CHARACTERIZATION OF *SALMONELLA ENTERICA* SEROTYPE ENTERITIDIS SUBPOPULATIONS USING COMPARATIVE GENOME SEQUENCING AND MUTATIONAL ANALYSIS

by

CESAR AGUSTO MORALES

(Under the Direction of Mark Harrison)

ABSTRACT

Salmonella enterica serotype Enteritidis continues to be the leading cause of salmonellosis both worldwide and in the United States, in part due to its unique ability to contaminate the internal contents of table eggs. Subpopulations that express varying phenotypes within the same phage type of *S.* Enteritidis are successful in different ecological niches but complement each other to complete the overall egg contamination pathway. Comparative Genomic Sequencing was used to determine the nucleotide sequence identity of all polymorphisms between two subpopulations of *S.* Enteritidis that differ in their lipopolysaccharide structure, ability produce biofilm, ability to achieve high-cell density growth, Phenotype Microarray results, and in their ability to contaminate eggs. A total of 247 regions of polymorphisms were found to differentiate the two subpopulations, ranging in size from one single nucleotide polymorphism to a 215 base pair deletion. Twelve polymorphisms were predicted to disrupt open reading frames (ORFs), 99 were predicted to result in amino acid substitutions, 82 were predicted to occur within an ORF but were synonymous, 38 polymorphisms were found in non-coding intergenic regions, and 12 were found in RNA genes.

Within these results, it is believed that the genetic variation reported is responsible for the phenotype variation observed between the two subpopulations. Three of the ORF-disrupting polymorphism were selected for further characterization based on their predicted genotype-phenotype relationships. Using site-directed mutagenesis, the genes *sen4316* and *dsdA* were confirmed to be necessary for biofilm formation and D-serine catabolism, respectively. Phenotype Microarray results further suggested a role of osmotic resistance for *dsdA.* The gene *sefD*, encoding a minor fimbrial subunit that is unique to *S.* Enteritidis, was analyzed in an egg production hen experiment for its role in chicken reproductive tract tropism. No significant difference in egg production was observed between the *sefD* mutant and wild type strains, but overexpression of *sefD* resulted in attenuation, suggesting a possible role for modulating illness in the chicken. The research herein represents the first steps towards determining genetic determinants for subpopulation heterogeneity, including the ability to contaminate the egg.

INDEX WORDS: *Salmonella*, Egg contamination, Comparative Genomic Sequencing, Single nucleotide polymorphism, Mutagenesis, Phenotype Microarray

CHARACTERIZATION OF *SALMONELLA ENTERICA* SEROTYPE ENTERITIDIS SUBPOPULATIONS USING COMPARATIVE GENOME SEQUENCING AND MUTATIONAL ANALYSIS

by

CESAR AGUSTO MORALES

B.S., The University of Georgia, 2001

M.S., The University of Georgia 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2010

CHARACTERIZATION OF *SALMONELLA ENTERICA* SEROTYPE ENTERITIDIS SUBPOPULATIONS USING COMPARATIVE GENOME SEQUENCING AND MUTATIONAL ANALYSIS

by

CESAR AGUSTO MORALES

Major Professor:	Mark Harrison
Committee:	Jinru Chen
	Joseph Frank
	Jean Guard
	Robert Maier

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2010

DEDICATION

I dedicate this work to my family, to my wife, Dagny, and my daughter, Stella. Thank you, my loves, for providing the motivation and support that I needed to complete this work.

ACKNOWLEDGEMENTS

Funding for this research and for graduate student support was provided by USDA-ARS-ESQRU CRIS project 6612-32000-004-00 Controlling Egg Contamination with *Salmonella enterica* by Understanding Its Evolution and Pathobiology. The research described herein is the property of the United States Department of Agriculture and is released at the discretion of the agency.

I would like to thank many scientists for setting me on this path as an undergraduate, including Dr. John Maurer, Dr. Margie Lee, and Dr. Charles Hofacre. I would like to thank my committee members for their specific guidance on this project, including Dr. Jinru Chen, Dr. Joe Frank, and Dr. Rob Maier. I would like to thank my major professor, Dr. Mark Harrison, for his patience and teachings during my graduate career. I would like to thank my fellow classmate, Dr. Jennifer Turpin, for being there every step of the way. I would like to thank my lab associates, Dr. Roxana Sanchez-Ingunza and Tod Stewart for their technical aid in this research. Above all, I would like to thank my mentor, Dr. Jean Guard, for teaching me how to be a scientist.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Nontyphoidal salmonellae are estimated to be the leading bacterial cause of gastroenteritis both worldwide and in the United States (11-12). In 2000, the U.S. Department of Health and Human Services set a national goal in their Healthy People 2010 agenda for a reduction in salmonellosis to 6.8 illness per 100,000 people by the year 2010 (13). However, active surveillance by the Center for Disease and Control (CDC) Foodborne Diseases Active Surveillance Network (FoodNet) reported an incidence of 15.19 illnesses per 100,000 people in 2009 and the incidence has never been below 12.3 per 100,000 people since reporting started in 1996 (1-11). The CDC further estimates that *Salmonella* species cause 1.4 million illnesses, 15,000 hospitalizations and 400 deaths each year in the U.S. and global estimates reported are as high as 93.8 million cases of salmonellosis and 155,000 deaths each year (21-22). Given the high impact to human health and the difficulty in controlling the pathogen, *Salmonella* remains an important food safety research focus in the areas of biology, virulence, epidemiology, and prevention.

The genus *Salmonella* is comprised of two species, *Salmonella enterica* and *Salmonella bongori* and consists of over 2,500 serotypes (18). *Salmonella enterica* is further divided into six subspecies, of which, serotypes of *Salmonella enterica* subspecies *enterica* (subspecies I) are primarily associated with the infection of warm-blooded animals (26). *Salmonella enterica* subspecies *enterica* serotype Enteritidis (*S.* Enteritidis) is the leading cause of salmonellosis worldwide and, since 2007, in the United States (7-8, 16). The pathogen has been associated

with a wide variety of food vectors but its unique ability to contaminate the internal contents of table eggs has been the primary source associated with *S.* Enteritidis infection (16). There has been extensive research into understanding mechanisms associated with egg contamination, but a clear-cut answer remains unknown. Likely, a multitude of factors contribute to the overall egg contamination pathway, and the results of various research studies identifying potential factors involved in colonization, invasion, and survival support this concept (14-15, 24). Therefore, comprehensive characterization of all contributing factors would provide a complete picture of mechanisms involved in the egg contamination pathway and lead to improved prevention strategies.

Initial DNA typing methods, such as pulsed field gel electrophoresis and ribotyping, lead to the belief that *S.* Enteritidis was a highly clonal organism (19). However, more powerful discriminatory methods have found greater genetic diversity than previously thought (20, 23). Research in subpopulation heterogeneity of isolates within the same phage type (PT) of *S.* Enteritidis has further demonstrated that not all strains are created equal and, in fact, can exhibit dramatically diverse phenotypes. One observed subpopulation produces high molecular mass (HMM) lipopolysaccharide (LPS) and can grow to high cell density, and it is associated with phenotypic characteristics that have been correlated with high incidence of egg contamination following intravenous infection of hens (15). However, infection only occurs at low incidence when this subpopulation is given orally (25). Another observed *S*. Enteritidis subpopulation produces low molecular mass (LMM) LPS and produces biofilm when grown at ambient temperatures. This subpopulation is capable of causing systemic infection in chickens following oral dosing, but it does not contaminate the egg (15, 17). These two subpopulations are poor or deficient at egg contamination when used alone in low-dose contact infection studies. However,

high incidence of egg contamination was achieved when the subpopulations were combined in a single inoculum, which demonstrated a complementation requirement for some isolates of *S.* Enteritidis (15). Additional characterization of the *S.* Enteritidis subpopulations revealed that they vary in phenotype by up to 20% of 1,920 test conditions assayed in a Phenotype Microarray experiment, but no differences in gene content were detected by DNA-DNA microarray hybridization (24). Since the DNA-DNA microarray hybridization platform used was only sensitive to large-scale nucleotide differences and a few single nucleotide polymorphisms were previously reported (23), it was concluded that the genetic determinants responsible for the varying phenotypes, including the ability to contaminate the egg, were small-scale nucleotide polymorphisms (24).

The primary objective of the research described herein is to determine the sequence identity of all small-scale genetic differences between the two subpopulations of *S.* Enteritidis and to characterize, via mutagenesis, selected polymorphisms that may account for physiological differences that aid completion of a complex infection pathway that results in egg contamination. Identifying such determinants could lead to improved and meaningful epidemiological tools, as well as potential new prevention strategies.

References

1. **Anonymous.** 2000. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 1999. MMWR Morb Mortal Wkly Rep **49:**201-5.

2. **Anonymous.** 2001. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2000. MMWR Morb Mortal Wkly Rep **50:**241-6.

3. **Anonymous.** 2002. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2001. MMWR Morb Mortal Wkly Rep **51:**325-9.

4. **Anonymous.** 2003. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2002. MMWR Morb Mortal Wkly Rep **52:**340-3.

5. **Anonymous.** 2005. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 sites, United States, 2004. MMWR Morb Mortal Wkly Rep **54:**352-6.

6. **Anonymous.** 2007. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2006. MMWR Morb Mortal Wkly Rep **56:**336-9.

7. **Anonymous.** 2008. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2007. MMWR Morb Mortal Wkly Rep **57:**366-70.

8. **Anonymous.** 2009. Preliminary FoodNet Data on the incidence of infection with pathogens transmitted commonly through food--10 States, 2008. MMWR Morb Mortal Wkly Rep **58:**333-7.

9. **Anonymous.** 2006. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 States, United States, 2005. MMWR Morb Mortal Wkly Rep **55:**392-5.

10. **Anonymous.** 2004. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--selected sites, United States, 2003. MMWR Morb Mortal Wkly Rep **53:**338-43.

11. **Anonymous.** 2010. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food - 10 states, 2009. MMWR Morb Mortal Wkly Rep **59:**418-22.

12. **Anonymous.** 2010. The Community Summary Report on Trends and Sources of Zooneses, Zoonotic Agents and Food-borne Outbreaks in the European Union in 2008. European Food Safety Authority **1496**.

13. **Anonymous.** 2000. Food Safety. *In* U.S. Department of Health and Human Services (ed.), Healthy People 2010: Understanding and Improving Health, 2nd ed. U.S. Government Printing Office, Washington, DC.

14. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, T. J. Humphrey, and F. Van Immerseel.** 2009. Mechanisms of egg contamination by *Salmonella* Enteritidis. FEMS Microbiol Rev **33:**718-38.

15. **Guard-Bouldin, J., R. K. Gast, T. J. Humphrey, D. J. Henzler, C. Morales, and K. Coles.** 2004. Subpopulation characteristics of egg-contaminating *Salmonella enterica* serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl. Environ. Microbiol. **70:**2756-63.

16. **Guard-Petter, J.** 2001. The chicken, the egg and *Salmonella* Enteritidis. Environ. Microbiol. **3:**421-30.

17. **Guard-Petter, J., L. H. Keller, M. M. Rahman, R. W. Carlson, and S. Silvers.** 1996. A novel relationship between O-antigen variation, matrix formation, and invasiveness of *Salmonella* Enteritidis. Epidemiol. Infect. **117:**219-31.

18. **Guibourdenche, M., P. Roggentin, M. Mikoleit, P. I. Fields, J. Bockemuhl, P. A. Grimont, and F. X. Weill.** 2010. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. Res Microbiol **161:**26-9.

19. **Liebana, E.** 2002. Molecular tools for epidemiological investigations of *S. enterica* subspecies *enterica* infections. Res. Vet. Sci. **72:**169-75.

20. **Lindstedt, B. A., T. Vardund, L. Aas, and G. Kapperud.** 2004. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J. Microbiol. Methods **59:**163-72.

21. **Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra.** 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50:**882-9.

22. **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:**607-25.

23. **Morales, C. A., M. Musgrove, T. J. Humphrey, C. Cates, R. Gast, and J. Guard-Bouldin.** 2007. Pathotyping of *Salmonella enterica* by analysis of single-nucleotide polymorphisms in cyaA and flanking 23S ribosomal sequences. Environ Microbiol **9:**1047-59.

