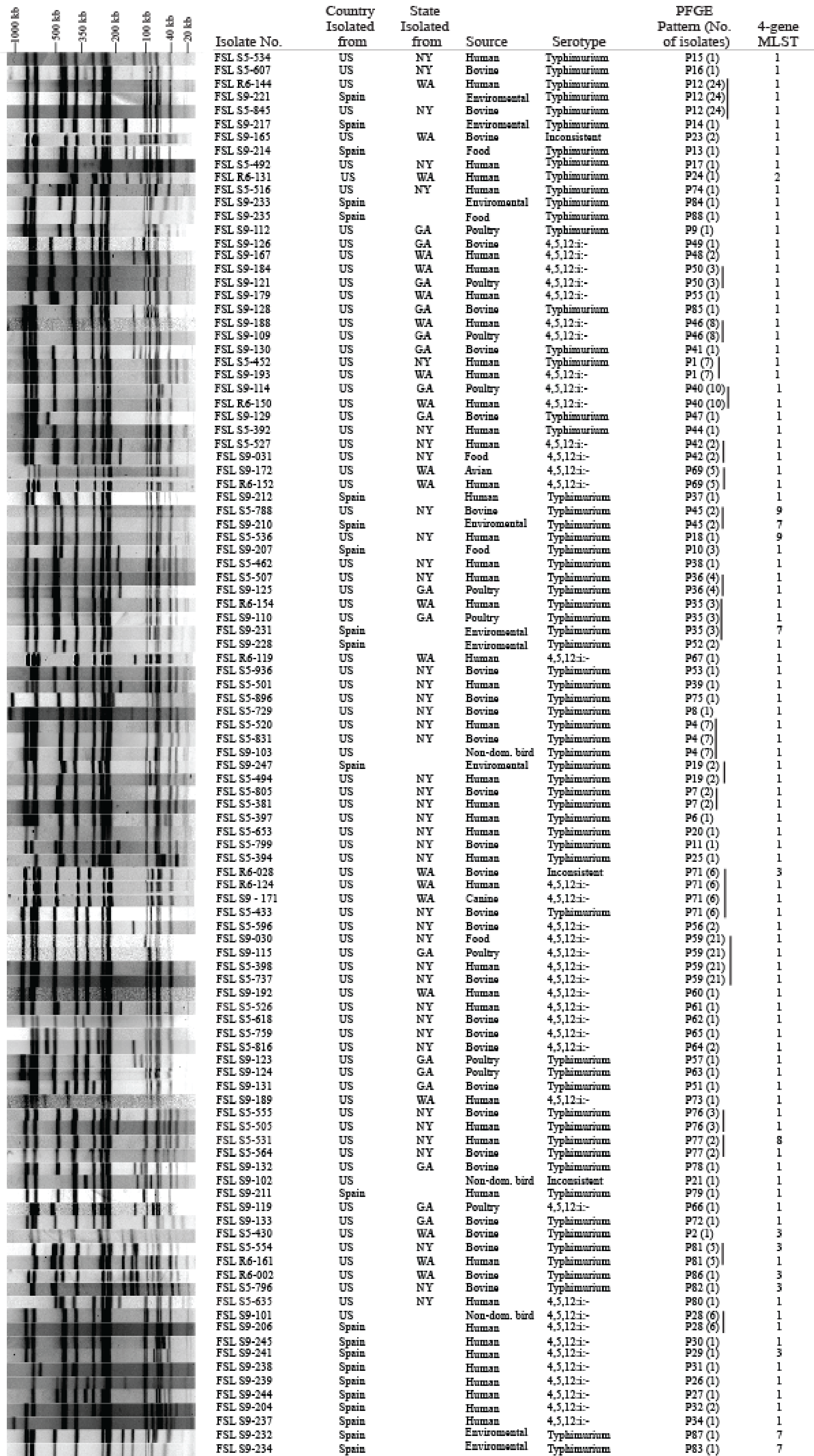
of studies (2, 14) that have shown that *Salmonella* Typhimurium is highly clonal. The observation that all serotype 4,5,12:i:- isolates characterized here share identical STs with Typhimurium isolates is consistent with a number of studies (e.g., references 1, 6, 13, 14, 16 and 38) that have shown, using different molecular subtyping methods (e.g. PFGE, MLST, phage typing), that *Salmonella* 4,5,12:i:- isolates are closely related to *Salmonella* Typhimurium. While, according to serological characterization, *Salmonella* 4,5,12:i:- is also closely related to *Salmonella* serotypes Lagos (4,5,12:i:1,5), Agama (4,12:i:1,6), Farsta (4,12:i:e,n,x), Tsevie (4,12:i:e,n,z_{15}), Cloucester (1,5,12,27:i:l,w), Tumodi (1,4,12:i:z_6), and an unnamed subspecies II serotype (4,5,27:i:z_{35}) (28), we are not aware of any data that suggest any of these closely related serotypes might be an ancestor of a 4,5,12:i:- strain. Echeita et al. (16) specifically reported that two genomic regions, i.e., a 1,000 bp *fliB-fliA* intergenic region and a 162 bp region specific for DT104 and DT U302 phage types, were absent in *S.* Lagos, but present in *S.* Typhimurium phage types DT104 and DT U302, as well as in Spanish 4,5,12:i:- isolates, suggesting that Spanish serotype 4,5,12:i:- isolates are closely related to *Salmonella* Typhimurium DT104 and DT U302, and are unlikely to have originated from a serotype Lagos ancestor. In our analysis of *aroC* allelic types (including *aroC* ATs obtained from the Max Planck Institute [MPI] MLST website http://web.mpiib-berlin.mpg.de/mlst/dbs/Senterica), we also found that the *aroC* AT (i.e., AROC146) for the only serotype Lagos isolate in this database was distinct from the *aroC* AT found among all 4,5,12:i:- isolates and all but one Typhimurium isolates characterized here (i.e., AT 10, which differs by 4 nucleotides from AROC146), further supporting that serotype Lagos is unlikely to be the ancestor of serotype 4,5,12:i:-. Similarly, the *aroC* AT for the one serotype Agama isolate represented in the MPI MLST database represents an *aroC* allelic type (AROC136), which is distinct from AT 10 (4 nt differences between AT10 and AROC136), suggesting that serotype

Agama is unlikely to be the ancestor of serotype 4,5,12:i:-. While, overall, these data suggesting that 4,5,12:i:- is a monophasic variant of serotype Typhimurium, the rare serotypes Farsta, Tsevie, Cloucester, Tumodi, and subspecies II serotype 4,5,27:i:z_{35} cannot be definitively excluded as ancestors of serotype 4,5,12:i:- until isolates representing these serotypes have been characterized by molecular methods and compared to serotype Typhimurium and 4,5,12:i:- isolates.

As PFGE has been shown to be a highly discriminatory subtyping method for a number of *Salmonella* serotypes (1, 14, 38), we further characterized all 190 *Salmonella* isolates using PFGE with the enzyme XbaI. Overall, we identified 79 PFGE patterns (SID =0.96) among all 190 isolates. A total of 29 and 50 PFGE types were differentiated among the 86 and 100 serotype 4,5,12:i:- and Typhimurium isolates (SID = 0.91 and 0.93, respectively); the four isolates with inconsistent serotypes represented three different PFGE patterns. Overall, these data support previous studies which have shown that serotype 4,5,12:i:- isolates represent considerable PFGE diversity (1, 21, 38) and that PFGE, in general, allows for more sensitive subtype discrimination among *Salmonella* isolates as compared to MLST (12, 17, 22).

Interestingly, two PFGE patterns (P1 and P71, see Figure 4.3) were shared by serotype 4,5,12:i:- and Typhimurium isolates. PFGE type P1 was found in 3 serotype 4,5,12:i:- and 4 serotype Typhimurium isolates from the US, while P71 represented four serotype 4,5,12:i:- and one serotype Typhimurium isolate as well as one isolate with an inconsistent serotype, all isolated in the US. These observations extend previous observations by de la Torre et al. (13) and Zamperini et al. (38). de la Torre et al. (13) showed that at least one XbaI and one BlnI PFGE type were shared between Spanish 4,5,12:i:- and Typhimurium isolates, even though these two serotypes never shared the same combined XbaI/BlnI PFGE type (13). However, Zamperini et al.

Figure 4. 3. Representative XbaI PFGE patterns for *Salmonella* 4,5,12:i:- and Typhimurium isolates as well as four isolates with inconsistent serotype data (i.e., isolates that were initially identified as 4,5,12:i:- , but were classified as Typhimurium when they were re-submitted for serotyping). PFGE types shown represent all 79 unique types found among the 190 isolates characterized. If identical PFGE types were found among isolates representing two serotypes, different sources (e.g., human and bovine), or different countries, one representative from each group was included in this figure; solid vertical lines indicate multiple isolates with identical PFGE patterns. For example PFGE pattern 28 (P28) was identified in five 4,5,12:i:- isolates from Spain and one 4,5,12:i:- isolate from a non-domestic bird in the US. Number of isolates with a given PFGE type is indicated in parentheses after the PFGE type designation.