24. **Morales, C. A., S. Porwollik, J. G. Frye, H. Kinde, M. McClelland, and J. Guard-Bouldin.** 2005. Correlation of phenotype with the genotype of egg-contaminating *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **71:**4388-99.

25. **Parker, C. T., B. Harmon, and J. Guard-Petter.** 2002. Mitigation of avian reproductive tract function by *Salmonella* Enteritidis producing high-molecular-mass lipopolysaccharide. Environ. Microbiol. **4:**538-45.

26. **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:**5883-98.

CHAPTER 2

LITERATURE REVIEW

Foodborne Disease

The impact of foodborne disease to human health and global economies is significant. In 2006, the World Health Organization (WHO) Department of Food Safety and Zoonoses began an initiative to estimate the global burden of foodborne disease. The assembled task force, referred to as the Foodborne Disease Burden Epidemiology Reference Group has set a target date of 2012 to report the estimate of foodborne illness caused by chemicals, parasites, and enteric infections on a global scale (100). Until then, the best estimate for the global burden of foodborne illness is 1.08 billion cases each year, which is based on an estimate of 2.8 billion cases of diarrheal illness each year world wide by Scallan *et al.* applied with the 36% foodborne transmission factor used by Mead *et al*. (78, 97). The Center for Disease Control and Prevention (CDC) estimates that in the United States (U.S.) alone there are 76 million illnesses, 325,000 hospitalizations, and 5,000 deaths caused by foodborne disease (78). The health-related costs of acute foodborne illness in the U.S. alone is estimated to be $152 billion each year, further highlighting its significant burden (98).

Foodborne disease is an illness whereby the identified etiological agent is transmitted by the consumption of a contaminated food. The number and diversity of microbial pathogens capable of contaminating food and water is considered vast and a majority are unknown (78). Bacteria, viruses, parasites, marine dinoflagellates, preformed toxins, and prions have all been implicated in foodborne disease. Technological advances in surveillance methods continue to

add to number of known foodborne pathogens (104, 107). Foodborne pathogens can be zoonotic, environmental, human-host-restricted, or have other routes of transmission (107). The challenge of protecting the public from foodborne illness is indeed multifaceted.

Newly recognized pathogens further compound the challenge related to foodborne disease. They are pathogens with previously little or no history of foodborne illness that are recently recognized or have increased incidence of illness. Factors that may contribute to the emergence of a new pathogen may be as simple as the implementation of a new sampling and testing methods or be the effect of a change in a food process, environment, technology, or even consumer fad. The pathogen itself may change by means of acquiring new virulence factors or metabolic processes that increase pathogenicity or survivability in a new ecological niche. Among the CDC's list of twenty-seven principle foodborne pathogens, thirteen emerged in the 1970's or later, indicative that newly recognized pathogens should be expected to emerge as time goes on (78, 104, 107).

Among foodborne pathogens, viruses are believed to be responsible for the highest number of foodborne related illnesses. Known foodborne viruses are estimated to cause over 9 million cases of foodborne disease each year in the U.S.. In comparison, bacterial pathogens are estimated to cause 4 million cases and parasitic pathogens cause an estimated 350,000 illnesses each year. Noroviruses are predicted to be responsible for over 99% of illnesses transmitted by foodborne viruses, while *Campylobacter* spp., *Salmonella* spp., and *Escherichia coli* are estimated to be responsible for 47%, 32%, and 4%, respectively, of bacterial illnesses transmitted by contaminated food (78). In contrast to estimates, the CDC's Foodborne Diseases Active Surveillance Network (FoodNet) reported that the top three foodborne pathogens with active surveillance were *Salmonella* spp. with 7,039 cases, *Campylobacter* spp. with 6,033 cases, and

Shigella spp. with 1,849 cases in 2009 (17). In other countries, Australia reported 9,484 illnesses caused by *Salmonella* spp., 16,984 illnesses caused by *Campylobacter* spp., and 597 cases caused by *Shigella* spp. in 2007 (39). In England and Wales during the period of 1992 to 2003, 39,625 people were affected by outbreaks, over half of which were by caused by *Salmonella* species (61).

Salmonella: A Persistent and Evolving Cause of Bacterial Foodborne Illness

Salmonella species are gram negative, rod-shaped, non-spore forming bacteria and are a leading cause of acute gastroenteritis worldwide. The global burden of non-typhoidal *Salmonella* infections is estimated to be 93.8 million cases of salmonellosis and 155,000 deaths each year, of which 80.3 million cases are estimated to be foodborne (75). In the United States, the CDC estimates 1.4 million annual cases of non-typhoidal *Salmonella* infections, resulting in 15,000 hospitalizations and 400 deaths each year (110). From 2001 to 2009, CDC's active surveillance program, FoodNet, has reported *Salmonella* as the bacterial pathogen causing the highest number of reported foodborne illnesses in the U.S. followed by *Campylobacter*, *Shigella*, *E. coli*, *Vibrio*, and *Listeria* (9-17). Symptoms of salmonellosis include diarrhea, headache, abdominal pain, nausea, chills, fever and vomiting. The onset of illness occurs within a few to 72 hours after ingestion, depending on the dose (89). The disease is most often self-limiting, but invasive infections occur at the average rate of 0.9 cases per 100,000 persons and are more likely to occur in young, elderly, and immuno-compromised individuals (112).

Salmonella is divided into two species: *S. bongori* and *S. enterica*. The species *Salmonella enterica* is further divided into six subspecies: *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI) (23). Further classification of *Salmonella* is performed by designating serotype based on the Kauffmann-White antigenic

typing scheme which generates antigen profiles based on agglutination reactions with homologous antisera that target lipopolysaccharide O-chain, flagellin H and, for a few serotypes, capsular Vi antigens. *Salmonella* consists of over 2,500 serotypes and continues to grow due to the hyper-variability of the genetic loci for the antigenic determinants (53). It is hypothesized that nearly all serotypes are capable of causing disease in humans; however, the epidemiology of salmonellosis indicates that only a handful of serotypes do so on a recurring basis (30). *Salmonella enterica* subspecies *enterica*, consisting of over 1,500 serotypes, is of particular human health and agricultural interest because it accounts for almost all *Salmonella* infections of warm-blooded mammals (24, 90).

Table 2.1 summarizes the incidence of *Salmonella* infections in the U.S. and their serotype proportions between the years 2000 and 2009 as reported by CDC's FoodNet active surveillance program. The top six serotypes during that time period were predominantly *S.* Enteritidis, *S.* Typhimurium, *S.* Newport, *S.* Heidelberg, *S.* Javiana, and serotype I 4,[5],12:i:-. Contamination by *Salmonella* serotypes has been attributed to a wide variety of food sources. While *Salmonella* contamination is typically associated with foods of animal origin, such as beef, poultry, eggs, and dairy products, it has also been implicated in outbreaks with fresh produce, juice beverages, and nuts. A few examples include *Salmonella* serotypes *S.* Muenchen and *S.* Hartford outbreaks associated with unpasteurized orange juice, serotypes *S.* Saintpaul and *S.* Kottbus outbreaks associated with alfalfa sprouts, serotype *S.* Enteritidis outbreak from consumption of raw almonds, and a serotype *S.* Tennessee outbreak associated with peanut butter (3-8). From 1973 through 1997, *Salmonella* was responsible for 48% of produce-associated outbreaks with a known bacterial pathogen (103). *Salmonella* is also found in fish and other

seafood with serotypes *S.* Weltevreden and *S.* Senftenberg being the most frequently isolated serotypes (55).

The host range of *Salmonella* spp. varies among the serotypes. A majority of the *Salmonella* subspecies *enterica* serotypes exhibit a broad, host range, however, host adapted and host specific serotypes are also observed. Host adapted serotypes are usually found only in a primary host, but they can occasionally be found in a secondary host. Host specific serotypes only infect a single host species (56). Serotypes *S.* Choleraesuis, *S.* Typhimurium variant (var.) Copenhagen, and *S.* Dublin are host adapted for porcine, pigeon, and bovine animals, respectively. Host-specific serotypes include human-specific serotypes *S.* Typhi, *S.* Paratyphi A, *S.* Paratyphi B, and *S.* Paratyphi C and poultry-specific serotypes *S.* Pullorum and *S.* Gallinarum (25, 56). Additionally, some serotypes can colonize unusual ecological niches, such as *Salmonella enterica* serotype Enteritidis, which contaminates the internal contents of eggs produced by otherwise healthy chickens (48, 77). Humans then become ill by consuming these contaminated eggs. Understanding how *Salmonella* evolves to colonize specific niches such as the egg is a primary goal of research described herein.

Salmonella enterica **subspecies** ***enterica*** **serotype Enteritidis**

Salmonella enterica subspecies *enterica* serotype Enteritidis (*S.* Enteritidis) is the leading cause of salmonellosis both worldwide (48) and in the U.S. from 2007 through 2009 (Table 2.1). It is a broad-range host pathogen that represents a considerable challenge to food safety with its unique ability to contaminate the internal contents of the egg without causing or showing signs of illness in the chicken (48, 89). Contamination of eggs in this manner correlates with this one serotype causing worldwide illness (38). Incidence of *S.* Enteritidis in the U.S. peaked from the mid-1980s to the mid-1990s with rates reaching near 10 cases per 100,000 persons. During this

time period, 80% of reported *S.* Enteritidis outbreaks with implicated food items were egg associated (87). During the period of 1993-2002, *S.* Enteritidis accounted for the largest number of outbreaks and outbreak-related cases among bacterial pathogens in the U.S. (74, 84). One theory suggests that the eradication of endemic *Salmonella* serotypes Pullorum and Gallinarum in the 1970s removed cross-immunity to *S.* Enteritidis in the flocks, thus, allowing the pathogen to flourish (92). The incidence steadily declined from 1996 to 1999, down to about 3 confirmed human cases per 100,000 individuals per year in the U.S though the incidence remained much higher in the European Union at 15.34 cases per 100,000 individuals for 2008 (18, 87). It is estimated that shell eggs contaminated with *S.* Enteritidis caused 182,060 illnesses in the U.S. in the year 2000 (99).

Additional routes of *S.* Enteritidis infection for humans include raw meat, poultry, dairy products, nuts, fruits and vegetables; however, infection from the ingestion of raw or under-cooked eggs, egg products or foods containing such eggs, is the strongest correlation (62, 88). The infection pathway to the egg involves interactions of the pathogen with multiple environments including the hen house, the bird, the egg, as well as the human host. Symptoms of disease in poultry are mostly seen in young chicks. The lack of symptoms in an infected, mature hen and the production of eggs with little or no discernable change are major factors in the persistence of the pathogen in the farm house and its dominance as a bacterial foodborne pathogen (62, 85). Poultry can be infected by both vertical transmission and horizontal transmission via shedding, rodents, feed, and airborne droplets or dust particles (20, 89). Furthermore, stress factors such as molting can increase the hen's susceptibility to infection, increase shedding of *S.* Enteritidis in the environment, and increase the proportion of contaminated eggs within the first five weeks post-molt (45, 59).

Risk assessment analysis of *Salmonella* in eggs by the World Health Organization and Food and Agriculture Organization of the United Nations summarizes that the risk of a contaminated egg is, in part, dependent on the prevalence of contaminated flocks (19). During peak infection rates in the U.S., the prevalence of *S.* Enteritidis-positive poultry houses for the Northern, Southeastern, and Central/Western states was 45%, 3%, and 17%, respectively (34). In 2004, sampled egg-laying farms in California had an *S.* Enteritidis prevalence of 10.5% (66). The current estimate of contaminated eggs in the U.S. as a result of *S.* Enteritidis persistence in the farm is about 2.3 million contaminated eggs of the 69 billion produced per year, or 1 in every 30,000 eggs (60).

Contamination of the eggshell by *S.* Enteritidis and other serotypes can occur as the egg contacts the fecal microflora during passage through the cloaca (oviposition), and this process is referred to as horizontal transmission (62, 89). *Salmonella* present on the egg shell before the proteinaceous cuticle layer is formed can enter the internal contents through open pores in the egg shell which is more easily penetrated immediately after the egg is laid (89). An environment of negative pressure created by the egg going from a 42°C chicken body temperature to the cooler 25°C room temperature facilitates penetration. Older eggs may also be more susceptible to bacterial penetration as the cuticle layer becomes dehydrated, exposing more pores (40). Cracks in the eggshell provide another means for bacterial entry into the internal contents, which was a major source of illness prior to the 1970s when mandatory inspection was started (48). However, vertical transmission into the preformed egg is believed to be the primary cause of *S.* Enteritidis outbreaks, which is most likely a consequence of reproductive tissue colonization in infected laying hens (30, 77, 111).

It appears that the egg-laying hen must become systemically infected by *S.* Enteritidis before the reproductive tract can be colonized and transovarian transmission can proceed. Fimbriae, adhesins, flagellin, and lipopolysaccharide (LPS) on the *Salmonella* cell surface can interact with Toll-like and other pattern recognition receptors to facilitate adhesion and invasion of host epithelial cells (63). *Salmonella* spp. contain in their genome pathogenicity islands (SPI) that contain a number of crucial virulence factors that are responsible for intestinal epithelial invasion and systemic spread (36). The SPI-1, in particular, encodes a type III secretion system (T3SS) that transports bacterial effector proteins into the cytosol of host cells and cause membrane ruffling, which results from rearrangement of the actin cytoskeleton. As the cytoskeleton rearranges, the host cell engulfs the *Salmonella* cell into a membrane-bound vacuole called the *Salmonella*-containing vacuole (SCV) (37, 63). A *Salmonella* cell is able to survive and replicate in the SCV, because its presence interferes with progression of an endosome to a phagolysosome. This prevents exposure of internalized *Salmonella* cells to the bactericidal components contained therein. Invasion of the intestinal epithelial cells attract host immune phagocytes, particularly monocytes and macrophages (blood borne monocytes), which phagocytize the altered host cell as well as the bacteria-containing SCV. *Salmonella* survival and replication in these phagocytes is dependent on a SPI-2 type III secretion system which injects effector proteins within the phagocytic vacuole to prevent fusion of the phagosome with the lysosomes and, thus, the bacteria are able to spread throughout the body by transportation via monocytes and macrophages (26). *Salmonella* can then induce macrophage apoptosis to be released when a target organ is reached (36-37).

While it is widely accepted that colonization of the avian reproductive tissue is required for intact shell-egg *Salmonella* Enteritidis contamination, the exact site or sites of colonization

remains unknown. The two reproductive domains of investigation are the ovary, which represents the origination site of the preformed egg, and oviduct, which is the tract that the developing egg travels during oviposition. Animal infection studies by different investigators present conflicting results with some saying the yolk is the most frequently contaminated, which would indicate ovarian colonization, while others report that the albumen is the most frequently contaminated, which would indicate oviduct colonization (32, 40). Regardless of colonization site preference, the conflicting studies at least suggest that any site of the reproductive tract can be colonized. Colonization of the ovary is believed to take place and remain on the membrane of the preovulatory follicle rather than fully penetrate into the yolk itself. Penetrating the follicle would lead to extensive *S.* Enteritidis growth in a nutrient-rich environment resulting in the degeneration of the follicle itself before oviposition could occur (67). This may be the pathology associated with production of granulomas of the ovary, commonly referred to as Pullorum lesions. The oviduct is comprised of five functional regions--the infundibulum, magnum, isthmus, uterus, and vagina--and colonization can occur from either descending from the ovary or ascending from the cloaca. Egg contamination from a colonized magnum is believed to result in *S.* Enteritidis in the albumen, while contamination from the isthmus will result in the pathogen being localized to the eggshell membranes. Contamination from the vagina is more likely to occur as eggshell penetration by negative pressure from the temperature differential rather than being originally inside the preformed egg (40).

Even with colonization of the reproductive tract, contamination of the egg is sporadic at best due to stringent selection pressures in the reproductive tract (48, 82). When contamination does occur, the number of *S.* Enteritidis cells in the fresh laid egg is low. Growth of the cells in the egg is influenced by time and temperature of storage after production. Growth in the

albumin is inhibited because it contains ovotransferrin, which limits the amount of available iron, and also contains lysozyme and at least eleven types of defensins, which bind to LPS. It has been shown in one animal infection study that one-third of preformed eggs were *S.* Enteritidis-positive compared with 0.6% of the freshly laid eggs, corroborating that antibacterial factors within the albumen can exert some control over contamination (40). During storage, the vitelline membrane that separates the yolk from the albumen eventually breaks down. Mixing of the yolk and albumen creates a more complete growth media with greater availability of carbon and nitrogen sources where rapid bacterial growth proceeds (62). Storage at ambient temperatures shortens the time for rapid growth to occur, although high numbers of cells are not needed to infect humans. Epidemiological data from a multi-state outbreak involving under-processed eggs in ice-cream showed that no more than 28 cells of *S.* Enteritidis are required to cause infection (111).

Salmonella Enteritidis Phage Type

A further distinguishing typing method for *Salmonella enterica* serotype Enteritidis is phage typing. Phage typing of *S.* Enteritidis is an internationally recognized epidemiological method differentiating isolates on the basis of their lysis patterns and characteristics by a specified scheme of bacteriophages (33, 113). Although there are over 30 phage types in the scheme (58), only a few appear to dominate in *S.* Enteritidis outbreaks. In Western Europe, *S.* Enteritidis phage type (PT) 4 is responsible for most outbreaks, comprising at least 50% of isolated *S.* Enteritidis (89). In the U.S., the dominant phage types are geographically separate. In the Northeast and South, PT8 is the primary phage type followed closely by PT13A. However, PT4 dominates in the Western region since 1993, when *S.* Enteritidis PT4 became the dominant

phage type in the U.S. (87). The reason why PT4 is correlated with higher incidence of human illness remains unknown (62).

The phage type system, developed by Ward et al., uses 16 bacteriophages that target specific epitopes on the bacterial cell surface. Characterization of the *S.* Enteritidis typing phages is marginal and still lacks comprehensive receptor-ligand information. The phages originate from lysogenic *Salmonella* isolates, from sewage, or are adapted from one another through multiple passages and belong to three families of bacteriophages: *Myoviridae*, *Siphoviridae*, and *Podoviridae*. Families *Siphoviridae* and *Podoviridea* have been shown to use LPS as an epitope but the *Myoviridae* bacteriophages use an unknown epitode for invasion (33). Without the identification and characterization of all known epitopes, the amount of biological information that phage typing can reveal is limited. Although LPS serves as an epitope for phage typing, it cannot be used to detect microheterogeneity in LPS. Furthermore, phage typing is not sensitive to phenotypic expression of biofilm-forming strains. Phage type designation can also be altered by sudden changes in cell surface components. Loss of O-chain, as seen in rough strains of *S.* Enteritidis, changes an isolates's phage type to PT7. Transformation of PT4 with a naturally isolated ampicillin resistance plasmid changes the phage type to PT6A, which occurs most likely due to a change in the LPS structure as the plasmid transformation results in a loss of complement resistance. Formation of a colanic acid capsule, a group IA capsule produced by many bacterial species at non-physiological temperatures to protect cells against desiccation, osmotic stress, and oxidative stress, results in an un-typeable phage type (49). The preceding examples reflect the limitations of phage typing, because it lacks sensitivity for detecting cell surface heterogeneity.

Nucleic Acid-based Typing Methods

As *S.* Enteritidis PT4 becomes globally dominant, new typing methods to discriminate beyond the level of phage type have become the focus of much research (68). Rather than the epitope-based typing of phage and serotyping, many new methods are nucleotide based and, therefore, are not subject to differences in cellular expression at different stages of growth. Specificity, reproducibility, and discriminatory power are key elements in devising a typing method. Some methods aim to provide high discriminatory power for the purposes of strain tracking in epidemiological investigations (71, 73), while others aim to be rapid, sensitive alternatives to phage typing and serotyping by targeting specific nucleotide sequences (1, 114). Currently, there is no consensus as to which method is most effective in differentiating strains of *Salmonella* (68), though an ideal typing system would confer meaningful biology about the test isolate in addition to discrimination.

Rapid Amplification of Polymorphic DNA. Random amplification of polymorphic DNA (RAPD) is a technique that uses non-specific, short oligonucleotide primers (8 to 12 base pairs) under non-stringent conditions in the polymerase chain reaction (PCR) to generate a DNA fingerprint. The short length of the primer increases the number of binding sites available, thereby, producing multiple bands. Genetic differences between different strains will have different binding sites, resulting in a different pattern of bands. The advantage of this technique is that no prior knowledge of the target organism's genome is needed since the primer is randomly designed. However, due to non-stringent conditions, reproducibility with this technique is poor and artifact bands are quite common as irreproducible patterns between different labs and even different PCR machines have been observed (68). Increasing stringency conditions decreases the discriminatory power, so optimization of the technique that creates a

balance between reliability and discrimination can be difficult. Nonetheless, some studies have successfully used the RAPD technique to discriminate *Salmonella* strains (73, 83).

Multi-locus Variable-number Tandem-repeats Analysis. Multiple-locus variable-number tandem-repeats analysis (MLVA) is based on repetitive nucleotide sequence motifs of about 10 to 100 base pairs in length (70). These repeated motifs are found at various loci throughout the genome and are subject to strain variation (94). Genome sequence is a prerequisite needed to locate the variable-number tandem-repeat (VNTR) loci. However, whole genome sequencing projects have made this knowledge increasingly available. The technique shows high discriminatory power when used on *S.* Typhimurium isolates (71, 94) and should prove useful in tracking strains during epidemiological investigations. However, high discrimination at the strain-level creates too much background noise when trying to identify markers that correlate with pathogenic potential, such as LPS structure.

Pulsed-field Gel Electrophoresis. Pulsed-field gel electrophoresis (PFGE) is the molecular typing tool used by the CDC for generating fingerprints of *E. coli*, *Listeria monocytogenes*, *Shigella*, and *Salmonella* serotypes (106). The method uses infrequent-cutter DNA restriction endonuclease enzymes, such as *Xba*I and *Bln*I, to digest genomic DNA into large fragments. The fragments are then fractionated by gel electrophoresis by cyclically altering the orientation of the electric field in order to get the best resolution of the large bands. Discrimination between strains depends on nucleotide polymorphisms at the site of DNA restriction (68, 106). The equipment required for PFGE is expensive and the technique itself labor intensive as well (68). While PFGE has been useful for other *Salmonella* serotypes, including *S.* Typhimurium, other studies show that PFGE does not have enough discriminatory power to differentiate *S.* Enteritidis strains (68-69).

Ribotype. Ribotyping is another commonly used technique (68, 79) that is based on the ribosomal RNA (rRNA) operons of bacteria. The operons typically consist of three rRNA genes: 16S, 23S, and 5S rRNA genes, in that order. The number of rRNA operons in the genome varies for bacteria; however, *Salmonella* serotypes contain 7 copies in their genome (56). The nucleotide sequences of these genes are highly conserved, serving as molecular chronometers that can be used convey evolutionary relationships between genera of bacteria. However, the rRNA sequences themselves are often too conservative to use as the basis for discriminating strains (65). However, the flanking and intergenic regions in the rRNA operons can be subject to variation by homologous recombination events (56, 72).

A ribotyping method used by the United Kingdom's Department of Environment, Food, and Rural Affairs (DEFRA) uses two restriction endonuclease enzymes, *Sph*I and *Pst*I, to digest genomic DNA. The choice of enzymes for ribotyping varies, but the two used in combination by DEFRA have shown to provide the most discriminatory power in *S*. Enteritidis (69, 79). The probe is the entire *rrnB* operon cloned into plasmid pBR322; therefore, signals are generated only on fragments that contain all or portions of the rRNA operon. As with PFGE, variations in band patterns are dependent upon nucleotide polymorphisms in the targeted restriction enzyme sites. However, the enzymes chosen are more frequent cutters than the ones used in PFGE analysis and the variable regions localized around the rRNA operons provide more discriminatory power However, the technique is labor-intensive, time consuming, and highly sensitive to user-error (68).

Intergenic Spacer Region. The intergenic spacer regions (ISR) between the rRNA genes in the ribosomal operons have been the subject of studies looking for species-specific nucleotide markers (28, 57, 114). Such markers can be used for rapid, specific methods such as real-time

PCR, as well as evolutionary studies based on the divergent sequences between closely related species (27, 42). ISRs occur in the 16S and 23S genes and the 23S and 5S genes. The high variability observed in these regions are attributed to homologous recombination (54). Most bacterial ISR studies focus on the 16S-23S ISR (46, 64, 96); however, few studies note higher discriminatory power in the 23S-5S ISR (57, 109). Morales et al. used the 23S-5S-*aspU* intergenic spacer regions to assess its ability to discriminate *Salmonella* serotypes and phenotypic variants within *S.* Enteritidis. The region was successful at distinguishing the serotypes, but could not distinguish phage types nor the subpopulation phenovars of *S.* Enteritidis (80).