Isolate No.	Country Isolated from	State Isolated from	Source	Serotype	PFGE Pattern (No. of isolates)	4-gene MLST
FSL S5-534	US	NY	Human	Typhimurium	P15 (1)	1
FSL S5-607	US	NY	Bovine	Typhimurium	P16 (1)	1
FSL R6-144	US	WA	Human	Typhimurium	P12 (24)	1
FSL S9-221	Spain		Enviromental	Typhimurium	P12 (24)	1
FSL S5-845	US	NY	Bovine	Typhimurium	P12 (24)	1
FSL S9-217	Spain		Enviromental	Typhimurium	P14 (1)	1
FSL S9-165	US	WA	Bovine	Inconsistent	P23 (2)	1
FSL S9-214	Spain		Food	Typhimurium	P13 (1)	1
FSL S5-492	US	NY	Human	Typhimurium	P17 (1)	1
FSL R6-131	US	WA	Human	Typhimurium	P24 (1)	2
FSL S5-516	US	NY	Human	Typhimurium	P74 (1)	1
FSL S9-233	Spain		Enviromental	Typhimurium	P84 (1)	1
FSL S9-235	Spain		Food	Typhimurium	P88 (1)	1
FSL S9-112	US	GA	Poultry	Typhimurium	P9 (1)	1
FSL S9-126	US	GA	Bovine	4,5,12:i:-	P49 (1)	1
FSL S9-167	US	WA	Human	4,5,12:i:-	P48 (2)	1
FSL S9-184	US	WA	Human	4,5,12:i:-	P50 (3)	1
FSL S9-121	US	GA	Poultry	4,5,12:i:-	P50 (3)	1
FSL S9-179	US	WA	Human	4,5,12:i:-	P55 (1)	1
FSL S9-128	US	GA	Bovine	Typhimurium	P85 (1)	1
FSL S9-188	US	WA	Human	4,5,12:i:-	P46 (8)	1
FSL S9-109	US	GA	Poultry	4,5,12:i:-	P46 (8)	1
FSL S9-130	US	GA	Bovine	Typhimurium	P41 (1)	1
FSL S5-452	US	NY	Human	Typhimurium	P1 (7)	1
FSL S9-193	US	WA	Human	4,5,12:i:-	P1 (7)	1
FSL S9-114	US	GA	Poultry	4,5,12:i:-	P40 (10)	1
FSL R6-150	US	WA	Human	4,5,12:i:-	P40 (10)	1
FSL S9-129	US	GA	Bovine	Typhimurium	P47 (1)	1
FSL S5-392	US	NY	Human	Typhimurium	P44 (1)	1
FSL S5-527	US	NY	Human	4,5,12:i:-	P42 (2)	1
FSL S9-031	US	NY	Food	4,5,12:i:-	P42 (2)	1
FSL S9-172	US	WA	Avian	4,5,12:i:-	P69 (5)	1
FSL R6-152	US	WA	Human	4,5,12:i:-	P69 (5)	1
FSL S9-212	Spain		Human	Typhimurium	P37 (1)	1
FSL S5-788	US	NY	Bovine	Typhimurium	P45 (2)	9
FSL S9-210	Spain		Enviromental	Typhimurium	P45 (2)	7
FSL S5-536	US	NY	Human	Typhimurium	P18 (1)	9
FSL S9-207	Spain		Food	Typhimurium	P10 (3)	1
FSL S5-462	US	NY	Human	Typhimurium	P38 (1)	1
FSL S5-507	US	NY	Human	Typhimurium	P36 (4)	1
FSL S9-125	US	GA	Poultry	Typhimurium	P36 (4)	1
FSL R6-154	US	WA	Human	Typhimurium	P35 (3)	1
FSL S9-110	US	GA	Poultry	Typhimurium	P35 (3)	1
FSL S9-231	Spain		Enviromental	Typhimurium	P35 (3)	7
FSL S9-228	Spain		Enviromental	Typhimurium	P52 (2)	1
FSL R6-119	US	WA	Human	4,5,12:i:-	P67 (1)	1
FSL S5-936	US	NY	Bovine	Typhimurium	P53 (1)	1
FSL S5-501	US	NY	Human	Typhimurium	P39 (1)	1
FSL S5-896	US	NY	Bovine	Typhimurium	P75 (1)	1
FSL S5-729	US	NY	Bovine	Typhimurium	P8 (1)	1
FSL S5-520	US	NY	Human	Typhimurium	P4 (7)	1
FSL S5-831	US	NY	Bovine	Typhimurium	P4 (7)	1
FSL S9-103	US		Non-dom. bird	Typhimurium	P4 (7)	1
FSL S9-247	Spain		Enviromental	Typhimurium	P19 (2)	1
FSL S5-494	US	NY	Human	Typhimurium	P19 (2)	1
FSL S5-805	US	NY	Bovine	Typhimurium	P7 (2)	1
FSL S5-381	US	NY	Human	Typhimurium	P7 (2)	1
FSL S5-397	US	NY	Human	Typhimurium	P6 (1)	1
FSL S5-653	US	NY	Human	Typhimurium	P20 (1)	1
FSL S5-799	US	NY	Bovine	Typhimurium	P11 (1)	1
FSL S5-394	US	NY	Human	Typhimurium	P25 (1)	1
FSL R6-028	US	WA	Bovine	Inconsistent	P71 (6)	3
FSL R6-124	US	WA	Human	4,5,12:i:-	P71 (6)	1
FSL S9 - 171	US	WA	Canine	4,5,12:i:-	P71 (6)	1
FSL S5-433	US	NY	Bovine	Typhimurium	P71 (6)	1
FSL S5-596	US	NY	Bovine	4,5,12:i:-	P56 (2)	1
FSL S9-030	US	NY	Food	4,5,12:i:-	P59 (21)	1
FSL S9-115	US	GA	Poultry	4,5,12:i:-	P59 (21)	1
FSL S5-398	US	NY	Human	4,5,12:i:-	P59 (21)	1
FSL S5-737	US	NY	Bovine	4,5,12:i:-	P59 (21)	1
FSL S9-192	US	WA	Human	4,5,12:i:-	P60 (1)	1
FSL S5-526	US	NY	Human	4,5,12:i:-	P61 (1)	1
FSL S5-618	US	NY	Bovine	4,5,12:i:-	P62 (1)	1
FSL S5-759	US	NY	Bovine	4,5,12:i:-	P65 (1)	1
FSL S5-816	US	NY	Bovine	4,5,12:i:-	P64 (2)	1
FSL S9-123	US	GA	Poultry	Typhimurium	P57 (1)	1
FSL S9-124	US	GA	Poultry	Typhimurium	P63 (1)	1
FSL S9-131	US	GA	Bovine	Typhimurium	P51 (1)	1
FSL S9-189	US	WA	Human	4,5,12:i:-	P73 (1)	1
FSL S5-555	US	NY	Bovine	Typhimurium	P76 (3)	1
FSL S5-505	US	NY	Human	Typhimurium	P76 (3)	1
FSL S5-531	US	NY	Human	Typhimurium	P77 (2)	8
FSL S5-564	US	NY	Bovine	Typhimurium	P77 (2)	1
FSL S9-132	US	GA	Bovine	Typhimurium	P78 (1)	1
FSL S9-102	US		Non-dom. bird	Inconsistent	P21 (1)	1
FSL S9-211	Spain		Human	Typhimurium	P79 (1)	1
FSL S9-119	US	GA	Poultry	4,5,12:i:-	P66 (1)	1
FSL S9-133	US	GA	Bovine	Typhimurium	P72 (1)	1
FSL S5-430	US	WA	Bovine	Typhimurium	P2 (1)	3
FSL S5-554	US	NY	Bovine	Typhimurium	P81 (5)	3
FSL R6-161	US	WA	Human	Typhimurium	P81 (5)	1
FSL R6-002	US	WA	Bovine	Typhimurium	P86 (1)	3
FSL S5-796	US	NY	Bovine	Typhimurium	P82 (1)	3
FSL S5-635	US	NY	Human	4,5,12:i:-	P80 (1)	1
FSL S9-101	US		Non-dom. bird	4,5,12:i:-	P28 (6)	1
FSL S9-206	Spain		Human	4,5,12:i:-	P28 (6)	1
FSL S9-245	Spain		Human	4,5,12:i:-	P30 (1)	1
FSL S9-241	Spain		Human	4,5,12:i:-	P29 (1)	3
FSL S9-238	Spain		Human	4,5,12:i:-	P31 (1)	1
FSL S9-239	Spain		Human	4,5,12:i:-	P26 (1)	1
FSL S9-244	Spain		Human	4,5,12:i:-	P27 (1)	1
FSL S9-204	Spain		Human	4,5,12:i:-	P32 (2)	1
FSL S9-237	Spain		Human	4,5,12:i:-	P34 (1)	1
FSL S9-232	Spain		Enviromental	Typhimurium	P87 (1)	7
FSL S9-234	Spain		Enviromental	Typhimurium	P83 (1)	7

found one combined XbaI/BlnI PFGE type shared by one 4,5,12:i- and one Typhimurium isolate, both isolated from poultry in the US (38). In general, these observations further support that *Salmonella* 4,5,12:i:- has evolved from a *Salmonella* Typhimurium ancestor.

A total of four PFGE patterns (P12, P19, P35, and P45) were found among both Spanish and US *Salmonella* Typhimurium isolates. P12 represented 14 and 10 *Salmonella* Typhimurium isolates from Spain and US, respectively. P35 represented two Typhimurium isolates from the US and one Typhimurium isolate from Spain. PFGE patterns P19 and P45 each represented one *Salmonella* Typhimurium isolate from the US and one isolate from Spain. Identification of identical XbaI PFGE patterns among *Salmonella* Typhimurium isolates from different continents is consistent with previous studies which have shown that some genetically closely related *Salmonella* strains are distributed worldwide (7, 19), including some other studies that have found *Salmonella* Typhimurium isolates with identical PFGE types in different countries and continents (7). While PFGE patterns for most Spanish 4,5,12:i:- isolates were similar and different from the patterns for US serotype 4,5,12:i:- isolates, one PFGE pattern (PFGE type P28) was shared among five Spanish serotype 4,5,12:i:- isolates and one 4,5,12:i:- isolate from the US (this US isolate was obtained from a free ranging owl in Georgia [38]). As most owls are non-migratory there is no apparent hypothesis as to the source of an infection with a “Spanish clone” 4,5,12:i:- isolate in this animal.

US and Spanish 4,5,12:i:- isolates show different patterns of gene deletion in the regions encoding phase 2 flagella and have different PFGE patterns. A previous genomic microarray study of four multidrug-resistant Spanish *Salmonella* 4,5,12:i:- isolates and *Salmonella* Typhimurium LT2 identified five genomic regions (clusters) that were absent in all four Spanish *Salmonella* 4,5,12:i:- isolates, but

present in *S.* Typhimurium LT2. As our initial PFGE data suggested that Spanish 4,5,12:i:- isolates may be genetically distinct from most US 4,5,12:i:- isolates, we analyzed an available genome sequence for a US 4,5,12:i:- isolate (strain CVM23701 [32]) for presence of these five clusters (i.e., clusters I to V). BLAST searches against the CVM23701 genome sequences showed that cluster I (STM0517-STM0529), which includes 13 genes most of which are involved in allantoin-glyoxylate pathway related functions, is present in the genome of this US 4,5,12:i:- isolate, even though this cluster appears to be absent from the four Spanish *Salmonella* 4,5,12:i:- isolates previously characterized by genomic microarrays (18). Cluster II (STM0893-STM0929), which includes 35 Fels-1 prophage genes and two adjacent genes, was reported to be absent from the four Spanish 4,5,12:i:- isolates and was not identified in the available unfinished genome of the US 4,5,12:i:- isolate. As the genes upstream and downstream of cluster II were located on a single contig, we conclude that this cluster is likely absent from the CVM23701 genome. These findings are consistent with the observation that the Fels1 prophage is present in LT2, but typically absent in other *Salmonella* Typhimurium isolates (25). Cluster III (STM2616-STM2617) encodes the Gifsy-1 prophage and was found in the genome of the US 4,5,12:i:- isolate CVM23701, but was reported to be absent from the four Spanish 4,5,12:i:- isolates previously characterized by genomic microarrays (18). The findings that clusters I to III were all absent from four Spanish 4,5,12:i:- isolates characterized by genomic microarray, while only cluster II was absent from the US 4,5,12:i:- isolate CVM23701, provide initial evidence that US and Spanish 4,5,12:i:- isolates may represent distinct genotypes.

Clusters IV (STM2694-STM2740) and V (STM2758-STM2773), which both were reported to be absent in the four Spanish 4,5,12:i:- isolates (Figure 1a), are located in close proximity to each other. While cluster IV contains 47 Fels-2 prophage

genes, cluster V contains a number of genes associated with different functions, including the *fljAB* operon (18). Notably, deletion of *fljAB* provides a functional explanation for the absence of phase 2 flagellar expression observed in serotype 4,5,12:i:-, as *fljB* encodes the phase 2 flagellar protein and *fljA* encodes a repressor of *fliC* transcription (which encodes the phase 1 flagellar protein). Initial BLAST searches against the genome sequence for the US 4,5,12:i:- isolate CVM23701 showed that both clusters IV and V were absent from the CVM23701genome (cluster IV and V as well as intervening genes were located on a single CVM23701 contig, NZ_ABA001000014.1). Two genes located in the 3' end of cluster V, including STM2772 (*hin*, encodes a recombinase that regulates the regulation of flagellar gene expression) and STM2773 (*iroB*, encodes glucosyl-transferase homolog protein), were present in the CVM23701genome, even though they were reported to be absent from the four Spanish 4,5,12:i:- isolates, based on genomic microarray data (18). Further analysis of the CVM23701 genome sequence indicated that the region between clusters IV and V (STM2739-STM2757) was also absent from the CVM23701 genome, indicating that this strain contains a larger deletion as compared to the four Spanish 4,5,12:i:- isolates; this deletion spans cluster IV and most of cluster V (except for two genes at the 3' end) as well as the region between these two clusters. Interestingly, in the genome sequence of the 4,5,12:i:- isolate CVM23701 an approximately 7 kb region is inserted into this deleted section of the genome. This insertion includes two partial Fels-2 genes (STM2704 and STM2706), and three genes homologous to STM1054, STM1053 and STM1997 (*umuC*), which encode two Gifsy-2 prohage genes and a component of DNA polymerase V (*umuC*) (Figure 4.1b). We will refer to this insertion as the "STM1053-1997" region; this region is not found in LT2. Presence of this region CVM23701 suggests the intriguing hypothesis that

deletion, in the US 4,5,12:i:- clone, of clusters IV and V and the intervening region may have been caused by abortive, imprecise excision of a prophage.

As our analyses detailed above suggest that the US 4,5,12:i: isolate CVM23701 shows distinct genomic gene presence/absence patterns as compared to four Spanish 4,5,12:i:- isolates previously characterized by genomic microarrays (18), we designed PCR primers to determine the absence/presence of 8 genes that are at the junctions of clusters IV and V (i.e., STM2692, STM2694, STM2740, STM2741, STM2757, STM2758, STM2773 (*iroB*) and STM2774; see Fig. 1 for primer locations) and three genes (i.e., *fljA, fljB,*and *hin*)*,* that are responsible for expression of phase-2 flagellar antigen. In addition, a set of primers was designed to allow for the detection of "STM1053-1997" region, which was found in the CVM23701 genome sequence. While negative PCR results may indicate absence of a gene or presence of a distinct allelic variant of a gene, which does not allow for PCR amplification, we surmised that, in this study, negative PCR results in 4,512:i- due to gene diversification (rather than gene absence) are extremely unlikely due to the high genetic similarity between 4,5,12:i- and Typhimurium isolates (e.g., as indicated by identical or highly similar MLST types for these two serotypes). Characterization of an initial 6 isolates (two 4,5,12:i:- isolates from each Spain and the US and one Typhimurium isolate from each Spain and the US) showed that Spanish *S.* 4,5,12:i:- had STM2692, STM2740, STM2741, STM2757and STM2774, but lacked STM2694, STM2758, STM2773 (*iroB*), *fljA, fljB, hin* and the STM1053-1997 region. These results confirmed the genes presence/absence patterns previously reported, based on genomic microarray data, for four Spanish 4,5,12:i:- isolates (18), except for the fact that the PCR primers for STM2740, which was previously reported as absent in Spanish *S.* 4,5,12:i:-, yielded positive results, suggesting presence of at least part of this gene. The PCR results on the two US *Salmonella* 4,5,12:i:- isolates were consistent with the

observations based on our analysis of the CVM23701 genome (representing a US *Salmonella* 4,5,12:i:- isolate). Specifically, the PCR data indicated that the two US *Salmonella* 4,5,12:i:- isolates (i) lacked clusters IV and V as well as the intervening region (as supported by negative PCR results for STM2694, STM2740, STM2741, STM2757, STM2758, *fljA,* and *fljB*), (ii) contain *hin* and *iroB* (which are located in the 3' end of cluster V and absent in the Spanish isolates), and (iii) contain an insertion upstream of *hin* gene (i.e., the STM1053-1997 region), which is absent in the Spanish *Salmonella* 4,5,12:i:- isolates. These data provided further support that Spanish 4,5,12:i:- isolates may be distinct from US 4,5,12:i:- isolates.