Salmonella Enteritidis Subpopulation Heterogeneity

An area of research that addresses how *S.* Enteritidis evolves to contaminate eggs focuses on the role of subpopulations of *Salmonella*. Subpopulations are defined as phenotypic variants that can be detected within serotype and phage type. The subpopulation biology of *S.* Enteritidis that impacts egg contamination was first defined as structural changes in the lipopolysaccharide, which is the major surface component of the outer cell membrane of gram-negative bacteria (47). Lipopolysaccharide consists of three sections: the hydrophobic lipid A, the non-repeating core oligosaccharide, and the distal polysaccharide O-chain that is composed of repeating units of sugars. An important function of LPS O-chain in the infection pathway is resistance to the host immune complement system. This is supported by infection studies where strains lacking O-chain are avirulent (50). Three classes of LPS have been described in *S.* Enteritidis. Class 1 lacks the O-chain, is rough, and is avirulent in animal models. Class 2 produces low molecular mass (LMM) LPS where 90% of LPS molecules have an average of five repeating units in the O-chain. Class 3 produces high molecular mass (HMM) LPS where over 50% of LPS molecules

have more than ten repeating O-chain units. The three classes correlate with differences in virulence as seen in animal experiments, with rough strains being non-virulent (49, 93). *S.* Enteritidis strains vary remarkably in their ability to make LMM and HMM LPS. This is in contrast to observations of *S. enterica* serotype Typhimurium strains, which has a homogenous LMM LPS structure (86).

High-dose (10^8 CFU) subcutaneous infections with a HMM LPS *S.* Enteritidis strain in hens result in reproductive tract involution, which provides evidence of specific tropism for the reproductive tract (85). This phenomenon correlates with the findings that isolates producing HMM LPS are more readily found in the egg contents in animal infection studies (43). Lower dose (10^7 CFU) intravenous infection with the same strain did not induce involution of the reproductive tract, but it did result in high incidence egg contamination, defined as 1% or greater contaminated eggs from experimentally infected hens collected over a 21 day period. In addition, the ability of *S.* Enteritidis to produce HMM LPS mitigates signs of illness in hens. The same strain administered orally, however, results in undetectable egg contamination (85). HMM LPS strongly correlates with high-cell density growth (2.5×10^{11} CFU) and spontaneous swarm cell migration, each of which are correlated with enhanced virulence (50-51). Isolates that produce LMM LPS are systemic as they are found in the kidney, liver, spleen, intestines, ceca, and ovary. However, they are found less frequently in the egg as compared to HMM LPS isolates (43). Gantois et al. identified strong induction of the gene *rfbH* when *S.* Enteritidis was inoculated in the egg albumen (41). The product of *rfbH* is responsible for LPS O-antigen biosynthesis and was determined to be necessary for survival in the albumen; thus, correlating with the association of HMM LPS and successful egg contamination.

Another major subpopulation characteristic is the temperature-dependent production of biofilm. Production of biofilm occurs on specific plate media after extended growth at room temperature. The expressed biofilm is characterized as an organic matrix of fimbriae, cellulose, gylcosylated flagella, and LMM LPS. Biofilm is an important factor for survival of gram-negative bacteria in the environment outside of the host. Strains that express this phenotype are more orally invasive than non-biofilm forming strains administered in the same way and they are readily recovered from the spleen and liver. However, biofilm-forming isolates have not been isolated from the egg (47, 52, 105).

Animal experiments that combine the poor orally invasive HMM LPS *S.* Enteritidis strain with the orally invasive biofilm-forming strain and administered by low-dose contact infection result in high incidence of egg contamination. Such a result is not observed when either of the two subpopulations are administered alone (47). It is believed that the animal host places different selection pressures at different points of the infection pathway, ranging from the initial point of infection (oral mucosa) to the end point (the egg). Thus, while the HMM LPS strain is particularly effective at contaminating the egg, it first needs the helper phenotype (biofilm-former) to initially infect the hen and become systemic. There is evidence that some strains have an intermediate phenotype, which may impart some selective advantage (82).

The house mouse, *Mus musculus*, has been identified as a source of subpopulation heterogeneity in the hen house. Subpopulation analysis of *S.* Enteritidis strains isolated from the spleens and intestines of naturally infected mice reveal LPS heterogeneity among the strains as well as the presence of the biofilm-forming phenotype (47, 51). Assessments of risks for egg contamination consistently identified the house mouse as a major contributor to the problem (19). Therefore, it is believed that naturally occurring subpopulation heterogeneity in the hen

house as propagated by the mouse is a critical contributor in the infection pathway that leads to egg contamination.

Morales et al. further characterized subpopulation heterogeneity by assessing phenotype and genotype differences. Phenotype Microarray is a high throughput array that uses 1,920 growth conditions in a single run to study cellular metabolic phenotypes. The 1,920 conditions monitor the respiratory activity under different sources of carbon, nitrogen, sulfur, phosphorous, antimicrobials, and antibiotics (22). Two *S.* Enteritidis subpopulations that differ in LPS structure and their ability to contaminate the egg were found to have significant growth differences in over 20% of the Phenotype Microarray test conditions (82). Genetic variation between the two subpopulations was detected using a two-restriction enzyme ribotype procedure. The four ribotype band differences were traced back to single nucleotide polymorphisms (SNPs) in the RNA operons *rrnA* and *rrnE* but no other meaningful genetic data could be inferred to explain the previous phenotypic differences observed (81).

Porwollik et al. developed a DNA microarray representing 4,442 genes (96.6%) of *S.* Typhimurium strain LT2, 4,348 genes (94.5%) of *S.* Typhi strain CT18 and un-annotated *S.* Enteritidis PT4 specific genes (91). DNA hybridization results using the two aforementioned *S.* Enteritidis subpopulations failed to show any differences in the gene content. However, the PCR amplicons used as probes on the microarray to represent each gene are only partial sequences and not sensitive to subtle nucleotide differences as any sequences with as high as 3% divergence will be called the same. Furthermore, only open reading frames (ORF) are represented on the microarray and, thus, would not be sensitive to any nucleotide differences in non-coding, intergenic regions. Therefore, it is suggested these subpopulations of *S.* Enteritidis

have undergone adaptive radiation evolution via small-scale nucleotide changes in response to the diverse ecological environments in the egg-contamination pathway (82).

Whole-Genome Resequencing Technologies

Identifying and characterizing specific genetic determinants in *S.* Enteritidis subpopulations that vary in their ability to contaminate the egg would further contribute to understanding their role in the egg-contamination pathway. A global nucleic acid survey would be necessary to account for all genetic diversity between subpopulations, but whole-genome sequencing using the gold-standard Sanger dye-terminator sequencing is not cost-effective for most research projects. However, due to advances in technology and reduction in reagent volume, new, lower-cost whole-genome analyses are now accessible. Factors to consider when choosing a next-generation sequencing technology include cost; throughput, accuracy, and completeness of the data.

Hybridization. Sequencing by hybridization is performed on a microarray platform using a high-density oligonucleotide probe set. The oligonucleotide probes are typically 25 base pairs in length and, therefore, have single-nucleotide resolution. The sequence of a related reference genome is required for construction of the ordered array and a minimum of four oligonucleotides to test each base in the target sequence is needed in areas of predicted divergence. Sequence information is derived from hybridization signal ratios discriminating exact matches from mismatches of the test DNA to the microarray probe set. A significant limitation to this resequencing technology is that it is not *de novo* sequencing and, therefore, can only survey genetic data present in the reference genome. Hybridization sequencing also cannot reliably interpret data from large regions of repetitive sequences since the small size of the oligonucleotide probes cannot span the unique flanking regions. However, the shrinking cost of

hybridization sequencing and growing availability of reference genomes still makes this an attractive whole-genome resequencing technology (21).

Roche Nimblegen, Inc. (www.nimblegen.com) offers a comparative genome sequencing service based on hybridization technology divided into two phases (2). In the first phase, mutation mapping, test and reference DNA samples are fluorescently labeled and hybridized to a whole-genome tiling array. Reference versus test sample DNA ratio signals are used to identify regions of divergence within a 7 base pair window. In the second phase, resequencing, a sequencing array is constructed based on the mutation mapping results. The sequencing array consists of high-density probe sets representing a minimum of 4 alternative alleles per identified genomic region of diversity in the mutation mapping phase. Only labeled test DNA is hybridized to the sequencing array and the nucleic acid sequenced is determined by highest signal peak ratios. The resulting sequences are compared to the reference sequence and SNPs are called and categorized as synonymous, non-synonymous, or non-coding.

Sequencing by Synthesis. In sequencing by synthesis technology, nucleic acid data is collected as deoxynucleoside triphosphates (dNTPs) extended from the 3' end of a primer complementary to a DNA template. Data collection at the time of synthesis is in contrast to dye-terminator sequencing, which creates complementary chains of random length terminated by a fluorescently labeled dideoxynucleoside triphosphate that are later resolved by electrophoresis to determine the original template DNA sequence. The data collection of sequence by synthesis technology does not proceed in real-time relative to DNA polymerase's reaction rate, but rather occurs in one base pair reading per cycle (21).

A significant advantage to sequence by synthesis technology is the capability for massive parallelization of sequencing reactions with some systems capable of over 1,000,000 reads in a

single run. The reactions take place in spatially separate locations in a highly parallel array and the technology is well suited for metagenomic studies of heterogeneous samples. A significant limitation to the technology is the short read lengths generated, generally between 25 to 100 base pairs depending on the system. Short read lengths are a challenge for *de novo* sequencing, especially for eukaryotic organisms, since DNA contig assembly accuracy decreases as sequence read lengths become shorter. Genome projects using sequence by synthesis technology often need a reference genome for successful assembly and may have to supplement gaps with dye-terminator sequencing (44). However, short read length assembly tends to be less problematic in prokaryotic systems due to fewer repetitive regions and the massive parallelization can generate over 40 megabases of data in a single run which can provide enough coverage for proper assembly. Pyrosequencing is an example of sequencing by synthesis technology, which is commercially available from 454 Life Sciences (Roche) (21, 76).

Single Molecule Sequencing. As the push for the $1,000 human genome continues, the rate of development of next-generation DNA sequencing technologies has increased dramatically, though very few are commercially available as of 2010 (21, 102). The primary goal of such technologies is to reduce cost by means of reagent volume reduction while increasing throughput and speed. Thus, these new technologies focus on single-molecule sequencing rather than depending on sample amplification to achieve minimum reagent costs. One such example is Pacific Biosciences Single Molecule Real Time (SMRT) DNA sequencing technology (35). The system places a single DNA template molecule with a polymerase in what is a called a zero-mode waveguide which is a hole, tens of nanometers in diameter, fabricated on a metal film providing a detection volume of 20 zeptoliters (10^{-21} liters). Sequence data is collected in real-time as DNA polymerase incorporates fluorescently labeled dNTPs into the

synthesis strand, which allows for sequencing to occur up to 20,000 times faster than pyrosequencing technology. While parallelization in SMRT sequencing is not as high as in pyrosequencing, the read length per reaction has been shown to be 1,000 base pairs or longer which significantly improves whole genome assembly.

Mutagenesis: Mining the Data Storm for Biological Significance

The abundant growth of genomic data as a result of increased whole genome sequencing has led to the prediction of many open reading frames (ORFs) of unknown function. In the published *Salmonella* Enteritidis strain P125109 genome alone, over 30% of the ORFs are of unknown function and a vast majority of ORFs with an annotated function are putative functions (108). One way of characterizing genes of undetermined function or physiological role is by mutagenesis. By altering the wild-type DNA sequence of a target gene, one may be able to study an altered phenotypic change under specific environmental conditions. Mutagenesis can be performed randomly across an entire microbial genome and screened for a desired phenotype or a specific gene can be targeted via site-directed mutagenesis (29, 101).