To further test the hypothesis that Spanish and US 4,5,12:i:- isolates represents different clonal groups with distinct genome deletion patterns we screened 59 representative *Salmonella* 4,5,12:i:- and Typhimurium isolates from these two countries (representing all PFGE patterns represented among 4,5,12:i:- isolates) for presence/absence of six genes (i.e., STM2740, STM2757, *fljA*, *fljB*, *hin*, and *iroB*) and the STM1053-1997 region (Table 4.2). These PCR targets were selected as they (i) allow for clear differentiation of Typhimurium and 4,5,12:i:- genotypes and (ii) allow for differentiation of the "Spanish" and "US" genomic deletion patterns in the cluster IV and V region of serotype 4,5,12:i:- isolates. The PCR data generated clearly indicated that (i) all Spanish 4,5,12:i:- isolates show a deletion of clusters IV and V, but presence of the intervening region (STM2740 to STM2757), and (ii) all but two US 4,5,12:i:- isolates show a deletion of clusters IV and V, including a deletion of the intervening region between clusters IV and V, as well as presence of *hin* and *iroB* (which are absent in the Spanish 4,5,12:i:- isolates) and presence of the STM1053-1997 region. We thus propose that *Salmonella* 4,5,12:i:- isolates from the US and Spain represent two distinct clones (i.e., the "Spanish" and the "US" clone). These findings are consistent with our observations that XbaI PFGE types of Spanish

4,5,12:i:- isolates generally are clearly distinct from the PFGE patterns for US serotype 4,5,12:i:- isolates. Interestingly, one 4,5,12:i:- isolate from the US (isolated from a free-ranging owl in Georgia) had the same deletion pattern as Spanish 4,5,12:i:- isolates. As this isolate also shared an identical PFGE pattern (P28; see Figure 4.3) with five Spanish 4,5,12:i:- isolates, we also provide initial evidence for intercontinental spread of the "Spanish" 4,5,12:i:- clone.

Interestingly, Matiasovicova et al. (25) suggested that multidrug resistant *Salmonella* Typhimurium might have evolved from a *Salmonella* Typhimurium ancestor that first lost the region including STM0517-0529 (designated as cluster I by Garaizar et al. [18]), allowing the utilization of allantoin as a sole nitrogen source, followed by acquisition of the *Salmonella* genomic island (GI)-1, which includes genes responsible for multi-drug resistance. Since multidrug-resistant isolates Spanish 4,5,12:i:- isolates lack cluster I (18), while the US *Salmonella* 4,5,12:i:- (CVM23701) contains this cluster, one might hypothesize that Spanish 4,5,12:i:- strains might have emerged from MDR *Salmonella* Typhimurium, while US *S.* 4,5,12:i:- might have emerged from non drug resistant *Salmonella* Typhimurium through independent events. Future studies on larger sets of multi-drug resistant and pansusceptible Typhimurium and 4,5,12:i:- isolates from different countries will be needed though to test this hypothesis.

In addition to two common *S.* 4,5,12:i:- clones (i.e., the "Spanish" and the "US" clone), we identified one rare 4,5,12:i:- genotype in North America. In addition to the common "Spanish" and "US" 4,5,12:i:- clones described in detail above, we also identified one rare 4,5,12:i:- genotype in North America. Specifically, a human 4,5,12:i:- isolate from New York State (isolate FSL S5-635; Table4. 4) was found to lack *hin* and the STM1053-1997 region, which are both present in the typical US 4,5,12:i:- isolates, but contained *iroB,* which is typically absent in the Spanish

Table 4. 4. Presence/absence of selected genes in isolates representing the Spanish and US *Salmonella* 4,5,12:i:- clones as well as other *Salmonella* isolates

Genes	Presence of genes in [a]				
	US 4,5,12:i:- Clone (n=30)	Spanish 4,5,12:i:- Clone (n=10)	FSL S5-635 (4,5,12:i:-; n=1)	Isolates with inconsistent serotype results (n=4)[b]	*Salmonella* Typhimurium (n=14)
2740	-	+	+	+	+
2757	-	+	+	+	+
1053-1997	+	-	-	-	-
fljA	-	-	-	+	+
fljB	-	-	-	+	+
hin	+	-	-	+	+
iroB	+	-	+	+	+

[a]Plus (+) and minus (-) signs designate positive and negative PCR results, indicating the presence or absence of a gene.

[b]these isolates were serotyped as 4,5,12:i:- in one replicate and Typhimurium in another replicate (including one isolates that was classified as 4,5,12:i:- in two replicates and Typhimurium in one replicate) and were thus designated as “inconsistent”

clone. This isolate also was positive in the PCR assays targeting STM2741 and 2757, suggesting that this isolate did maintain the genomic region between clusters IV and V, which is present in the “Spanish”, but absent in the “US” 4,5,12:i:- clone. This isolate thus seems to be similar to the Spanish clone, but shows a deletion pattern different from Spanish clones isolates at the 3’ end of cluster V (see Table 4.4). Further characterization of this isolate will be needed to determine whether it represents a third emergence event, independent of both the emergence of the “Spanish” and the “US” clone of 4,5,12:i:- or whether it represents an evolutionary intermediate related to the Spanish 4,5,12:i:- clone. While this isolate represents a unique PFGE pattern not found among any other 4,5,12:i:- or Typhimurium isolates, it was classified as ST 1, the same sequence type that represented the majority of Typhimurium isolates (84/100) as well as the majority of Spanish and US clone 4,5,12:i:- isolates (12/13 and 73/73, respectively). This indicates that this strain is closely related to Typhimurium and most likely also emerged from a *Salmonella* Typhimurium ancestor. Overall, our findings suggest that serotype 4,5,12:i:- represent multiple genotypes, possibly indicating a strong selective pressure for loss of phase 2 flagella expression.

We also identified four isolates from the US (FSL S9-102, FSL S9-165, FSL S9-166 and FSL R6-084) that were initially determined to be serotype 4,5,12:i:-, but were found in the PCR screens to contain *fljA*, *fljB***,** and *hin*, three genes critical for phase-2 flagellar expression (Table 4.4). PCR screens for other genes in clusters IV and V indicated that both of these clusters were present in these four isolates. As Zamperini et al. (2007) suggested that mutations in *fljB* (the gene encoding phase 2 flagella) may also cause a serotype 4,5,12:i:- phenotype, we sequenced the 1521 nt *fljB* ORF in three of these isolates (isolates FSL S9-165 and FSL S9-166 showed the same PFGE type [P23] and thus *fljB* was only sequenced for one of these isolates). All

of these isolates had an identical *fljB* sequence, which showed one synonymous single nucleotide polymorphisms as compared to *Salmonella* Typhimurium LT 2 (the representative sequence for FSL S9-102 was deposited in GenBank with accession no. FJ763347), these isolates did not show any non-synonymous changes or other mutations that would explain a lack of phase 2 flagella expression. These four isolates were thus submitted to NVSL (USDA APHIS VS, Ames, IA) for serotype confirmation. While isolates FSL S9-166 and FSL R6-084 were re-serotyped as Typhimurium, FSL S9-102 was re-serotyped twice, once as 4,5,12:i:- and once as Typhimurium and FSL S9-165 was re-serotyped as Typhimurium twice. These results are consistent with previous reports (26) that serotyping of *Salmonella* may sometimes be difficult to reproduce and suggest that Typhimurium isolates may sometimes be misclassified as 4,5,12:i:- (and vice versa). While the four specific isolates with inconsistent serotype results characterized here appear to represent serotype Typhimurium (based on genetic evidence for presence of intact phase 2 genes and at least on serotype result characterizing them as Typhimurium), it is tempting to speculate that these isolates may show reduced phase 2 flagella expression, which could be responsible for the inconsistent serotype data. This hypothesis would need to be tested further by expression analyses (e.g., quantitative RT-PCR analysis).

CONCLUSIONS

Overall, our observations suggest that *Salmonella* 4,5,12:i:- evolved through multiple independent emergence events, most likely from serotype Typhimurium ancestors. Serotype 4,5,12:i:- isolates from Spain and the US appear to represent two different clones with distinct geographical distributions. This hypothesis is supported by multiple independent pieces of evidence. First, different genome-wide deletion patterns were found in four Spanish *Salmonella* 4,5,12:i:- isolates (as previously

determined by genomic microarrays [18]) and one US *Salmonella* 4,5,12:i:- isolate (based on an available whole genome sequence [32]). In particular, clusters I and III were present in the US 4,5,12:i:- isolate (CVM23701), even though these clusters were reported to be absent in Spanish *S.* 4,5,12:i:- isolates (18). Second, genome analyses and detailed PCR based mapping (of 31 US and 10 Spanish 4,5,12:i:- isolates) showed clearly distinct deletion patterns in the genome region up and downstream of the genes encoding proteins critical for phase 2 flagella and phase variation (i.e., *fljA*, *fljB*, and *hin*) in all Spanish and all but two US 4,5,12:i:- isolates. Specifically, the Spanish isolates showed two deletions (of clusters IV and V), while the majority of US isolates showed a single larger deletion (encompassing both clusters IV and V as well as the intervening region) with a 3' junction different from that observed in the Spanish isolates. These findings provide another example of a *Salmonella* serotype of considerable public health relevance that represents at least two independent genetic lineages. For example, *Salmonella* Newport has previously been shown to represent two distinct genetic lineages, including one lineage that contains predominantly pansusceptible isolates and one that predominantly contains multi-drug resistant isolates (2, 8, 22). In addition, multiple independent emergence of serotype 4,5,12:i:- and subsequent ecological success of multiple lineages (as evidenced by common isolation from human clinical cases in both Spain and the US), suggesting a strong selective pressure for loss of phase 2 flagella or a closely linked genotype. Future efforts to define the possible selection for loss of phase 2 flagella and to understand the specific 4,5,12:i- genotypes circulating in countries other than the US and Spain will be critical for understanding of the ecology and evolution of human disease associated non-typhoidal *Salmonella*.

ACKNOWLEDGEMENTS

We thank Javier Garaizar, University of the Basque Country, Vitoria-Gasteiz, Spain for providing *Salmonella* isolates from Spain and David White, FDA, for providing *Salmonella* food isolates. Support for this project was provided by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract N01-AI-30054 - ZC-006-07 (to Lorin Warnick). Andrea Moreno was supported by National Research Initiative Competitive Grant 2006-55212-17250 from the USDA Cooperative State Research, Education, and Extension Service Epidemiological Approaches to Food Safety program awarded to Dr. Randall Singer.

REFERENCES

1. **Agasan, A., J. Kornblum, G. Williams, C. C. Pratt, P. Fleckenstein, M. Wong, and A. Ramon.** 2002. Profile of *Salmonella enterica* subsp. *enterica* (subspecies I) serotype 4,5,12:I:- strains causing food-borne infections in New York city. J. Clin. Microbiol. **40:**1924-1929**.**
2. **Alcaine, S. D., Y. Soyer, L. D. Warnick, W. L. Su, S. Sukhnanand, J. Richards, E. D. Fortes, P. McDonough, T. P. Root, N. B. Dumas, Y. Grohn, and M. Wiedmann.** 2006. Multilocus sequence typing supports the hypothesis that cow- and human-associated *Salmonella* isolates represent distinct and overlapping populations. Appl. Environ. Microbiol. **72:**7575-7585.
3. **Alcaine, S. D., S. S. Sukhnanand, L. D. Warnick, W. L. Su, P. McGann, P. McDonough, and M. Wiedmann.** 2005. Ceftiofur-resistant *Salmonella* strains isolated from dairy farms represent multiple widely distributed subtypes that evolved by independent horizontal gene transfer. Antimicrob. Agents Chemother. **49:**4061-4067.
4. **Aldridge, P. D., C. Wu, J. Gnerer, J. E. Karlinsey, K. T. Hughes, and M. S. Sachs.** 2006. Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. Proc. Natl. Acad. Sci. U.S.A **103:**11340-11345.
5. **Alvarez, J., M. Sota, A. B. Vivanco, I. Perales, R. Cisterna, A. Rementeria, and J. Garaizar.** 2004. Development of a multiplex PCR technique for detection and epidemiological typing of *Salmonella* in human clinical samples. J. Clin. Microbiol. **42:**1734-1738.
6. **Amavisit, P., W. Boonyawiwat, and A. Bangtrakulnont.** 2005. Characterization of *Salmonella enterica* serovar Typhimurium and monophasic

Salmonella serovar 1,4,[5],12:I:- isolates in Thailand. J. Clin. Microbiol. **43:**2736-2740.

7. **Baggesen, D. L., D. Sandvang, and F. M. Aarestrup.** 2000. Characterization of *Salmonella enterica* serovar Typhimurium DT104 isolated from Denmark and comparison with isolates from Europe and the United States. J. Clin. Microbiol. **38:**1581-1586**.**

8. **Berge, A. C., J. M. Adaska, and W. M. Sischo.** 2004. Use of antibiotic susceptibility patterns and pulsed-field gel electrophoresis to compare historic and contemporary isolates of multi-drug-resistant *Salmonella enterica* subsp. *Enterica* serovar Newport. Appl. Environ. Microbiol. **70:**318-323.

9. **Boonmar, S., A. Bangtrakulnonth, S. Pornruangwong, S. Samosornsuk, K. Kaneko, and M. Ogawa.** 1998. Significant increase in antibiotic resistance of *Salmonella* isolates from human beings and chicken meat in Thailand. Veterinary Microbiology **62:**73-80.

10. **CDC.** 2008. *Salmonella* surveillance: Annual summary, 2006. Atlanta, Georgia: US Department of Health and Human Services, CDC.

11. **Chevance, F. F., and K. T. Hughes.** 2008. Coordinating assembly of a bacterial macromolecular machine. Nat. Rev. Microbiol.**6:**455-65.

12. **Cooke, F. J., D. J. Brown, M. Fookes, D. Pickard, A. Ivens, J. Wain, M. Roberts, R. A. Kingsley, N. R. Thomson, and G. Dougan.** 2008. Characterization of the genomes of a diverse collection of *Salmonella enterica* serovar Typhimurium definitive phage type 104. J. Bacteriol. **190:**8155-8162.

13. **de la Torre. E., D. Zapata, M. Tello, W. Mejia, N. Frias, F. J. Garcia Pena, E. M. Mateu, and E. Torre.** 2003. Several *Salmonella enterica* subsp. *enterica* serotype 4,5,12:I:- phage types isolated from swine samples originate from serotype Typhimurium DT U302. J. Clin. Microbiol. **41:**2395-2400.

14. **Delgado Ronda, N., J. L. Muñoz Bellido, M. I. García García, R. Ibáñez Pérez, S. Muñoz Criado, R. Serrano Heranz, M. C. Sáenz González, J. A. García Rodríguez.** 2006. Molecular epidemiology of drug-resistant Salmonella Typhimurium in Spain Rev. Esp. Quimioterap. **19:**152-160.

15. **Echeita, M. A., A. Aladuena, S. Cruchaga, and M. A. Usera.** 1999**.** Emergence and spread of an atypical *Salmonella enterica* subsp *enterica* serotype 4,5,12 : I :- strain in Spain. J. Clin. Microbiol. **37:**3425-3425.

16. **Echeita, M. A., S. Herrera, and M. A. Usera.** 2001. Atypical, *fljB*-negative *Salmonella enterica* subsp. e*nterica* strain of serovar 4,5,12:I:- appears to be a monophasic variant of serovar Typhimurium. J. Clin. Microbiol. **39:**2981-2983.

17. **Fakhr, M. K., L. K. Nolan, and C. M. Logue.** 2005. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. J. Clin. Microbiol. **43:**2215-2219.

18. **Garaizar, J., S. Porwollik, A. Echeita, A. Rementeria, S. Herrera, R. M. Y. Wong, J. Frye, M. A. Usera, and M. McClelland.** 2002. DNA microarray-based typing of an atypical monophasic *Salmonella enterica* serovar. J. Clin. Microbiol. **40:**2074-2078.

19. **Gatto, A. J., T. M. Peters, J. Green, I. S. Fisher, O. N. Gill, J. S. O'Brien, C. Maguire, C. Berghold, I. Lederer, P. Gerner-Smidt, M. Torpdahl, A. Siitonen, S. Lukinmaa, H. Tschape, R. Prager, I. Luzzi, A. M. Dionisi, W. K. van der Zwaluw, M. Heck, J. Coia, D. Brown, M. Usera, A. Echeita, and E. J. Threlfall.** 2006. Distribution of molecular subtypes within *Salmonella enterica* serotype Enteritidis phage type 4 and *S.* Typhimurium

definitive phage type 104 in nine European countries, 2000-2004: results of an international multi-centre study. Epidemiol. Infect. **134:**729-736.

20. **Grimont, P., and F. Weil.** 2007. Antigenic Formulae of the *Salmonella* Servovars, 9th ed. Pasteur Institute, Paris: World Health Organization Centre for Reference and Research on *Salmonella*.

21. **Guerra, B., I. Laconcha, S. M. Soto, M. A. Gonzalez-Hevia, and M. C. Mendoza.** 2000. Molecular characterisation of emergent multiresistant *Salmonella enterica* serotype [4,5,12:I:-] organisms causing human salmonellosis. FEMS Microbiol. Lett. **190:**341-347.

22. **Harbottle, H., D. G. White, P. F. McDermott, R. D. Walker, and S. Zhao.** 2006. Comparison of multilocus sequence typing, pulsed-field gel electrophoresis, and antimicrobial susceptibility typing for characterization of *Salmonella enterica* serotype Newport isolates. J. Clin. Microbiol. **44:**2449-2457.

23. **Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: An application of Simpson's index of diversity. J. Clin. Microbiol. **26:**2465-2466.

24. **Maiden, M. C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U.S.A. **95:**3140-3145.

25. **Matiasovicova, J., P. Adams, P. A. Barrow, H. Hradecka, M. Malcova, R. Karpiskova, E. Budinska, L. Pilousova, and I. Rychlik.** 2007. Identification of putative ancestors of the multidrug-resistant *Salmonella enterica* serovar

Typhimurium DT104 clone harboring the *Salmonella* genomic island 1. Arch. Microbiol. **187:**415-24.

26. **McQuiston, J. R., R. Parrenas, M. Ortiz-Rivera, L. Gheesling, F. Brenner, and P. I. Fields.** 2004. Sequencing and comparative analysis of flagellin genes *fliC, fljB*, and *flpA* from *Salmonella*. J. Clin. Microbiol. **42:**1923-1932.

27. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg. Infec. Dis. **5:**607-625.

28. **Moreno Switt, A. I., Y. Soyer, L. D. Warnick, and M. Wiedmann.** 2009. Emergence, distribution and molecular and phenotypic characteristics of *Salmonella enterica* serotype 4,5,12:I:-. Foodborne Pathog. Dis. in press.

29. **Mossong, J., P. Marques, C. Ragimbeau, P. Huberty-Krau, S. Losch, G. Meyer, G. Moris, C. Strottner, W. Rabsch, and F. Schneider.** 2007. Outbreaks of monophasic *Salmonella enterica* serovar 4,[5],12:I:- in Luxembourg, 2006. Euro. Surveill. **12:**E11-E12.

30. **Posada, D., and K. A. Crandall.** 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics. **14:**817-818.

31. **Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan, and T. J. Barrett.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. Foodborne Pathog. Dis. **3:**59-67.

32. **Rosovitz, M. J., McDermott, P., White, D., LeClerc, J. E., Mammel, M. K., Cebula, T. A. and Ravel, J.** Annotation of *Salmonella enterica* subsp. *enterica* serovar 4,[5],12:I:- str. CVM23701. The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

http://msc.jcvi.org/Salmonella/Salmonella_enterica_subsp__enterica_serovar_4__5__12_i___str__cvm23701/index.shtml.

33. **Sukhnanand, S., S. Alcaine, L. D. Warnick, W. L. Su, J. Hof, M. P. Craver, P. McDonough, K. J. Boor, and M. Wiedmann.** 2005. DNA sequence-based subtyping and evolutionary analysis of selected *Salmonella enterica* serotypes. J. Clin. Microbiol. **43:**3688-3698.

34. **Swaminathan, B., P. Gerner-Smidt, and T. Barrett.** 2006. Focus on *Salmonella*. Foodborne Pathog. Dis. **3:**154-156.

35. **Swofford, D. L.** 2002. Paup*: Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

36. **Tavechio, A. T., A. C. Ghilardi, and S. A. Fernandes.** 2004. "Multiplex PCR" Identification of the atypical and monophasic *Salmonella enterica* subsp. *enterica* serotype 1,4,[5],12:I:- in Sao Paulo State, Brazil: Frequency and antibiotic resistance patterns. Rev. Inst. Med. Trop. Sao Paulo. **46:**115-117.

37. **Yamamoto, S., and K. Kutsukake.** 2006. *fljA*-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica* serovar Typhimurium. J. Bacteriol. **188:**958-967**.**

38. **Zamperini, K., V. Soni, D. Waltman, S. Sanchez, E. C. Theriault, J. Bray, and J. J. Maurer.** 2007. Molecular characterization reveals *Salmonella enterica* serovar 4,[5],12:I:- from poultry is a variant Typhimurium serovar. Avian. Dis. **51:**958-64

CHAPTER 5

CONCLUSIONS

Salmonella serotypes can cause a wide range of diseases, from self-limiting gastroenteritis to severe systemic infections. *Salmonella* nontyphoid serotypes are the major cause of known foodborne diseases worldwide. In the US alone, the annual economic burden due to *Salmonella* was estimated at $2.8 billion, which does not include any economic loss due to recalls in the food industry. Since *Salmonella* remains to be a public health concern and causes huge economic loss in the food industry, it is important to determine the transmission, ecology and evolution of *Salmonella* in different host species from different geographical regions to assure public health worldwide.

The study reported in chapter 2 was conducted to provide a better understanding of the genetic relationship and epidemiology of human and cattle associated *Salmonella*. In this study, we used a total of 335 human and bovine *Salmonella* clinical isolates, collected from New York and a neighboring state (Vermont), and we characterized them by serotyping and pulsed field gel electrophoresis (PFGE), as well as multilocus sequence typing (MLST). My work showed that:(i) PFGE provides higher discrimination for human and bovine *Salmonella* isolates than serotyping and MLST; (ii) PFGE can be used to differentiate host specific and unrestricted *Salmonella* subtypes, as well as widely distributed *Salmonella* subtypes, and (iii) persistent *Salmonella* isolates may cause infection more than once on the same farm. Overall, the combination of these subtyping methods provides a better understanding of ecology and transmission of *Salmonella*.

Using different subtyping methods, we built a database which will be used to determine host specific, unrestricted, geographically clustered and spatially persistent *Salmonella* subtypes. The data we generated in this study, as well as isolate

information (e.g. isolation year, source, antibiotic resistance, county) are publicly available in Pathogen Tracker (www.pathogentracker.net) and can be used by other researchers for their academic purposes and for their epidemiological investigations. We also found that the larger database, which includes *Salmonella* isolates from various sources from different regions of the US, as well as different countries and continents, in combination with more sensitive molecular biology techniques, will enhance our abilities to detect outbreaks and link the outbreak to a specific food faster. This improves our ability to secure a safe food supply, thus safeguards the public health.

Further work is needed to assess which mechanisms are responsible for *Salmonella's* ability to survive in different host species as well as the mechanisms that determines its the pathogenic potential in *Salmonella* host adapted and unrestricted subtypes determined in this study. The data provided in Chapter 2 can be used in future studies to further determine the distribution of subtypes of *Salmonella* nontyphoidal serotypes from different geographical regions world wide, as well as from different years. The data can also be used to detect which *Salmonella* subtypes among host adapted or unrestricted subtypes are more likely to gain multi drug resistance.