Transposon-Mediated Mutagenesis. Transposons are mobile genetic elements that can move to another position in the genome in a processed called transposition (29). Different classes of transposons vary in their structural arrangements that govern their mobility, though one of the first transposons to be characterized and widely used as a molecular tool is Tn5, a Class I transposon. Tn5 is a 5,280 base pair element consisting of two inverted repeat sequences of 1,534 base pairs flanking three antibiotic resistance genes. The flanking, inverted repeat insertion sequences (IS) contain genes responsible for transposition. Tn5 has a low insertional specificity and high frequency of transposition; therefore, it can be used to generate random insertion knockout mutations on a genome-wide scale. To create Tn5 insertion mutations in the

genome of a bacterium, the transposon must be introduced into the host cell through an unstable or suicide vector. The vector can either be a bacteriophage or plasmid, depending on compatibility with the host cell. After introduction of the Tn5 transposon into the host cell, insertion mutants can be selected by screen on an appropriate antibiotic media (95).

Red-Mediated Recombination. The Red recombinase system developed by Datsenko and Wanner, an example of site-directed mutagenesis, utilizes plasmid-borne phage genes to facilitate the homologous recombination of a target gene sequence with a desired knockout sequence. Three λ phage genes, γ, β, and *exo* that encode for Gam, Bet, and Exo, respectively, were cloned into a helper plasmid under the control of an inducible arabinose promoter. Gam is responsible for inhibiting host exonuclease activity that would otherwise degrade foreign DNA, while Bet and Exo are responsible for promoting the recombination of the knockout cassette into the host genome. The knockout cassette is created by using a plasmid, containing the antibiotic resistance gene sequence, as template DNA in a PCR reaction with primers designed to have homology to a target sequence in the host genome. The resulting product is transformed into a host cell carrying the helper plasmid with the induced recombinase genes. Genetic material is exchanged, resulting in a knockout mutation with a selectable antibiotic resistance marker (31).

The disruption of a gene by knockout mutagenesis may have polar effects, particularly if the gene is part of an operon transcribed in the same mRNA strand. The Red recombinase system has a feature in place to overcome any undesired downstream effects. The antibiotic resistance cassette inserted into the host genome contains flippase recognition target (FRT) sites flanking the resistance gene. A helper plasmid containing a flippase gene can be transformed into the mutant cell and express flippase to specifically excise the sequence within the FRT sites. The resulting "scar" nucleotide sequence contains stop codons in all six reading frames and a

ribosome binding site to resume any downstream gene expression (31). All helper plasmids in the Red recombinase system are based on temperature-sensitive replicons and, therefore, can be simply removed from the host cell.

The *S.* Enteritidis Challenge: Correlating Genotype to Phenotype

The question of how *Salmonella* Enteritidis is the only human pathogen to contaminate the internal contents of intact eggs, resulting in outbreaks on a global scale, still remains unanswered. Several studies have focused on locating genes unique to *S.* Enteritidis that are responsible for survival in egg contents (40), but the egg contamination pathway involves overcoming multiple hurdles in varying environments. Thus, the infection pathway is likely to be more complex than can be explained by a unique-gene theory. Identifying and understanding the genetic determinants of subpopulation biology in *S.* Enteritidis on a genome-wide level may offer insight on how it contaminates the internal contents of eggs.

References

1. **Agron, P. G., R. L. Walker, H. Kinde, S. J. Sawyer, D. C. Hayes, J. Wollard, and G. L. Andersen.** 2001. Identification by subtractive hybridization of sequences specific for *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **67:**4984-91.

2. **Albert, T. J., D. Dailidiene, G. Dailide, J. E. Norton, A. Kalia, T. A. Richmond, M. Molla, J. Singh, R. D. Green, and D. E. Berg.** 2005. Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*. Nat Methods **2:**951-3.

3. **Anonymous.** 2009. Multistate outbreak of *Salmonella* infections associated with peanut butter and peanut butter-containing products--United States, 2008-2009. MMWR Morb Mortal Wkly Rep **58:**85-90.

4. **Anonymous.** 2007. Multistate outbreak of *Salmonella* serotype Tennessee infections associated with peanut butter--United States, 2006-2007. MMWR Morb Mortal Wkly Rep **56:**521-4.

5. **Anonymous.** 2004. Outbreak of *Salmonella* serotype Enteritidis infections associated with raw almonds--United States and Canada, 2003-2004. MMWR Morb Mortal Wkly Rep **53:**484-7.

6. **Anonymous.** 2002. Outbreak of *Salmonella* serotype Kottbus infections associated with eating alfalfa sprouts--Arizona, California, Colorado, and New Mexico, February-April 2001. MMWR Morb Mortal Wkly Rep **51:**7-9.

7. **Anonymous.** 1999. Outbreak of *Salmonella* serotype Muenchen infections associated with unpasteurized orange juice--United States and Canada, June 1999. MMWR Morb Mortal Wkly Rep **48:**582-5.

8. **Anonymous.** 2008. Outbreak of *Salmonella* serotype Saintpaul infections associated with multiple raw produce items--United States, 2008. MMWR Morb Mortal Wkly Rep **57:**929-34.

9. **Anonymous.** 2002. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2001. MMWR Morb Mortal Wkly Rep **51:**325-9.

10. **Anonymous.** 2003. Preliminary FoodNet data on the incidence of foodborne illnesses--selected sites, United States, 2002. MMWR Morb Mortal Wkly Rep **52:**340-3.

11. **Anonymous.** 2005. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 sites, United States, 2004. MMWR Morb Mortal Wkly Rep **54:**352-6.

12. **Anonymous.** 2007. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2006. MMWR Morb Mortal Wkly Rep **56:**336-9.

13. **Anonymous.** 2008. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2007. MMWR Morb Mortal Wkly Rep **57:**366-70.

14. **Anonymous.** 2009. Preliminary FoodNet Data on the incidence of infection with pathogens transmitted commonly through food--10 States, 2008. MMWR Morb Mortal Wkly Rep **58:**333-7.

15. **Anonymous.** 2006. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 States, United States, 2005. MMWR Morb Mortal Wkly Rep **55:**392-5.

16. **Anonymous.** 2004. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--selected sites, United States, 2003. MMWR Morb Mortal Wkly Rep **53:**338-43.

17. **Anonymous.** 2010. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food - 10 states, 2009. MMWR Morb Mortal Wkly Rep **59:**418-22.

18. **Anonymous.** 2010. The Community Summary Report on Trends and Sources of Zooneses, Zoonotic Agents and Food-borne Outbreaks in the European Union in 2008. European Food Safety Authority **1496**.

19. **Anonymous (ed.).** 2002, posting date. Risk assessments of *Salmonella* in eggs and broiler chickens., vol. 2, [Online.] ftp://ftp.fao.org/docrep/fao/005/y4392e/y4392e00.pdf.

20. **Baskerville, A., T. J. Humphrey, R. B. Fitzgeorge, R. W. Cook, H. Chart, B. Rowe, and A. Whitehead.** 1992. Airborne infection of laying hens with *Salmonella* Enteritidis phage type 4. Vet. Rec. **130:**395-8.

21. **Bentley, D. R.** 2006. Whole-genome re-sequencing. Curr Opin Genet Dev **16:**545-52.

22. **Bochner, B. R., P. Gadzinski, and E. Panomitros.** 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. **11:**1246-55.

23. **Brenner, F. W., R. G. Villar, F. J. Angulo, R. Tauxe, and B. Swaminathan.** 2000. *Salmonella* nomenclature. J. Clin. Microbiol. **38:**2465-7.

24. **Brown, E. W., M. K. Mammel, J. E. LeClerc, and T. A. Cebula.** 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. Proc. Natl. Acad. Sci. U. S. A. **100:**15676-81.

25. **Chadfield, M. S., D. J. Brown, S. Aabo, J. P. Christensen, and J. E. Olsen.** 2003. Comparison of intestinal invasion and macrophage response of *Salmonella* Gallinarum and other host-adapted *Salmonella enterica* serovars in the avian host. Vet. Microbiol. **92:**49-64.

26. **Chappell, L., P. Kaiser, P. Barrow, M. A. Jones, C. Johnston, and P. Wigley.** 2009. The immunobiology of avian systemic salmonellosis. Vet Immunol Immunopathol **128:**53-9.

27. **Chen, C. C., L. J. Teng, and T. C. Chang.** 2004. Identification of clinically relevant viridans group streptococci by sequence analysis of the 16S-23S ribosomal DNA spacer region. J. Clin. Microbiol. **42:**2651-7.

28. **Chiu, T. H., T. R. Chen, W. Z. Hwang, and H. Y. Tsen.** 2005. Sequencing of an internal transcribed spacer region of 16S-23S rRNA gene and designing of PCR primers for the detection of *Salmonella* spp. in food. Int. J. Food Microbiol. **97:**259-65.

29. **Choi, K. H., and K. J. Kim.** 2009. Applications of transposon-based gene delivery system in bacteria. J Microbiol Biotechnol **19:**217-28.

30. **Cogan, T. A., and T. J. Humphrey.** 2003. The rise and fall of *Salmonella* Enteritidis in the UK. J. Appl. Microbiol. **94 Suppl:**114S-119S.

31. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A **97:**6640-5.

32. **De Buck, J., F. Van Immerseel, F. Haesebrouck, and R. Ducatelle.** 2004. Colonization of the chicken reproductive tract and egg contamination by *Salmonella*. J Appl Microbiol **97:**233-45.

33. **De Lappe, N., G. Doran, J. O'Connor, C. O'Hare, and M. Cormican.** 2009. Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. J Med Microbiol **58:**86-93.

34. **Ebel, E. D., M. J. David, and J. Mason.** 1992. Occurrence of *Salmonella* Enteritidis in the U.S. commercial egg industry: report on a national spent hen survey. Avian Dis. **36:**646-54.

35. **Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, and S. Turner.** 2009. Real-time DNA sequencing from single polymerase molecules. Science **323:**133-8.

36. **Fierer, J., and D. G. Guiney.** 2001. Diverse virulence traits underlying different clinical outcomes of *Salmonella* infection. J. Clin. Invest. **107:**775-80.

37. **Finlay, B. B., and J. H. Brumell.** 2000. *Salmonella* interactions with host cells: in vitro to in vivo. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **355:**623-31.

38. **Fisher, I. S.** 1999. *Salmonella* Enteritidis in Western Europe 1995-98 - a surveillance report from Enter-net. Euro. Surveill. **4:**56.

39. **Fullerton, K.** 2008. Monitoring the incidence and causes of diseases potentially transmitted by food in Australia: annual report of the OzFoodNet Network, 2007. Commun Dis Intell **32:**400-24.

40. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, T. J. Humphrey, and F. Van Immerseel.** 2009. Mechanisms of egg contamination by *Salmonella* Enteritidis. FEMS Microbiol Rev **33:**718-38.

41. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, and F. Van Immerseel.** 2009. The *Salmonella* Enteritidis lipopolysaccharide biosynthesis gene *rfbH* is required for survival in egg albumen. Zoonoses Public Health **56:**145-9.

42. **Garcia-Martinez, J., S. G. Acinas, A. I. Anton, and F. Rodriguez-Valera.** 1999. Use of the 16S--23S ribosomal genes spacer region in studies of prokaryotic diversity. J. Microbiol. Methods **36:**55-64.

43. **Gast, R. K., J. Guard-Bouldin, and P. S. Holt.** 2004. Colonization of reproductive organs and internal contamination of eggs after experimental infection of laying hens with *Salmonella* Heidelberg and *Salmonella* Enteritidis. Avian Dis. **48:**863-9.

44. **Gilmour, M. W., M. Graham, G. Van Domselaar, S. Tyler, H. Kent, K. M. Trout-Yakel, O. Larios, V. Allen, B. Lee, and C. Nadon.** 2010. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics **11:**120.

45. **Golden, N. J., H. H. Marks, M. E. Coleman, C. M. Schroeder, N. E. Bauer, Jr., and W. D. Schlosser.** 2008. Review of induced molting by feed removal and contamination of eggs with *Salmonella enterica* serovar Enteritidis. Vet Microbiol **131:**215-28.

46. **Gonzalez, N., J. Romero, and R. T. Espejo.** 2003. Comprehensive detection of bacterial populations by PCR amplification of the 16S-23S rRNA spacer region. J. Microbiol. Methods **55:**91-7.