Results from chapter 3 clearly indicate that positive selection and homologous recombination among core genome genes play an important role in the evolution of bacterial pathogens. These are, in addition, to the well established importance of gene acquisition and deletion in conjunction with previous genome wide studies on positive selection in other pathogens. Our analyses clearly show that both homologous recombination and positive selection (particularly lineage-specific positive selection) contribute critically to the evolution of the *Salmonella* core genome. Genes with evidence of positive selection identified here may provide promising targets for future

mutational studies aimed at further identifying mechanisms that contribute to *Salmonella* diversification, including its adaptation to specific host species. The relevance of the lineage-specific positive selection patterns identified is supported by the convergence of the positive selection patterns identified in the *Salmonella* Typhi lineage (i.e., for genes encoding proteins involved in purine, pyrimidine, nucleotide and nucleotide biosynthesis) and experimental evidence that genes involved in *de novo* nucleotide biosynthesis are essential for bacterial growth in blood.

Further studies are needed with additional *Salmonella* genomes, with different serotypes, source and multidrug resistant patterns to further improve our understanding of bacterial pathogens. In addition, further studies are needed to assess the roles on virulence of the genes found to have evidence of under positive selection and recombination in my study in different host species.

The study described in chapter 4 was conducted to better understand the evolution and ecology of *Salmonella* serotype 4,5,12:i:-, a newly emerged serotype, in the US and Spain using a variety of molecular methods. Our data indicate that (i) *Salmonella* 4,5,12:i:- and Typhimurium represent a highly clonal group, which can be differentiated by PFGE, (ii) US and Spanish 4,5,12:i:- isolates show different patterns of gene deletion in the regions encoding phase 2 flagella and represent distinct PFGE patterns, and (iii) in addition to two common 4,5,12:i:- genotypes (designated in this study as the "Spanish" and the "US" 4,5,12:i:- clone), other 4,5,12:i:- genotypes exist. The novel finding in Chapter 4 showed that *Salmonella* 4,5,12:i:- most likely represents multiple clones that emerged through independent deletion events.

These findings provide another example of a *Salmonella* serotype of considerable public health relevance that represents at least two independent genetic lineages. Multiple independent emergence of serotype 4,5,12:i:- and the subsequent ecological success of multiple lineages (as evidenced by common isolation from

human clinical cases in both Spain and the US), suggest a strong selective pressure for loss of phase 2 flagella on *Salmonella* Typhimurium or a closely linked genotype. Future efforts to define the possible selection for loss of phase 2 flagella and to understand the specific 4,5,12:i- genotypes circulating in countries other than the US and Spain will be critical for understanding the ecology and evolution of human disease associated non-typhoidal *Salmonella.*

In conclusion, my Ph.D. work provides important contributions to our understanding of the transmission and evolution of *Salmonella.* My work provides additional information on the *Salmonella* core genome that is important for understanding the evolution of *Salmonella* to help the food industry to detect and differentiation of *Salmonella* species (e.g. developing rapid molecular techniques to detect *Salmonella*). Further, this study provides data that may contribute to identification of novel targets for drugs for salmonellosis and/or typhoid fever or may aid in vaccine development to prevent typhoid fever in developing countries.

APPENDIX ONE

LIST OF SUPPLEMENTAL TABLES

Table A1 [S2.1]. A total of 335 spatially and temporally matched *Salmonella enterica* spp. e*nterica* nontyphoidal isolates and their characteristics

FSL No.[a]	County	State	Farm No.	Serotype	Source	ST	PFGE Type No.	PFGE Pattern Designation
FSL S5-546	Wyoming	NY	259	Muenster	Cattle	44	2	NYCU.TDSX01.0004
FSL S5-425	Wyoming	NY	259	Muenster	Cattle	44	4	NYCU.TDSX01.0005
FSL S5-387	Franklin	NY	909	Muenster	Cattle	44	5	NYCU.TDSX01.0002
FSL S5-432	Wyoming	NY	907	Muenster	Cattle	44	5	NYCU.TDSX01.0002
FSL S5-868	Wyoming	NY	259	Muenster	Cattle	44	6	NYCU.TDSX01.0006
FSL S5-789	Orleans	NY	853	Muenster	Cattle	44	7	NYCU.TDSX01.0003
FSL S5-917	Livingston	NY	329	Muenster	Cattle	44	7	NYCU.TDSX01.0003
FSL S5-798	Livingston	NY	329	Muenster	Cattle	44	7	NYCU.TDSX01.0003
FSL S5-838	Livingston	NY	329	Muenster	Cattle	44	7	NYCU.TDSX01.0003
FSL S5-547	Lewis	NY	54	Agona	Cattle	1	13	NYCU.JABX01.0006
FSL S5-549	Rensselaer	NY	120	Havana	Cattle	69	36	NYCU.TDLX01.0001
FSL S5-427	Oneida	NY	61	Oranienburg	Cattle	53	41	NYCU.JJXX01.0002
FSL S5-551	Washington	NY	186	Adelaide	Cattle	75	44	NYCU.TDAX01.0002
FSL S5-429	Cortland	NY	911	Typhimurium	Cattle	6	60	NYCU.JPXX01.0024
FSL S5-845	Clinton	NY	764	Typhimurium	Cattle	6	60	NYCU.JPXX01.0024
FSL S5-933	Clinton	NY	243	Typhimurium	Cattle	6	60	NYCU.JPXX01.0024
FSL S5-784	Clinton	NY	764	Typhimurium	Cattle	6	60	NYCU.JPXX01.0024
FSL S5-788	Niagara	NY	904	Typhimurium	Cattle	49	63	NYCU.JPXX01.0027
FSL S5-607	Rensselaer	NY	584	Typhimurium	Cattle	6	64	NYCU.JPXX01.0031
FSL S5-555	Tompkins	NY	100	Typhimurium	Cattle	6	65	NYCU.JPXX01.0005
FSL S5-564	St. Lawrence	NY	208	Typhimurium	Cattle	6	66	NYCU.JPXX01.0006
FSL S5-556	St. Lawrence	NY	208	Typhimurium	Cattle	6	66	NYCU.JPXX01.0006
FSL S5-620	Wyoming	NY	415	Typhimurium	Cattle	6	70	NYCU.JPXX01.0010
FSL S5-747	Wyoming	NY	691	Typhimurium	Cattle	6	70	NYCU.JPXX01.0010
FSL S5-831	Wyoming	NY	837	Typhimurium	Cattle	6	70	NYCU.JPXX01.0010
FSL S5-805	Orleans	NY	826	Typhimurium	Cattle	6	72	NYCU.JPXX01.0012
FSL S5-799	Genesee	NY	821	Typhimurium	Cattle	6	73	NYCU.JPXX01.0013
FSL S5-936	Rensselaer	NY	125	Typhimurium	Cattle	6	79	NYCU.JPXX01.0019
FSL S5-931	Rensselaer	NY	125	Typhimurium	Cattle	6	79	NYCU.JPXX01.0019
FSL S5-729	Washington	NY	804	Typhimurium	Cattle	6	84	NYCU.JPXX01.0028
FSL S5-896	Rensselaer	NY	318	Typhimurium	Cattle	6	85	NYCU.JPXX01.0029
FSL S5-737	Wyoming	NY	261	4,5,12:i:-	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-433	Ontario	NY	820	Typhimurium	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-558	Wyoming	NY	261	4,5,12:i:-	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-597	Wyoming	NY	261	4,5,12:i:-	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-615	Wyoming	NY	261	4,5,12:i:-	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-622	Wyoming	NY	261	4,5,12:i:-	Cattle	6	89	NYCU.JPXX01.0033
FSL S5-580	Wyoming	NY	261	4,5,12:i:-	Cattle	6	90	NYCU.JPXX01.0035
FSL S5-596	Clinton	NY	524	4,5,12:i:-	Cattle	6	90	NYCU.JPXX01.0035
FSL S5-618	Wyoming	NY	261	4,5,12:i:-	Cattle	6	91	NYCU.JPXX01.0034
FSL S5-816	Niagara	NY	510	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-891	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-744	Wyoming	NY	261	Typhimurium	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-777	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-782	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-785	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-820	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-829	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-841	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001

Table A1 [S2.1]. (Continued)

FSL S5-865	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-876	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-924	Wyoming	NY	261	4,5,12:i:-	Cattle	6	94	NYCU.JPXX01.0001
FSL S5-759	Wyoming	NY	261	4,5,12:i:-	Cattle	6	95	NYCU.JPXX01.0038
FSL S5-431	Erie	NY	901	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-889	Wyoming	NY	261	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-822	Wyoming	NY	261	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-840	Wyoming	NY	261	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-864	Wyoming	NY	261	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-873	Wyoming	NY	261	Kentucky	Cattle	17	96	NYCU.JGPX01.0001
FSL S5-947	Lewis	NY	11	Mbandaka	Cattle	65	100	NYCU.TDRX01.0004
FSL S5-796	Chenango	NY	105	Typhimurium	Cattle	8	102	NYCU.JPXX01.0002
FSL S5-554	Cayuga	NY	97	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-786	Washington	NY	225	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-800	Washington	NY	186	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-916	Washington	NY	186	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-550	Cayuga	NY	97	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-797	Washington	NY	186	Typhimurium	Cattle	8	104	NYCU.JPXX01.0004
FSL S5-428	Oswego	NY	906	Rough o:i: 1,2	Cattle	8	105	NYCU.YABX01.0001
FSL S5-430	Oswego	NY	906	Typhimurium	Cattle	8	106	NYCU.JPXX01.0041
FSL S5-839	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-553	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-560	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-571	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-575	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-582	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-584	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-593	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-617	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-725	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-743	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-750	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-760	Saratoga	NY	223	Infantis	Cattle	60	107	NYCU.JFXX01.0001
FSL S5-566	Saratoga	NY	223	Infantis	Cattle	60	108	NYCU.JFXX01.0002
FSL S5-590	Saratoga	NY	223	Infantis	Cattle	60	109	NYCU.JFXX01.0003
FSL S5-734	Cayuga	NY	163	Infantis	Cattle	60	114	NYCU.JFXX01.0009
FSL S5-559	Wyoming	NY	415	Montevideo	Cattle	9	119	NYCU.JIXX01.0005
FSL S5-757	Livingston	NY	329	Montevideo	Cattle	9	119	NYCU.JIXX01.0005
FSL S5-630	Washington	NY	521	Montevideo	Cattle	9	120	NYCU.JIXX01.0006
FSL S5-836	Genesee	NY	838	Bardo	Cattle	11	121	NYCU.TEGX01.0002
FSL S5-436	Oneida	NY	902	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-544	Oneida	NY	228	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-545	Wyoming	NY	260	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-611	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-612	Erie	NY	522	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-621	Wyoming	NY	791	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-739	Niagara	NY	827	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-776	Niagara	NY	637	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-842	Chenango	NY	105	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-629	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-730	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-740	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-746	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-752	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-758	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-762	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-780	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-808	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-833	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001

Table A1 [S2.1]. (Continued)

FSL S5-846	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-879	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-882	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-887	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-903	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-910	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-914	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-920	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-938	Niagara	NY	510	Newport	Cattle	11	121	NYCU.JJPX01.0001
FSL S5-850*	Niagara	NY	510	Newport	Cattle	11	122	NYCU.JJPX01.0022
FSL S5-570	Franklin, VT	VT	488	Bardo	Cattle	11	126	NYCU.TEGX01.0001
FSL S5-420	Seneca	NY	903	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-424	St. Lawrence	NY	204	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-557	Franklin, VT	VT	303	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-578	Franklin, VT	VT	320	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-601	Franklin, VT	VT	218	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-605	Clinton	NY	524	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-610	Cayuga	NY	163	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-619	Onondaga	NY	152	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-624	Franklin, VT	VT	488	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-626	Chittenden, VT	VT	490	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-548	Franklin, VT	VT	303	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-561	Franklin, VT	VT	488	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-567	Franklin, VT	VT	488	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-577	Franklin, VT	VT	320	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-623	Onondaga	NY	152	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-631	Onondaga	NY	152	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-722	Cayuga	NY	163	Newport	Cattle	11	126	NYCU.JJPX01.0011
FSL S5-732	Onondaga	NY	152	Newport	Cattle	11	126	NYCU.JJPX01.0003
FSL S5-552	Chittenden, VT	VT	489	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-562	Lamoille, VT	VT	359	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-594	Clinton	NY	524	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-628	Chittenden, VT	VT	490	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-625	Clinton	NY	524	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-627	Chittenden, VT	VT	490	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-766	Clinton	NY	524	Newport	Cattle	11	127	NYCU.JJPX01.0004
FSL S5-591	Chittenden, VT	VT	490	Newport	Cattle	11	129	NYCU.JJPX01.0007
FSL S5-602	Franklin, VT	VT	300	Newport	Cattle	11	132	NYCU.JJPX01.0011
FSL S5-715	Lewis	NY	438	Newport	Cattle	11	132	NYCU.JJPX01.0011
FSL S5-419	Cortland	NY	910	Newport	Cattle	11	150	NYCU.JJPX01.0013
FSL S5-790	Cattaraugus	NY	680	4, 12:i:-	Cattle	6	153	NYCU.JPXX01.0039
FSL S5-761	Niagara	NY	679	Thompson	Cattle	62	157	NYCU.JP6X01.0001

Table A1 [S2.1]. (Continued)

FSL S5-765	Livingston	NY	329	Thompson	Cattle	62	157	NYCU.JP6X01.0001
FSL S5-386	Livingston	NY	905	Thompson	Cattle	43	159	NYCU.JP6X01.0003
FSL S5-437	Tompkins	NY	908	Agona	Cattle	2	164	NYCU.JABX01.0004
FSL S5-748	Rensselaer	NY	584	Agona	Cattle	2	165	NYCU.JABX01.0005
FSL S5-867	Allegany	NY	599	Agona	Cattle	2	166	NYCU.JABX01.0007
FSL S5-872	Rensselaer	NY	584	Agona	Cattle	2	166	NYCU.JABX01.0007
FSL S5-565	Clinton	NY	308	Newport	Cattle	11	168	NYCU.JJPX01.0005
FSL S5-660	Clinton	NY		Urbana	Human	52	1	NYCU.JQGX01.0003
FSL S5-485	Erie	NY		Saintpaul	Human	38	3	NYCU.JN6X01.0004
FSL S5-401	Chemung	NY		Muenster	Human	44	8	NYCU.TDSX01.0001
FSL S5-481	Erie	NY		Pomona	Human	29	9	NYCU.POMX01.0002
FSL S5-487	Unknown	NY		Give	Human	30	10	NYCU.JEXX01.0001
FSL S5-661	Schenectady	NY		Urbana	Human	52	11	NYCU.JQGX01.0002
FSL S5-512	Monroe	NY		Saintpaul	Human	81	12	NYCU.JN6X01.0003
FSL S5-456	Orleans	NY		Schwarzen-grund	Human	4	14	NYCU.JM6X01.0001
FSL S5-458	Orleans	NY		Schwarzen-grund	Human	4	14	NYCU.JM6X01.0001
FSL S5-379	Suffolk	NY		Panama	Human	37	15	NYCU.JKGX01.0001
FSL S5-454	Chenango	NY		Panama	Human	37	16	NYCU.JKGX01.0002
FSL S5-652	Erie	NY		Javiana	Human	21	17	NYCU.JGGX01.0004
FSL S5-406	Bronx	NY		Javiana	Human	21	18	NYCU.JGGX01.0003
FSL S5-500	Nassau	NY		Panama	Human	19	19	NYCU.JKGX01.0003
FSL S5-395	Richmond	NY		Javiana	Human	19	20	NYCU.JGGX01.0001
FSL S5-665	Nassau	NY		Javiana	Human	19	21	NYCU.JGGX01.0002
FSL S5-453	Nassau	NY		Arechavaleta	Human	37	22	NYCU.AREX01.0001
FSL S5-502	Nassau	NY		Berta	Human	82	23	NYCU.JAXX01.0001
FSL S5-474	Suffolk	NY		Montevideo	Human	56	24	NYCU.JIXX01.0003
FSL S5-478	Suffolk	NY		Montevideo	Human	56	24	NYCU.JIXX01.0003
FSL S5-382	Chautauqua	NY		Montevideo	Human	57	25	NYCU.JIXX01.0004
FSL S5-414	Erie	NY		Enteritidis	Human	14	26	NYCU.JEGX01.0001
FSL S5-415	Broome	NY		Enteritidis	Human	14	26	NYCU.JEGX01.0001
FSL S5-443	Dutchess	NY		Enteritidis	Human	14	26	NYCU.JEGX01.0001
FSL S5-445	Suffolk	NY		Enteritidis	Human	14	26	NYCU.JEGX01.0001
FSL S5-371	Erie	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-376	Nassau	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-377	Nassau	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-402	Unknown	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-486	Onondaga	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-514	Suffolk	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-522	Westchester	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-528	Erie	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-641	Westchester	NY		Enteritidis	Human	14	27	NYCU.JEGX01.0002
FSL S5-460	Monroe	NY		Enteritidis	Human	14	28	NYCU.JEGX01.0003
FSL S5-492	Orange	NY		Typhimurium	Human	6	29	NYCU.JPXX01.0043
FSL S5-467	Kings	NY		Enteritidis	Human	14	30	NYCU.JEGX01.0004
FSL S5-539	Onondaga	NY		Enteritidis	Human	14	30	NYCU.JEGX01.0004
FSL S5-645	Orange	NY		Enteritidis	Human	14	30	NYCU.JEGX01.0004
FSL S5-538	Nassau	NY		Enteritidis	Human	14	31	NYCU.JEGX01.0005
FSL S5-416	Albany	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-444	Tompkins	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-459	Erie	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-461	Delaware	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-496	Monroe	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-508	Nassau	NY		Enteritidis	Human	36	32	NYCU.JEGX01.0006
FSL S5-483	Westchester	NY		Enteritidis	Human	36	33	NYCU.JEGX01.0007
FSL S5-497	Monroe	NY		Enteritidis	Human	14	34	NYCU.JEGX01.0008
FSL S5-407	Orange	NY		Dublin	Human	23	35	NYCU.JDXX01.0001
FSL S5-439	Suffolk	NY		Dublin	Human	23	35	NYCU.JDXX01.0001

Table A1 [S2.1]. (Continued)

FSL S5-662	Allegany	NY	1,7:-:1,5	Human	58	37	NYCU.SC1X01.0003
FSL S5-373	Westchester	NY	Braenderup	Human	61	38	NYCU.JBPX01.0001
FSL S5-374	Monroe	NY	4,5,12:i:-	Human	40	39	NYCU.JPXX01.0040
FSL S5-642	Nassau	NY	Oranienburg	Human	53	40	NYCU.JJXX01.0001
FSL S5-404	Monroe	NY	Paratyphi C	Human	80	42	NYCU.YESX01.0001
FSL S5-646	Albany	NY	Adelaide	Human	75	43	NYCU.TDAX01.0001
FSL S5-499	Nassau	NY	Paratyphi B	Human	48	45	NYCU.JKXX01.0001
FSL S5-637	Oneida	NY	Berta	Human	83	46	NYCU.JAXX01.0002
FSL S5-388	Schenectady	NY	Urbana	Human	52	47	NYCU.JQGX01.0001
FSL S5-410	Schenectady	NY	Urbana	Human	52	47	NYCU.JQGX01.0001
FSL S5-648	Kings	NY	Blockley	Human	27	48	NYCU.JBGX01.0001
FSL S5-417	Albany	NY	Agbeni	Human	63	49	NYCU.JRFX01.0001
FSL S5-465	Nassau	NY	Poona	Human	28	50	NYCU.JL6X01.0002
FSL S5-442	Onondaga	NY	Poona	Human	55	51	NYCU.JL6X01.0001
FSL S5-517	Suffolk	NY	Agona	Human	1	52	NYCU.JABX01.0001
FSL S5-667	Suffolk	NY	Agona	Human	1	53	NYCU.JABX01.0002
FSL S5-647	Bronx	NY	Agona	Human	1	54	NYCU.JABX01.0003
FSL S5-480	New York	NY	Heidelberg	Human	3	55	NYCU.JF6X01.0001
FSL S5-491	Suffolk	NY	Heidelberg	Human	26	56	NYCU.JF6X01.0002
FSL S5-383	Dutchess	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-440	Franklin	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-448	Erie	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-455	Monroe	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-475	Nassau	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-495	Suffolk	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-655	Oneida	NY	Heidelberg	Human	3	57	NYCU.JF6X01.0003
FSL S5-466	Nassau	NY	4,12:r:-	Human	3	58	NYCU.YSMX01.0001
FSL S5-482	Suffolk	NY	4,12:r:-	Human	3	59	NYCU.YSMX01.0002
FSL S5-370	Chautauqua	NY	Typhimurium	Human	6	60	NYCU.JPXX01.0024
FSL S5-375	Kings	NY	Typhimurium	Human	6	60	NYCU.JPXX01.0024
FSL S5-488	Westchester	NY	Typhimurium	Human	6	60	NYCU.JPXX01.0024
FSL S5-509	Washington	NY	Typhimurium	Human	6	60	NYCU.JPXX01.0024
FSL S5-535	Suffolk	NY	Typhimurium	Human	6	60	NYCU.JPXX01.0024
FSL S5-534	Albany	NY	Typhimurium	Human	6	61	NYCU.JPXX01.0025
FSL S5-653	Suffolk	NY	Typhimurium	Human	6	62	NYCU.JPXX01.0026
FSL S5-531	Steuben	NY	Typhimurium	Human	7	66	NYCU.JPXX01.0006
FSL S5-505	Steuben	NY	Typhimurium	Human	6	67	NYCU.JPXX01.0007
FSL S5-532	Cortland	NY	Typhimurium	Human	6	67	NYCU.JPXX01.0007
FSL S5-473	Suffolk	NY	Typhimurium	Human	6	68	NYCU.JPXX01.0008
FSL S5-511	Chemung	NY	Typhimurium	Human	6	68	NYCU.JPXX01.0008
FSL S5-520	Steuben		Typhimurium	Human	6	68	NYCU.JPXX01.0008
FSL S5-397	St. Lawrence	NY	Typhimurium	Human	6	69	NYCU.JPXX01.0009
FSL S5-381	Washington	NY	Typhimurium	Human	6	72	NYCU.JPXX01.0012
FSL S5-394	Dutchess	NY	Typhimurium	Human	6	74	NYCU.JPXX01.0014
FSL S5-494	Bronx	NY	Typhimurium	Human	6	75	NYCU.JPXX01.0015
FSL S5-452	Nassau	NY	Typhimurium	Human	6	76	NYCU.JPXX01.0016
FSL S5-633	Onondaga	NY	Typhimurium	Human	6	76	NYCU.JPXX01.0016
FSL S5-640	Franklin	NY	Typhimurium	Human	6	76	NYCU.JPXX01.0016
FSL S5-663	Tompkins	NY	Typhimurium	Human	6	76	NYCU.JPXX01.0016
FSL S5-392	Westchester	NY	Typhimurium	Human	6	77	NYCU.JPXX01.0017
FSL S5-501	Nassau	NY	Typhimurium	Human	6	78	NYCU.JPXX01.0018
FSL S5-462	Suffolk	NY	Typhimurium	Human	6	80	NYCU.JPXX01.0020
FSL S5-493	Suffolk	NY	Typhimurium	Human	47	81	NYCU.JPXX01.0021
FSL S5-507	Nassau	NY	Typhimurium	Human	6	82	NYCU.JPXX01.0022
FSL S5-536	Jefferson	NY	Typhimurium	Human	49	83	NYCU.JPXX01.0023
FSL S5-369	Monroe	NY	Saintpaul	Human	38	86	NYCU.JN6X01.0001
FSL S5-405	Monroe	NY	Saintpaul	Human	38	86	NYCU.JN6X01.0001
FSL S5-649	Bronx	NY	Saintpaul	Human	38	87	NYCU.JN6X01.0002

Table A1 [S2.1]. (Continued)