47. **Guard-Bouldin, J., R. K. Gast, T. J. Humphrey, D. J. Henzler, C. Morales, and K. Coles.** 2004. Subpopulation characteristics of egg-contaminating *Salmonella enterica* serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl. Environ. Microbiol. **70:**2756-63.

48. **Guard-Petter, J.** 2001. The chicken, the egg and *Salmonella* Enteritidis. Environ. Microbiol. **3:**421-30.

49. **Guard-Petter, J.** 1999. Phage type and other outer-membrane characteristics of *Salmonella enterica* serovar Enteritidis associated with virulence, p. 221-232. *In* A.M. Saeed, R.K Gast, M.E Potter, P.G. Wall (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control., 1st ed. Iowa State University, Ames.

50. **Guard-Petter, J.** 1998. Variants of smooth *Salmonella enterica* serovar Enteritidis that grow to higher cell density than the wild type are more virulent. Appl. Environ. Microbiol. **64:**2166-72.

51. **Guard-Petter, J., D. J. Henzler, M. M. Rahman, and R. W. Carlson.** 1997. On-farm monitoring of mouse-invasive *Salmonella enterica* serovar Enteritidis and a model for its association with the production of contaminated eggs. Appl. Environ. Microbiol. **63:**1588-93.

52. **Guard-Petter, J., L. H. Keller, M. M. Rahman, R. W. Carlson, and S. Silvers.** 1996. A novel relationship between O-antigen variation, matrix formation, and invasiveness of *Salmonella* Enteritidis. Epidemiol. Infect. **117:**219-31.

53. **Guibourdenche, M., P. Roggentin, M. Mikoleit, P. I. Fields, J. Bockemuhl, P. A. Grimont, and F. X. Weill.** 2010. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. Res Microbiol **161:**26-9.

54. **Gurtler, V.** 1999. The role of recombination and mutation in 16S-23S rDNA spacer rearrangements. Gene **238:**241-52.

55. **Heinitz, M. L., R. D. Ruble, D. E. Wagner, and S. R. Tatini.** 2000. Incidence of *Salmonella* in fish and seafood. J Food Prot **63:**579-92.

56. **Helm, R. A., S. Porwollik, A. E. Stanley, S. Maloy, M. McClelland, W. Rabsch, and A. Eisenstark.** 2004. Pigeon-associated strains of *Salmonella enterica* serovar Typhimurium phage type DT2 have genomic rearrangements at rRNA operons. Infect. Immun. **72:**7338-41.

57. **Herpers, B. L., B. M. de Jongh, K. van der Zwaluw, and E. J. van Hannen.** 2003. Real-time PCR assay targets the 23S-5S spacer for direct detection and differentiation of *Legionella* spp. and *Legionella pneumophila*. J. Clin. Microbiol. **41:**4815-6.

58. **Hickman-Brenner, F. W., A. D. Stubbs, and J. J. Farmer, 3rd.** 1991. Phage typing of *Salmonella* Enteritidis in the United States. J. Clin. Microbiol. **29:**2817-23.

59. **Holt, P. S.** 2003. Molting and *Salmonella enterica* serovar Enteritidis infection: the problem and some solutions. Poult. Sci. **82:**1008-10.

60. **Hope, B. K., R. Baker, E. D. Edel, A. T. Hogue, W. D. Schlosser, R. Whiting, R. M. McDowell, and R. A. Morales.** 2002. An overview of the *Salmonella* Enteritidis risk assessment for shell eggs and egg products. Risk Anal. **22:**203-18.

61. **Hughes, C., I. A. Gillespie, and S. J. O'Brien.** 2007. Foodborne transmission of infectious intestinal disease in England and Wales, 1992-2003. Food Control **18:**766-772.

62. **Humphrey, T. J.** 1999. Contamination of eggs and poultry meat with *Salmonella enterica* serovar Enteritidis, p. 183-192. *In* A.M. Saeed, Gast, R.K., Potter, M.E., Wall, P.G. (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control., 1st ed. Iowa State University Press, Ames.

63. **Ibarra, J. A., and O. Steele-Mortimer.** 2009. *Salmonella*--the ultimate insider. *Salmonella* virulence factors that modulate intracellular survival. Cell Microbiol **11:**1579-86.

64. **Iteman, I., R. Rippka, N. Tandeau De Marsac, and M. Herdman.** 2000. Comparison of conserved structural and regulatory domains within divergent 16S rRNA-23S rRNA spacer sequences of cyanobacteria. Microbiology **146 (Pt 6):**1275-86.

65. **Jaspers, E., and J. Overmann.** 2004. Ecological significance of microdiversity: identical 16S rRNA gene sequences can be found in bacteria with highly divergent genomes and ecophysiologies. Appl. Environ. Microbiol. **70:**4831-9.

66. **Kinde, H., D. M. Castellan, P. H. Kass, A. Ardans, G. Cutler, R. E. Breitmeyer, D. D. Bell, R. A. Ernst, D. C. Kerr, H. E. Little, D. Willoughby, H. P. Riemann, J. A. Snowdon, and D. R. Kuney.** 2004. The occurrence and distribution of *Salmonella* Enteritidis and other serovars on California egg laying premises: a comparison of two sampling methods and two culturing techniques. Avian Dis **48:**590-4.

67. **Kinde, H., H. L. Shivaprasad, B. M. Daft, D. H. Read, A. Ardans, R. Breitmeyer, G. Rajashekara, K. V. Nagaraja, and I. A. Gardner.** 2000. Pathologic and bacteriologic findings in 27-week-old commercial laying hens experimentally infected with *Salmonella* Enteritidis, phage type 4. Avian Dis **44:**239-48.

68. **Liebana, E.** 2002. Molecular tools for epidemiological investigations of *S. enterica* subspecies *enterica* infections. Res. Vet. Sci. **72:**169-75.

69. **Liebana, E., L. Garcia-Migura, M. F. Breslin, R. H. Davies, and M. J. Woodward.** 2001. Diversity of strains of *Salmonella enterica* serotype Enteritidis from English poultry farms assessed by multiple genetic fingerprinting. J. Clin. Microbiol. **39:**154-61.

70. **Lindstedt, B. A., E. Heir, E. Gjernes, and G. Kapperud.** 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J. Clin. Microbiol. **41:**1469-79.

71. **Lindstedt, B. A., T. Vardund, L. Aas, and G. Kapperud.** 2004. Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. J. Microbiol. Methods **59:**163-72.

72. **Liu, S. L., and K. E. Sanderson.** 1998. Homologous recombination between rrn operons rearranges the chromosome in host-specialized species of *Salmonella*. FEMS Microbiol. Lett. **164:**275-81.

73. **Lopes, V. C., B. T. Velayudhan, D. A. Halvorson, D. C. Lauer, R. K. Gast, and K. V. Nagaraja.** 2004. Comparison of methods for differentiation of *Salmonella enterica* serovar Enteritidis phage type 4 isolates. Am. J. Vet. Res. **65:**538-43.

74. **Lynch, M., J. Painter, R. Woodruff, and C. Braden.** 2006. Surveillance for foodborne-disease outbreaks--United States, 1998-2002. MMWR Surveill Summ **55:**1-42.

75. **Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra.** 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50:**882-9.

76. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P.**

Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376-80.

77. **Mazurek, J., L. Holbert, M. K. Parrish, and E. Salehi.** 2005. Raw eggs--lessons learned from an outbreak of *Salmonella* serotype Enteritidis infections associated with meringue pie. J. Public Health Manag. Pract. **11:**201-7.

78. **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:**607-25.

79. **Mendoza, M. C. a. E. L.** 1999. Molecular epidemiology methods for differentiation of *Salmonella enterica* serovar Enteritidis strains, p. 125-140. *In* A.M. Saeed, R.K Gast, M.E Potter, P.G. Wall (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control, 1st ed. Iowa State University Press, Ames.

80. **Morales, C. A., R. Gast, and J. Guard-Bouldin.** 2006. Linkage of avian and reproductive tract tropism with sequence divergence adjacent to the 5S ribosomal subunit rrfH of *Salmonella enterica*. FEMS Microbiol Lett **264:**48-58.

81. **Morales, C. A., M. Musgrove, T. J. Humphrey, C. Cates, R. Gast, and J. Guard-Bouldin.** 2007. Pathotyping of *Salmonella enterica* by analysis of single-nucleotide polymorphisms in cyaA and flanking 23S ribosomal sequences. Environ Microbiol **9:**1047-59.

82. **Morales, C. A., S. Porwollik, J. G. Frye, H. Kinde, M. McClelland, and J. Guard-Bouldin.** 2005. Correlation of phenotype with the genotype of egg-contaminating *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **71:**4388-99.

83. **Muresu, E., A. Piana, A. Azara, I. Maida, A. Nastasi, S. D. Sajid, and S. Rubino.** 2001. Clonal relations among *Salmonella* Enteritidis phage type 3 outbreak isolates traced by DNA fingerprinting. New Microbiol. **24:**371-7.

84. **Olsen, S. J., L. C. MacKinnon, J. S. Goulding, N. H. Bean, and L. Slutsker.** 2000. Surveillance for foodborne-disease outbreaks--United States, 1993-1997. MMWR CDC Surveill Summ **49:**1-62.

85. **Parker, C. T., B. Harmon, and J. Guard-Petter.** 2002. Mitigation of avian reproductive tract function by *Salmonella* Enteritidis producing high-molecular-mass lipopolysaccharide. Environ. Microbiol. **4:**538-45.

86. **Parker, C. T., E. Liebana, D. J. Henzler, and J. Guard-Petter.** 2001. Lipopolysaccharide O-chain microheterogeneity of *Salmonella* serotypes Enteritidis and Typhimurium. Environ. Microbiol. **3:**332-42.

87. **Patrick, M. E., P. M. Adcock, T. M. Gomez, S. F. Altekruse, B. H. Holland, R. V. Tauxe, and D. L. Swerdlow.** 2004. *Salmonella* Enteritidis infections, United States, 1985-1999. Emerg. Infect. Dis. **10:**1-7.

88. **Penteado, A. L., B. S. Eblen, and A. J. Miller.** 2004. Evidence of *Salmonella* internalization into fresh mangos during simulated postharvest insect disinfestation procedures. J. Food Prot. **67:**181-4.

89. **Poppe, C.** 1999. Epidemiology of *Salmonella enterica* serovar Enteritidis, p. 3-18. *In* A.M. Saeed, R.K Gast, M.E Potter, P.G. Wall (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control, 1st ed. Iowa State University Press, Ames.

90. **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:**5883-98.

91. **Porwollik, S., J. Frye, L. D. Florea, F. Blackmer, and M. McClelland.** 2003. A non-redundant microarray of genes for two related bacteria. Nucleic Acids Res. **31:**1869-76.

92. **Rabsch, W., B. M. Hargis, R. M. Tsolis, R. A. Kingsley, K. H. Hinz, H. Tschape, and A. J. Baumler.** 2000. Competitive exclusion of Salmonella enteritidis by *Salmonella* Gallinarum in poultry. Emerg Infect Dis **6:**443-8.

93. **Rahman, M. M., J. Guard-Petter, and R. W. Carlson.** 1997. A virulent isolate of *Salmonella* Enteritidis produces a *Salmonella* Typhi-like lipopolysaccharide. J. Bacteriol. **179:**2126-31.

94. **Ramisse, V., P. Houssu, E. Hernandez, F. Denoeud, V. Hilaire, O. Lisanti, F. Ramisse, J. D. Cavallo, and G. Vergnaud.** 2004. Variable number of tandem repeats in *Salmonella enterica* subsp. *enterica* for typing purposes. J. Clin. Microbiol. **42:**5722-30.

95. **Rossbach, S., and F. J. d. Bruijn.** 2007. Transposon Mutagenesis, p. 684-708. *In* C. A. Reddy (ed.), Methods for General and Molecular Microbiology, 3rd ed. ASM Press, Washington, DC.

96. **Roth, A., M. Fischer, M. E. Hamid, S. Michalke, W. Ludwig, and H. Mauch.** 1998. Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences. J. Clin. Microbiol. **36:**139-47.

97. **Scallan, E., S. E. Majowicz, G. Hall, A. Banerjee, C. L. Bowman, L. Daly, T. Jones, M. D. Kirk, M. Fitzgerald, and F. J. Angulo.** 2005. Prevalence of diarrhoea in the community in Australia, Canada, Ireland, and the United States. Int J Epidemiol **34:**454-60.