FSL S5-516	Otsego	NY	Typhimurium	Human	6	88	NYCU.JPXX01.0030
FSL S5-390	Monroe	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-398	Westchester	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-409	Erie	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-498	Nassau	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-656	Westchester	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-666	Warren	NY	4,5,12:i:-	Human	6	89	NYCU.JPXX01.0033
FSL S5-526	Suffolk	NY	4,5,12:i:-	Human	6	92	NYCU.JPXX01.0036
FSL S5-527	Onondaga	NY	4,5,12:i:-	Human	6	93	NYCU.JPXX01.0037
FSL S5-519	Erie	NY	Mbandaka	Human	73	97	NYCU.TDRX01.0001
FSL S5-521	Erie		Mbandaka	Human	73	98	NYCU.TDRX01.0002
FSL S5-657	Onondaga	NY	Mbandaka	Human	64	99	NYCU.TDRX01.0003
FSL S5-451	Oneida	NY	Mbandaka	Human	64	101	NYCU.TDRX01.0005
FSL S5-664	Suffolk	NY	Heidelberg	Human	50	103	NYCU.JF6X01.0003
FSL S5-408	Niagara	NY	Stanley	Human	84	110	NYCU.JNGX01.0001
FSL S5-533	Saratoga	NY	Infantis	Human	60	111	NYCU.JFXX01.0005
FSL S5-438	Monroe	NY	Weltevreden	Human	79	112	NYCU.JPQX01.0001
FSL S5-506	Kings	NY	Infantis	Human	60	113	NYCU.JFXX01.0008
FSL S5-372*			Infantis	Human	60	115	NYCU.JFXX01.0007
FSL S5-391	Westchester	NY	Kintambo	Human	59	116	NYCU.JRNX01.0001
FSL S5-503	Genesee	NY	Newport	Human	33	117	NYCU.JJPX01.0010
FSL S5-441	Monroe	NY	Abony	Human	45	118	NYCU.ABOX01.0001
FSL S5-449	Unknown	NY	Newport	Human	11	123	NYCU.JJPX01.0002
FSL S5-632	Erie	NY	Cubana	Human	71	124	NYCU.JDGX01.0001
FSL S5-399	Tompkins	NY	Hadar	Human	41	125	NYCU.TDKX01.0001
FSL S5-543	Dutchess	NY	Hadar	Human	41	125	NYCU.TDKX01.0001
FSL S5-413	Chenango	NY	Newport	Human	11	126	NYCU.JJPX01.0003
FSL S5-525	Westchester	NY	Newport	Human	11	126	NYCU.JJPX01.0003
FSL S5-541	Ulster	NY	Newport	Human	11	128	NYCU.JJPX01.0006
FSL S5-396	Suffolk	NY	Newport	Human	11	130	NYCU.JJPX01.0008
FSL S5-446	Saratoga	NY	Newport	Human	11	131	NYCU.JJPX01.0009
FSL S5-479	Suffolk	NY	Muenchen	Human	35	133	NYCU.JJ6X01.0003
FSL S5-537	Erie	NY	Newport	Human	33	134	NYCU.JJPX01.0014
FSL S5-490	Warren	NY	Worthington	Human	68	135	NYCU.TDYX01.0001
FSL S5-489	Suffolk	NY	Newport	Human	76	136	NYCU.JJPX01.0019
FSL S5-650	Suffolk	NY	Newport	Human	46	137	NYCU.JJPX01.0020
FSL S5-513	Franklin	NY	Newport	Human	46	138	NYCU.JJPX01.0021
FSL S5-484	Onondaga	NY	Schwarzen-grund	Human	4	139	NYCU.JM6X01.0002
FSL S5-524	Herkimer	NY	Newport	Human	78	140	NYCU.JJPX01.0015
FSL S5-639	Orange	NY	Newport	Human	78	141	NYCU.JJPX01.0016
FSL S5-651	Onondaga	NY	Newport	Human	78	142	NYCU.JJPX01.0017
FSL S5-643	Nassau	NY	Newport	Human	13	143	NYCU.JJPX01.0018
FSL S5-380	Niagara	NY	Litchfield	Human	31	144	NYCU.JGXX01.0001
FSL S5-504	Suffolk	NY	Muenchen	Human	74	145	NYCU.JJ6X01.0001
FSL S5-636	Saratoga	NY	Muenchen	Human	74	146	NYCU.JJ6X01.0002
FSL S5-635	Suffolk	NY	4,5,12:i:-	Human	6	147	NYCU.JPXX01.0032
FSL S5-477	Orleans	NY	Rubislaw	Human	54	148	NYCU.JLPX01.0001
FSL S5-403*		NY	Montevideo	Human	88	149	NYCU.JIXX01.0002
FSL S5-529	Erie	NY	Anatum	Human	25	151	NYCU.JAGX01.0001
FSL S5-530	Erie	NY	Anatum	Human	25	151	NYCU.JAGX01.0001
FSL S5-540	Erie	NY	Anatum	Human	25	151	NYCU.JAGX01.0001
FSL S5-464	Nassau	NY	Stanley	Human	39	152	NYCU.JNGX01.0002
FSL S5-510	Washington	NY	Hartford	Human	77	154	NYCU.JHAX01.0001
FSL S5-658	Suffolk	NY	Senftenberg	Human	18	155	NYCU.JMPX01.0001
FSL S5-457*		NY	Montevideo	Human	67	156	NYCU.JIXX01.0001
FSL S5-470	Nassau	NY	Montevideo	Human	67	156	NYCU.JIXX01.0001
FSL S5-542	Dutchess	NY	1,7:-:1,5	Human	43	157	NYCU.SC1X01.0002
FSL S5-471*	Nassau	NY	Thompson	Human	62	157	NYCU.JP6X01.0001

Table A1 [S2.1]. (Continued)

FSL S5-472	Nassau	NY	Thompson	Human	62	157	NYCU.JP6X01.0001
FSL S5-411	Wayne	NY	Thompson	Human	62	158	NYCU.JP6X01.0002
FSL S5-412	Saratoga	NY	Thompson	Human	62	158	NYCU.JP6X01.0002
FSL S5-450	Ulster	NY	Thompson	Human	62	158	NYCU.JP6X01.0002
FSL S5-523	Schenectady	NY	Thompson	Human	43	158	NYCU.JP6X01.0002
FSL S5-378	Nassau	NY	1,7:-:1,5	Human	42	160	NYCU.SC1X01.0001
FSL S5-393	Putnam	NY	Newport	Human	11	161	NYCU.JJPX01.0012
FSL S5-447	Erie	NY	Paratyphi B var. Java	Human	32	162	NYCU.JKXX01.0001
FSL S5-468	Bronx	NY	Paratyphi B var. Java	Human	32	163	NYCU.JKXX01.0002
FSL S5-654	Onondaga	NY	Nyanza	Human	66	167	NYCU.YSSX01.0001
FSL S5-515	Niagara	NY	Newport	Human	11	168	NYCU.JJPX01.0005
FSL S5-518	Wayne	NY	Newport	Human	11	168	NYCU.JJPX01.0005

*Isolates are not included in the previous study (Alcaine et al. 2006)

APPENDIX TWO

LIST OF SUPPLEMENTAL TABLES

Table A2 [S4.1]. A total of 190*Salmonella* 4,5,12:i:- and Typhimurium isolates used and their characterists

Isolate No.	State	Serotypea	Source	fimA	mdh	manB	aroC	3-gene ST	4-gene ST	PFGE Pattern No	Presence of fljA	1053-1997	fljB	hin	2740	2757	2773	Isolation year
FSL S9-177	Washington	4,5,12:i:-	Human	4	**5**	5	10	6	1	P1								2005
FSL S9-187	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P1								2006
FSL S9-193	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P1								2006
FSL S9-239	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P26	-	-	-	-	+	+	-	1999
FSL S9-244	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P27	-	-	-	-	+	+	-	1999
FSL S9-101	non-domestic bird	4,5,12:i:-	Owl (free-ranging)	4	5	5	10	6	1	P28	-	-	-	-	+	+	-	
FSL S9-206	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P28	-	-	-	-	+	+	-	1997
FSL S9-240	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P28								1999
FSL S9-242	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P28								1999
FSL S9-243	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P28	-	-	-	-	+	+	-	1999
FSL S9-246	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P28	-	-	-	-	+	+	-	1999
FSL S9-241	Spain	4,5,12:i:-	Human	4	5	7	10	8	3	P29	-	-	-	-	+	+	-	1999
FSL S9-245	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P30	-	-	-	-	+	+	-	1999
FSL S9-238	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P31	-	-	-	-	+	+	-	1999
FSL S9-204	Spain	4,5,12:i:-	Pork meat	4	5	5	10	6	1	P32	-	-	-	-	+	+	-	1998
FSL S9-205	Spain	4,5,12:i:-	Sausage	4	5	5	10	6	1	P32								1998
FSL S9-237	Spain	4,5,12:i:-	Human	4	5	5	10	6	1	P34	-	-	-	-	+	+	-	1999
FSL R6-125	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2004
FSL R6-150	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40	-	+	-	+	-	-	+	2005
FSL S9-114	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P40	-	+	-	+	-	-	+	2005
FSL S9-174	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2005
FSL S9-178	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2005
FSL S9-185	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2006
FSL S9-190	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2006
FSL S9-196	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2007
FSL S9-198	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2007
FSL S9-199	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P40								2007
FSL S5-527	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P42	-	+	-	+	-	-	+	2004
FSL S9-031	FDA	4,5,12:i:-	Food-chicken breast	4	5	5	10	6	1	P42	-	+	-					2003

Table A2 [S2.1]. (Continued)

FSL S9-109	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P46	-	+	-	+	-	-	+	2005
FSL S9-120	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P46	-	+	-	+				2006
FSL S9-188	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46	-	+	-	+	-	-	+	2006
FSL S9-191	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46								2006
FSL S9-194	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46								2006
FSL S9-195	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46								2006
FSL S9-197	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46								2007
FSL S9-200	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P46								2007
FSL S9-167	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P48								2004
FSL S9-180	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P48	-	+	-	+	-	-	+	2005
FSL S9-126	Georgia	4,5,12:i:-	Bovine	4	5	5	10	6	1	P49	-	+	-	+	-	-	+	2003
FSL S9-121	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P50	-	+	-	+	-	-	+	2005
FSL S9-122	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P50	-	+	-	+	-	-	+	2006
FSL S9-184	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P50								2006
FSL S9-179	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P55	-	+	-	+	-	-	+	2005
FSL S5-580	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P56								2004
FSL S5-596	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P56	-	+	-	+	-	-	+	2004
FSL S5-390	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2004
FSL S5-398	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2004
FSL S5-409	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2004
FSL S5-498	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2004
FSL S5-656	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2005
FSL S5-666	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2005
FSL S5-737	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2004
FSL S9-030	FDA	4,5,12:i:-	Food-chicken	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2003
FSL S9-115	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2005
FSL S9-116	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2006
FSL S9-117	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P59								2005
FSL S9-118	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P59								2005
FSL S9-127	Georgia	4,5,12:i:-	Bovine	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2003
FSL S9-168	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59	-	+	-	+	-	-	+	2004
FSL S9-169	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2004
FSL S9-181	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2005
FSL S9-182	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2006

Table A2 [S2.1]. (Continued)

FSL S9-183	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2006
FSL S9-186	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2006
FSL S9-201	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2007
FSL S9-202	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P59								2007
FSL S9-192	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P60	-	+	-	+	-	-	+	2006
FSL S5-526	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P61	-	+	-	+	-	-	+	2004
FSL S5-618	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P62	-	+	-	+	-	-	+	2004
FSL S5-816	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P64	-	+	-	+	-	-	+	2004
FSL S5-891	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P64								2004
FSL S5-759	New York	4,5,12:i:-	Bovine	4	5	5	10	6	1	P65	-	+	-	+	-	-	+	2004
FSL S9-119	Georgia	4,5,12:i:-	Poultry	4	5	5	10	6	1	P66	-	+	-	+	-	-	+	2006
FSL R6-119	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P67	-	+	-	+	-	-	+	2004
FSL R6-152	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P69	-	+	-	+	-	-	+	2005
FSL R6-153	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P69								2005
FSL S9-170	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P69								2004
FSL S9-172	Washington	4,5,12:i:-	Avian	4	5	5	10	6	1	P69								2005
FSL S9-176	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P69								2005
FSL R6-124	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P71	-	+	-	+	-	-	+	2004
FSL S9-171	Washington	4,5,12:i:-	Canine	4	5	5	10	6	1	P71	-	+	-	+	-	-	+	2005
FSL S9-173	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P71								2005
FSL S9-175	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P71								2005
FSL S9-189	Washington	4,5,12:i:-	Human	4	5	5	10	6	1	P73	-	+	-	+	-	-	+	2006
FSL S5-635	New York	4,5,12:i:-	Human	4	5	5	10	6	1	P80	-	-	-	-	+	+	+	2004
FSL S9-102	non-domestic bird	Inconsistent	Heron (free-ranging)	4	5	5	10	6	1	P21	+	-	+	+	+	+	+	
FSL S9-165	Washington	Inconsistent	Bovine	4	5	5	10	6	1	P23	+	-	+	+	+	+	+	1996
FSL S9-166	Washington	Inconsistent	Bovine	4	5	5	10	6	1	P23	+	-	+	+	+	+	+	1996
FSL R6-084	Washington	Inconsistent	Bovine	4	5	7	10	8	3	P71	+	-	+	+	+	+	+	2005
FSL S5-430	New York	Typhimurium	Bovine	4	5	7	10	8	3	P2								2004
FSL S9-110	Georgia	Typhimurium	Poultry	4	5	5	10	6	1	P35	+	-	+	+	+	+	+	2006
FSL S5-554	New York	Typhimurium	Bovine	4	5	7	10	8	3	P81								2004
FSL S5-786	New York	Typhimurium	Bovine	4	5	7	10	8	3	P81								2004
FSL S5-800	New York	Typhimurium	Bovine	4	5	7	10	8	3	P81								2004