98. **Scharff, R. L.** 2010. Health-related Costs from foodborne illness in the United States. Produce Safety Project. **http://www.producesafetyproject.org/reports?id=0008**.

99. **Schroeder, C. M., A. L. Naugle, W. D. Schlosser, A. T. Hogue, F. J. Angulo, J. S. Rose, E. D. Ebel, W. T. Disney, K. G. Holt, and D. P. Goldman.** 2005. Estimate of illnesses from *Salmonella* Enteritidis in eggs, United States, 2000. Emerg Infect Dis **11:**113-5.

100. **Senior, K.** 2009. Estimating the global burden of foodborne disease. Lancet Infect Dis **9:**80-1.

101. **Shen, B.** 2002. PCR approaches to DNA mutagenesis and recombination. An overview. Methods Mol Biol **192:**167-74.

102. **Shendure, J., and H. Ji.** 2008. Next-generation DNA sequencing. Nat Biotechnol **26:**1135-45.

103. **Sivapalasingam, S., C. R. Friedman, L. Cohen, and R. V. Tauxe.** 2004. Fresh produce: a growing cause of outbreaks of foodborne illness in the United States, 1973 through 1997. J Food Prot **67:**2342-53.

104. **Skovgaard, N.** 2007. New trends in emerging pathogens. Int J Food Microbiol **120:**217-24.

105. **Solano, C., B. Garcia, J. Valle, C. Berasain, J. M. Ghigo, C. Gamazo, and I. Lasa.** 2002. Genetic analysis of *Salmonella* Enteritidis biofilm formation: critical role of cellulose. Mol. Microbiol. **43:**793-808.

106. **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:**382-9.

107. **Tauxe, R. V.** 2002. Emerging foodborne pathogens. International Journal of Food Microbiology **78:**31-41.

108. **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:**1624-37.

109. **Tilsala-Timisjarvi, A., and T. Alatossava.** 2001. Characterization of the 16S-23S and 23S-5S rRNA intergenic spacer regions of dairy propionibacteria and their identification with species-specific primers by PCR. Int. J. Food Microbiol. **68:**45-52.

110. **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:**S127-34.

111. **Vought, K. J., and S. R. Tatini.** 1998. *Salmonella* Enteritidis contamination of ice cream associated with a 1994 multistate outbreak. J. Food Prot. **61:**5-10.

112. **Vugia, D. J., M. Samuel, M. M. Farley, R. Marcus, B. Shiferaw, S. Shallow, K. Smith, and F. J. Angulo.** 2004. Invasive *Salmonella* infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. Clin. Infect. Dis. **38 Suppl 3:**S149-56.

113. **Ward, L. R., J. D. de Sa, and B. Rowe.** 1987. A phage-typing scheme for *Salmonella* Enteritidis. Epidemiol. Infect. **99:**291-4.

114. **Zhu, Q., C. K. Lim, and Y. N. Chan.** 1996. Detection of *Salmonella* Typhi by polymerase chain reaction. J. Appl. Bacteriol. **80:**244-51.

TABLE 2.1. *Salmonella* and serotype incidence data from the Center For Disease Control and Prevention Foodborne Disease Active Surveillance Network from 2000-2009a

Year of data collection	Total laboratory confirmed foodborne illnesses processed for causative agent	Number (%) of cases that were confirmed as *Salmonella*	% serotypes reported for *Salmonella* isolates that were serotyped					
			Enteritidis	Typhimurium	Newport	Heidelburg	Javiana	I 4,[5],12:I:-
2009	17,468	7,039 (40.3%)	19.2	16.1	12.1	3.6	8.5	3.1
2008	18,499	7,444 (40.2%)	20.1	16.0	10.1	2.9	6.3	4.0
2007	17,883	6,790 (38.0%)	16.9	16.0	10.4	3.9	5.5	5.7
2006	17,252	6,655 (38.6%)	19.0	19.0	9.0	4.0	5.0	4.0
2005	16,614	6,471 (38.9%)	18.0	19.0	10.0	6.0	5.0	3.0
2004	15,806	6,464 (40.9%)	15.0	20.0	10.0	5.0	7.0	NR
2003	15,600	6,017 (38.6%)	14.0	20.0	12.0	6.0	6.0	NR
2002	16,580	6,028 (36.4%)	15.0	19.0	14.0	NR	NR	NR
2001	13,705	5,198 (37.9%)	15.0	25.0	12.0	7.0	5.0	NR
2000	12,631	4,237 (33.5%)	15.0	23.0	11.0	7.0	NR	NR

[a]See Anonymous(9-16)

CHAPTER 3

WHOLE GENOME COMPARISON FOR DETECTION OF SUBPOPULATION HETEROGENEITY IN *SALMONELLA* ENTERITIDIS [1]

[1] Morales, C.A and J.Y. Guard. To be submitted to *Applied and Environmental Microbiology*

[2] Morales, C.A., J.Y. Guard and M.A. Harrison. To be submitted to *Applied and Environmental*

Abstract

Salmonella enterica subspecies *enterica* serotype Enteritidis is the leading cause of salmonellosis both worldwide and in the United States. Its persistence as a foodborne pathogen is due in part to its ability to mitigate signs of illness in egg-laying hens while successfully contaminating the internal contents of eggs. Subpopulations of *S.* Enteritidis phage type 13a have been shown to vary in their phenotype and in their ability to contaminate the internal contents of the egg. Comparative Genome Sequencing was used to identify nucleotide polymorphisms on a genome-wide level between two subpopulations: isolate 21046, an egg contaminator with no biofilm formation; and 21027, positive for biofilm but negative for egg contamination. Results support the hypothesis that small-scale nucleotide polymorphisms are responsible for diversity between the two subpopulations. A total of 247 polymorphisms, ranging in size from 1 base pair to 215 contiguous base pairs, were identified that differentiate the two subpopulations. Twelve polymorphisms were predicted to disrupt open reading frames, 99 were non-synonymous amino acid substitutions, 82 were synonymous within an open reading frame, 38 polymorphisms were in non-coding intergenic regions, and 12 were in RNA genes. Open reading frame-disrupting and non-synonymous polymorphisms predominantly occurred in genes predicted to be involved in metabolism, though isolate 21027 contained a frame-shift mutation in *sefD*, a putative fimbrial minor subunit gene of the O-antigen Group D *Salmonella*. Creating a comprehensive database of the polymorphisms between the two subpopulations on a genome-wide scale facilitates identifying their role in the infection pathway that results in egg contamination and results in human disease.

Introduction

Non-typhoidal *Salmonella* species can cause self-limiting gastroenteritis in healthy adults and can become systemic and cause additional sequelae in the young, elderly, and immuno-compromised individuals (47). *Salmonella* spp. are the leading cause of acute bacterial gastroenteritis worldwide and are estimated to cause 1.4 million cases of salmonellosis each year in the United States (26). *Salmonella* spp. are also the leading cause of foodborne illness in the United States and, currently *Salmonella enterica* subspecies *enterica* serotype Enteritidis (*S.* Enteritidis) is the serotype reported to be of the highest incidence (2). Incidence of *S.* Enteritidis in the U.S. peaked from the mid 1980s to the mid 1990s with rates reaching approximately 10 cases per 100,000 persons. During this period, 80% of reported *S.* Enteritidis outbreaks with implicated food items were egg-associated (29). The unique ability of *S.* Enteritidis to contaminate and survive within the internal contents of eggs without causing illness in hens has, in part, led to its success as a foodborne pathogen (17, 22).

The egg contamination pathway is complex as it includes other animal vectors, survival in the farm environment, systemic infection of the egg-laying hen, tropism towards the reproductive tract, and then survival and growth in an inhibitory albumen environment (12, 17). It is expected that multiple genes that encode proteins with optimal amino acid content and levels of expression play a role in the overall pathway in contrast to one single virulence gene being responsible for egg contamination (28). Many studies have searched and identified possible genetic determinants that may lead to successful egg contamination. The *rfbH* gene, which encodes a putative dehydratase necessary for lipopolysaccharide (LPS) O-antigen synthesis was found to be necessary for *S.* Enteritidis growth and survival in the egg-white (14). Additionally, the ability to produce high-molecular mass (HMM) LPS has been associated with high-incidence

of egg contamination (16); thus, suggesting an important role for LPS in the egg contamination pathway. Another study found DNA repair genes *yafD* and *xthA* also to be necessary for survival in the albumen (25). Further genes involving amino acid and nucleic acid metabolism, motility, cell wall integrity, and stress response were found to be up-regulated when colonizing the oviduct (13). Apart from survival mechanisms, studies have also looked at *S.* Enteritidis specific fimbriae and their involvement in oviduct tissue attachment (31). However, many of these studies only examine one point in the egg-contamination pathway or focus on comparing the *S.* Enteritidis genetic elements with other serotypes or genera, which is a comparative approach that lacks the stringency required to detect evolution specifically impacting egg contamination.

The role of subpopulation heterogeneity within *S.* Enteritidis in the egg-contamination pathway has emerged as a research area of interest (16, 20, 28). Analysis of the *S.* Enteritidis cell surface has shown that certain strains of different phage types (PT) are able to produce high-molecular mass (HMM) LPS, forming a capsular-like O-chain region. This phenotype correlates well with high-incidence of egg contamination in contrast to strains that produce low-molecular mass (LMM) LPS (16). Another virulence factor known to vary in *S.* Enteritidis, as well as other serotypes, is biofilm formation. Biofilm is an extracellular organic matrix of fimbriae, cellulose, glycosylated flagella, and LPS and is an important factor for survival in the environment outside of the host (20). *Salmonella* Enteritidis strains that are capable of producing biofilm are more orally invasive than non-biofilm forming strains and exhibit strong systemic invasion (16). Field isolation studies have found subpopulations of *S.* Enteritidis together, within the same rodent carrier, which vary in their LPS structure and ability to produce biofilm (17, 19).

Two particular subpopulation isolates of *S.* Enteritidis PT13a have been the subjects of recent studies due to their varying ability to contaminate the egg (16, 20, 28). One subpopulation produces HMM LPS but not biofilm, while the other produces LMM LPS and biofilm but cannot contaminate the egg (16). Subsequent animal infection studies have shown that the two subpopulations can complement each other in the infection pathway as the HMM LPS producing subpopulation is not as efficient at oral invasion and the LMM LPS subpopulation, while virulent, does not appear to contaminate the egg. Low-dose contact infections with single subpopulation inoculums do not result in detectable egg contamination. However, challenging hens with a low-dose mixture of the two *S.* Enteritidis subpopulations reliably results in high-incidence egg-contamination, demonstrating the necessity of multiple virulence factors to complete the egg contamination pathway (15-16).

Genetic analysis of the two *S.* Enteritidis PT13a subpopulations thus far has only revealed single nucleotide polymorphisms (SNPs) in two 23S ribosomal genes, contributing to the different ribotype patterns observed, as well as two SNPs identified in the *cyaA* gene (27). A DNA-DNA microarray hybridization study looking for discrete gene differences between the two subpopulations was unable to identify any difference in gene content, implying that only small-scale nucleotide polymorphisms differentiate their genomes (28). Biolog's Phenotype Microarray technology was able to identify over 380 biochemical conditions that differentiated the growth between the two subpopulations (4, 28). Considering the known phenotypic differences between the two subpopulations in LPS structure, colonial morphology, and metabolic pathways it is unlikely that the few identified SNPs are responsible for all of the observed phenotype differences. Therefore, there is still a need to identify genetic determinants

responsible for subpopulation heterogeneity and to characterize how they contribute to the egg contamination pathway.

The research described herein uses Comparative Genome Sequencing technology to resequence the genomes of the two described *S.* Enteritidis PT13a subpopulations. Using a combination of Nimblegen's mutation mapping custom array and Sanger dye-terminator confirmatory sequencing, the objective is to reveal all nucleotide polymorphisms on a genome-wide scale between the two subpopulations in a first-step approach towards identifying specific genetic determinants that contribute to the ability of *S.* Enteritidis to internally contaminate the egg.