Table A2 [S2.1]. (Continued)

FSL S5-452	New York	Typhimurium	Human	4	5	5	10	6	1	P1	+	-	+	+	+	+	+	2004
FSL S5-633	New York	Typhimurium	Human	4	5	5	10	6	1	P1								2004
FSL S5-640	New York	Typhimurium	Human	4	5	5	10	6	1	P1								2004
FSL S5-663	New York	Typhimurium	Human	4	5	5	10	6	1	P1								2005
FSL S9-207	Spain	Typhimurium	Egg	4	5	5	10	6	1	P10	+	-	+	+	+	+	+	1986
FSL S9-227	Spain	Typhimurium	Chicken sausage	4	5	5	10	6	1	P10								1986
FSL S9-229	Spain	Typhimurium	Human	4	5	5	10	6	1	P10								1986
FSL S5-799	New York	Typhimurium	Bovine	4	5	5	10	6	1	P11								2004
FSL R6-100	Washington	Typhimurium	Human	4	5	5	10	6	1	P12								2004
FSL R6-144	Washington	Typhimurium	Human	4	5	5	10	6	1	P12	+	-	+	+	+	+	+	2005
FSL S5-370	New York	Typhimurium	Human	4	5	5	10	6	1	P12	+	-	+	+	+	+	+	2004
FSL S5-375	New York	Typhimurium	Human	4	5	5	10	6	1	P12								2004
FSL S5-429	New York	Typhimurium	Bovine	4	5	5	10	6	1	P12								2004
FSL S5-488	New York	Typhimurium	Human	4	5	5	10	6	1	P12								2004
FSL S5-509	New York	Typhimurium	Human	4	5	5	10	6	1	P12								2004
FSL S5-535	New York	Typhimurium	Human	4	5	5	10	6	1	P12								2004
FSL S5-845	New York	Typhimurium	Bovine	4	5	5	10	6	1	P12								2004
FSL S5-933	New York	Typhimurium	Bovine	4	5	5	10	6	1	P12								2004
FSL S9-209	Spain	Typhimurium	Chicken s	4	5	5	10	6	1	P12								1989
FSL S9-213	Spain	Typhimurium	Human	4	5	5	10	6	1	P12								1986
FSL S9-215	Spain	Typhimurium	Human	4	5	5	10	6	1	P12								1988
FSL S9-216	Spain	Typhimurium	Meat	4	5	5	10	6	1	P12								1989
FSL S9-218	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1991
FSL S9-219	Spain	Typhimurium	River	4	5	5	10	6	1	P12								1992
FSL S9-220	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1992
FSL S9-221	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12	+	-	+	+	+	+	+	1992
FSL S9-222	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1992
FSL S9-223	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1993
FSL S9-224	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1994
FSL S9-225	Spain	Typhimurium	Beach	4	5	5	10	6	1	P12								1996
FSL S9-226	Spain	Typhimurium	Pork saus.	4	5	5	10	6	1	P12								1996
FSL S9-230	Spain	Typhimurium	Chicken	4	5	5	10	6	1	P12								1988
FSL S9-214	Spain	Typhimurium	Meat	4	5	5	10	6	1	P13								1988
FSL S9-217	Spain	Typhimurium	Beach	4	5	5	10	6	1	P14								1991
FSL S5-534	New York	Typhimurium	Human	4	5	5	10	6	1	P15								2004

Table A2 [S2.1]. (Continued)

FSL S5-607	New York	Typhimurium	Bovine	4	5	5	10	6	1	P16								2004
FSL S5-492	New York	Typhimurium	Human	4	5	5	10	6	1	P17								2004
FSL S5-536	New York	Typhimurium	Human	36	5	5	10	49	9	P18	+	-	+	+	+	+	+	2004
FSL S5-494	New York	Typhimurium	Human	4	5	5	10	6	1	P19								2004
FSL S9-247	Spain	Typhimurium	Beach	4	5	5	10	6	1	P19								1996
FSL S5-653	New York	Typhimurium	Human	4	5	5	10	6	1	P20								2005
FSL R6-131	Washington	Typhimurium	Human	4	5	5	18	6	2	P24								2005
FSL S5-394	New York	Typhimurium	Human	4	5	5	10	6	1	P25								2004
FSL R6-154	Washington	Typhimurium	Human	4	5	5	10	6	1	P35								2005
FSL S9-231	Spain	Typhimurium	Beach	34	5	5	10	47	7	P35	+	-	+	+	+	+	+	1994
FSL R6-136	Washington	Typhimurium	Human	4	5	5	10	6	1	P36								2004
FSL S5-493	New York	Typhimurium	Human	34	5	5	10	47	7	P36								2004
FSL S5-507	New York	Typhimurium	Human	4	5	5	10	6	1	P36	+	-	+	+	+	+	+	2004
FSL S9-125	Georgia	Typhimurium	Poultry	4	5	5	10	6	1	P36	+	-	+	+	+	+	+	
FSL S9-212	Spain	Typhimurium	Human	4	5	5	10	6	1	P37								1992
FSL S5-462	New York	Typhimurium	Human	4	5	5	10	6	1	P38								2004
FSL S5-501	New York	Typhimurium	Human	4	5	5	10	6	1	P39								2004
FSL S5-473	New York	Typhimurium	Human	4	5	5	10	6	1	P4								2004
FSL S5-511	New York	Typhimurium	Human	4	5	5	10	6	1	P4								2004
FSL S5-520	New York	Typhimurium	Human	4	5	5	10	6	1	P4								2004
FSL S5-620	New York	Typhimurium	Bovine	4	5	5	10	6	1	P4								2004
FSL S5-747	New York	Typhimurium	Bovine	4	5	5	10	6	1	P4								2004
FSL S5-831	New York	Typhimurium	Bovine	4	5	5	10	6	1	P4								2004
FSL S9-103	non-domestic bird	Typhimurium	FL Emu (captive)	4	5	5	10	6	1	P4	+	-	+	+	+	+	+	
FSL S9-130	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P41								1999
FSL S5-392	New York	Typhimurium	Human	4	5	5	10	6	1	P44								2004
FSL S5-788	New York	Typhimurium	Bovine	36	5	5	10	49	9	P45								2004
FSL S9-210	Spain	Typhimurium	Water	34	5	5	10	47	7	P45	+	-	+	+	+	+	+	1990
FSL S9-129	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P47								2003
FSL S9-131	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P51								2002
FSL S9-208	Spain	Typhimurium	Water	4	5	5	10	6	1	P52								1986
FSL S9-228	Spain	Typhimurium	Water	4	5	5	10	6	1	P52	+	-	+	+	+	+	+	1986
FSL S5-936	New York	Typhimurium	Bovine	4	5	5	10	6	1	P53								2004

Table A2 [S2.1]. (Continued)

FSL S9-123	Georgia	Typhimurium	Poultry	4	5	5	10	6	1	P57								
FSL S5-397	New York	Typhimurium	Human	4	5	5	10	6	1	P6								2004
FSL S9-124	Georgia	Typhimurium	Poultry	4	5	5	10	6	1	P63								2000
FSL S5-381	New York	Typhimurium	Human	4	5	5	10	6	1	P7								2004
FSL S5-805	New York	Typhimurium	Bovine	4	5	5	10	6	1	P7								2004
FSL S5-433	New York	Typhimurium	Bovine	4	5	5	10	6	1	P71								2004
FSL S9-133	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P72								2000
FSL S5-516	New York	Typhimurium	Human	4	5	5	10	6	1	P74								2004
FSL S5-896	New York	Typhimurium	Bovine	4	5	5	10	6	1	P75								2004
FSL S5-505	New York	Typhimurium	Human	4	5	5	10	6	1	P76								2004
FSL S5-532	New York	Typhimurium	Human	4	5	5	10	6	1	P76								2004
FSL S5-555	New York	Typhimurium	Bovine	4	5	5	10	6	1	P76								2004
FSL S5-531	New York	Typhimurium	Human	4	5	6	10	7	8	P77								2004
FSL S5-564	New York	Typhimurium	Bovine	4	5	5	10	6	1	P77								2004
FSL S9-132	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P78								1995
FSL S9-211	Spain	Typhimurium	Human	4	5	5	10	6	1	P79								1991
FSL S5-729	New York	Typhimurium	Bovine	4	5	5	10	6	1	P8								2004
FSL R6-161	Washington	Typhimurium	Human	4	5	5	10	6	1	P81	+	-	+	+	+	+	+	2005
FSL S5-916	New York	Typhimurium	Bovine	4	5	7	10	8	3	P81								2004
FSL S5-796	New York	Typhimurium	Bovine	4	5	7	10	8	3	P82								2004
FSL S9-234	Spain	Typhimurium	Beach	34	5	5	10	47	7	P83								1996
FSL S9-233	Spain	Typhimurium	Beach	4	5	5	10	6	1	P84								1994
FSL S9-128	Georgia	Typhimurium	Bovine	4	5	5	10	6	1	P85								2002
FSL R6-002	Washington	Typhimurium	Bovine	4	5	7	10	8	3	P86								2004
FSL S9-232	Spain	Typhimurium	Beach	34	5	5	10	47	7	P87								1994
FSL S9-235	Spain	Typhimurium	Minced beef	4	5	5	10	6	1	P88								1996
FSL S9-112	Georgia	Typhimurium	Poultry	4	5	5	10	6	1	P9								2006

[a] these isolates were serotyped as 4,5,12:i:- in one replicate and Typhimurium in another replicate (including one isolates that was classified as 4,5,12:i:- in two replicates and Typhimurium in one replicate) and were thus designated as “inconsistent”

Table A2 [S4.2]. Primers and PCR conditions

Gene	Size of amplicon, bp	Primers[a]	Reaction parameters[b]
STM2692	919	F: 5’- ATA TTC AGC GTG AAC GGG CG-3’ R: 5’- ACG CCG TCA AGC CCG CCG-3’	95°C for 10 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for .15 min (30x); 72 °C for 7 min (1x)
STM2694	214	F: 5’- TGA ACT GTC CAG AGT GCG G-3’ R: 5’- TCA GAA ACT CAT GTG GCC TTG ACC-3’	95°C for 10 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for .15 min (30x); 72 °C for 7 min (1x)
STM2741	543	F: 5’- AAG CGC GGC ATC TCG CCC-3’ R: 5’- AAG CCC ATC CGA CGG C-3’	95°C for 10 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for .15 min (30x); 72 °C for 7 min (1x)
STM2758	653	F: 5’- ATT GCC ATG CTG CCT GCC GC-3’ R: 5’- AGC CAG AAC GTC GGC C-3’	95°C for 10 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for .15 min (30x); 72 °C for 7 min (1x)
STM2774	642	F: 5’- TCG GTT GAA GGT CAG ATT ATC GGG C-3’ R: 5’- TAA CTG CAG TGT TGA ACG GCG G-3’	95°C for 10 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for .15 min (30x); 72 °C for 7 min (1x)
manB	790	F: 5'-CAT AAY CCG ATG GAC TAC AAC G-3' 5'-ACC AGC AGC CAC GGG ATC AT-3'	95°C for 10 min (1x); 95°C for 45 sec, TD 55°C -45 °C for 45 sec, 72°C for 1 min (40x); 72°C for 7 min (1x)
fimA	760	F: 5'- TCA GGG GAG AAA CAG AAA ACT AAT -3' R: 5'- TCC CCG ATA GCC TCT TCC -3'	95°C for 10 min (1x); 95°C for 45 sec, 57.1°C for 45 sec, 72°C for 1 min (45x); 72°C for 7 min (1x)
mdh	849	F: 5'-GAT GAA AGTCGC AGT CCT CG-3' R: 5'-TAT CCA GCA TAG CGT CCA GC-3	95°C for 10 min (1x); 95°C for 45 sec, TD 55°C -45°C for 45 sec, 72°C for 1 min (40x); 72°C for 7 min (1x)
aroC	501	F: 5'-GGCACCAGTATTGGCCTGCT-3' R: 5'-CATATGCGCCACAATGTGTTG-3'	95°C for 5 min (1x); 95°C for 1 min, 55°C for 45 sec, 72°C for 1 min (30x); 72°C for 7 min (1x)

[a]R: reverse primer; F: forward: primer

[b]TD: Touch down PCR; annealing temperatures decreased 0.5 ºC/cycle during the first 20 cycles, followed by 20 cycles at 55°C