Materials and Methods

Strains used for analysis. Three *Salmonella* Enteritidis isolates were analyzed by Nimblegen's (Madison, WI) Comparative Genome Sequencing technology (CGS). Accession numbers for the isolates are 21046 (phage type 13a), 21027 (phage type 13a), and 22079 (phage type 4). Isolates 21046 and 21027 were originally derived from a single parent culture, SE6, and strains with similar variant phenotypes were later isolated from the spleen and intestines, respectively, of infected farm mice. Therefore, both phenotypes appeared common within the poultry environment. Isolate properties were previously described (16, 28). In brief, isolate 21046 contaminates eggs but does not form biofilm and exhibits poor colonization in White Leghorn hens when infected orally, whereas isolate 21027 forms biofilm and does not contaminate eggs but exhibits good organ colonization in hens when infected orally. Isolate 22079 was used as the control reference strain for CGS and is able to both contaminate eggs and form biofilm. The ability to form biofilm was determined as follows. Cultures were spread on brilliant green agar (Acumedia-Neogen, Lansing, MI, USA) at a density of between 25 and 50

CFU and then incubated for 72 h at ambient temperatures between 23 and 25°C. Colonies that produced a distinctive organic matrix was indicative of biofilm.

DNA isolation for confirmatory sequencing and mutation mapping. Single colonies of *S.* Enteritidis were grown in 10 ml of brain heart infusion broth (Difco BD, Franklin Lakes, NJ) at 37°C for 16 h. Bacterial cells were pelleted in a Sorvall RC5B Plus centrifuge at 5000 x *g* for 15 min in a Sorvall Super-lite SLA 600TC rotor. For confirmatory sequencing, total DNA was extracted using a Qiagen Genomic-tip 100/G kit following the protocol designated for bacteria (Qiagen, Valencia, CA). Precipitated DNA was dissolved in 200 µl of Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8]) and stored at -20°C. For mutation mapping, total DNA was extracted using a Qiagen Genomic-tip 500/G kit following the protocol designated for bacteria. Precipitated DNA was dissolved in 150 µl of Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA [pH 8]) and stored at -20°C. Spectrometer readings using a NanoDrop 1000 (Wilmington, DE) were performed to ensure an $OD_{260/280}$ ratio greater than 1.7 and a genomic DNA concentration of 1 µg/µl as required by Nimblegen for the mutation mapping service.

Comparative Genome Sequencing phase 1: mutation mapping. Comparative Genome Sequencing is comprised of two phases, the first being mutation mapping. The protocol for mutation mapping was previously described (1). In brief, purified whole DNA from isolates 21046, 21027, and 22079 were submitted to Nimblegen for mutation mapping analysis. A custom high-density oligonucleotide microarray was constructed based on the sequenced *S.* Enteritidis PT4 genome (GenBank accession: AM933172). Oligonucleotide probes were designed in a high-overlap, tile-array format, each with a length of 29 to 39 base pairs (bp) and designed from every 7 bp window for both strands along the 4.68 Mbp genome. The small length of the probes enables the array to be sensitive to single nucleotide polymorphisms (SNPs)

as well as other small or large-scale nucleotide polymorphisms. Two separate hybridization experiments were performed where each test strain, 21046 and 21027, was hybridized against the reference strain 22079. Test and reference samples were independently labeled with fluorescent dyes, Cy3 and Cy5 respectively, hybridized against the custom mutation mapping array, and scanned for signal intensities. Raw data reporting signal ratios for every probe and predicted regions of nucleotide difference between the test and reference samples was returned in a tab-delimited text file format.

Comparative Genome Sequencing phase 2: confirmatory sequencing. Sanger dye-terminator sequencing was used to confirm and identify polymorphisms predicted by the mutation mapping phase. Probe locations on the mutation map array were based on the sequenced *S.* Enteritidis PT4 genome. The 300 bp sequence upstream and downstream (600 bp total) of all probe positions predicted to contain nucleotide polymorphisms between *S.* Enteritidis PT13a subpopulations 21027 and 21046 were used to design primers for confirmatory sequencing. Primers were designed using Integrated DNA Technologies SciTools PrimerQuest program (http://www.idtdna.com/Scitools). A total of 434 primer pairs were designed for confirmatory sequencing. The remainder primer pairs that did not confirm a polymorphism are not reported. Polymerase chain reactions (PCR) performed on all three isolates included 100 nM of each primer and 1X concentration of Applied Biosystems (Carlsbad, CA) GeneAmp Fast PCR Master Mix in a final reaction volume of 20 µl. The PCR reactions were performed in an Applied Biosystems Veriti 96 Well Fast Thermal Cylcer system using annealing temperatures calculated from PimerQuest and default cycle conditions for Applied Biosystem's Fast PCR protocol. Single-band, amplified PCR products of expected size were confirmed by applying an unspecified, proprietary voltage on Invitrogen precast 2% agarose E-Gel48 gels (Carlsbad, CA)

for 20 minutes on an Invitrogen E-Gel electrophoresis system and visualizing under ultraviolet light of 302 nm wavelength. Confirmed PCR products were purified using a Qiagen QIAquick PCR Purification kit and submitted with primers to Retrogen Inc (San Diego, California) for dye-terminator sequencing.

Mutation mapping data analysis. Raw data from the mutation mapping experiment was imported into Microsoft Office Excel 2003 spreadsheet program (Redmond, Washington) for ease of sorting the data. For each probe on the array, signal ratio values between the test and reference samples within the probe sequence were returned. A signal ratio value significantly greater than 1 indicated the presence of a polymorphism within the probe between the test and reference samples. Probes with a signal ratio value that was not significantly greater than one did not indicate the presence of a polymorphism and were filtered out of the data set. Since *S.* Enteritidis subpopulations 21027 and 21046 could not be directly compared on the same mutation mapping array, predicted differences between the two subpopulations were indirectly inferred by comparing their respective differences with reference isolate *S.* Enteritidis PT4 22079. If 21027 and 21046 had a predicted polymorphism compared to 22079 at the same probe, it was not considered to be a polymorphism between the two subpopulations but rather an *S.* Enteritidis PT13a versus PT4 polymorphism. Therefore, only probes with a difference value of 1 and that did not share the same predicted polymorphism compared to the reference 22079 strain were considered likely polymorphisms between the two subpopulations of *S.* Enteritidis.

Confirmatory sequencing data analysis. Returned sequence data of the three isolates from each primer set was aligned with the *S.* Enteritidis PT4 genome sequence using Gene Codes Sequencher version 4.8 alignment software (Ann Arbor, Michigan). Polymorphisms were confirmed to be either present or undetected by 2X sequence coverage for each of these isolates.

Polymorphism locations respective to predicted open reading frames (ORFs) were based on the annotated *S.* Enteritidis PT4 genome sequence. Any polymorphisms located outside of an ORF were designated as intergenic. Polymorphisms identified within an ORF were further analyzed for potential amino acid sequence changes. New ORF sequences representing 21046 and 21027 with their respective polymorphisms were aligned and translated using DNASTAR Lasergene MegAlign version 8.0 to identify amino acid changes. Polymorphisms were then designated as synonymous for no amino acid change, non-synonymous if there was an amino acid change, and ORF-disrupting if there was a frame-shift deletion or terminating codon introduced by the polymorphism. Functional information on ORF translated products was obtained from the *S.* Enteritidis PT4 genome annotations and Basic Local Alignment Search Tool (http://blast.ncbi.nlm.nih.gov) (45). A circular, genomic map allowing the visualization of the polymorphisms along the *S.* Enteritidis PT4 genome was created using DNAPlotter software (6).

Results

Mutation mapping results. The mutation mapping array was constructed using overlapping oligonucleotide probes and, therefore, multiple adjacent probes could produce a signal to the same polymorphism. A single region of polymorphism as predicted by mutation mapping is defined as one or more contiguous probes predicting a polymorphism uninterrupted by a probe not predicting a polymorphism. The number of contiguous probes predicting polymorphism was often proportional to the size of the polymorphism. Thus, a region of polymorphism ranged in size from a single nucleotide polymorphism to a 215 bp deletion.

A total of 439 regions were predicted by mutation mapping to contain nucleotide polymorphisms between the two *S.* Enteritidis PT13a subpopulations 21027 and 21406. Isolate 21046 was predicted to have 164 regions polymorphisms in the chromosome and 10

polymorphisms in the large virulence plasmid relative to the reference *S.* Enteritidis PT4 22079 strain. Isolate 21027 was predicted to have 234 regions of chromosomal polymorphisms and 31 polymorphism regions in the large virulence plasmid relative to the reference PT4 strain.

The limitation of Nimblegen's mutation mapping technology is that it is unable to describe polymorphism details. The oligonucleotide probes are sensitive to SNPs, but a positive signal predicting a polymorphism only suggests whether the polymorphisms is likely to be large or small and only approximates where the polymorphism is located along the reference genome. Nimblegen offers a secondary service, Resequencing, which creates a custom microarray based from the mutation mapping results with oligonucleotide probes. It includes all possible nucleotide positions at a predicted polymorphism region. The probe with the highest signal upon hybridization with a test sample would represent the true nucleic acid sequence. This service was originally performed along with mutation mapping for this study. However, results included a high frequency of false-negatives (data not shown), including SNPs that were previously reported between the two subpopulations (27). Therefore, Sanger method dye-terminator sequencing (40) was instead used as the second phase of Comparative Genomic Sequencing to confirm and identify polymorphisms predicted by mutation mapping.

Confirmatory Sequencing. A summary of the overall results of confirmatory sequencing can be seen in Table 3.1. In the context of confirmatory sequencing, a polymorphism is defined as one or more contiguous base pairs and ranges in size from a SNP to a 215 bp deletion. In total, 247 polymorphisms were discovered between two subpopulations of *S.* Enteritidis that differ in their ability to contaminate the egg. Relative to the *S.* Enteritidis PT4 reference strain, subpopulation isolate 21046 contained 111 unique polymorphisms. The polymorphisms consisted of two nucleotide deletion events, two termination codons introduced

inside an ORF, 44 non-synonymous mutations resulting in amino acids changes, 8 polymorphisms in RNA genes, 16 polymorphisms in non-coding regions, and 39 DNA substitutions within an ORF but did not change the predicted amino acid sequence. Isolate 21046 also contained the only discovered virulence plasmid polymorphisms, which were 3 intergenic SNPs. *S.* Enteritidis PT13a subpopulation isolate 21027 contained 136 unique polymorphisms relative to the reference strain. The polymorphisms included 8 deletion events, three terminating codons introduced within an ORF, one terminating codon removed from an ORF, 58 amino acid substitutions, four polymorphisms within RNA genes, 22 intergenic polymorphisms, and 43 synonymous mutations within a gene. Visualization of the polymorphisms relative to their position on the PT4 reference genome can be seen in Figure 3.1. All confirmed polymorphisms were submitted to the National Center for Biotechnology Information (NCBI) dbSNP database (http://www.ncbi.nlm.nih.gov/SNP) and assigned a dbSNP ID number (Tables 3.2-3.11). A master list of all the polymorphisms was published online and can be found at http://www.ncbi.nlm.nih.gov/genomes/static/Salmonella_SNPS.html.

Polymorphisms that disrupt open reading frames. Confirmatory sequencing revealed a total of 12 polymorphisms between *S.* Enteritidis PT13a subpopulations 21046 and 21027 that interrupted open reading frames either by frame shift deletion, introduction of a termination codon within the ORF, or removal of a termination codon. Tables 3.2 and 3.3 detail the ORF-disrupting polymorphisms for isolates 21046 and 21027, respectively.

S. *Enteritidis PT13a isolate 21046 ORF-disrupting polymorphisms*. The PT13a 21046 isolate contained three polymorphisms that appeared to disrupt an ORF. The gene *fhuA* encodes a 747 amino acid protein described as a ferrichrome-iron receptor and transporter located in the outer membrane of the cell. It also serves as a target receptor for bacteriophages and certain

antimicrobials (5). *S.* Enteritidis isolate 21046 contains a SNP that introduces an early termination codon at glutamine residue 470. Such a truncation of the protein would likely affect the isolate's ability to thrive in an iron-limiting environment. Experimental mutations of *fhuA* performed in *Escherichia coli* have demonstrated increased resistance to phage infections and the antimicrobials albomycin, rifamycin derivative CGP 4832, and colicin M (5, 11). However, it is unknown if the mutation in the *S.* Enteritidis *fhuA* would provide similar resistances considering the specificity of bacteriophages (18).

The PT13a subpopulation isolate contains a 92 bp deletion in the ORF *sen3898* which is predicted to encode a 503 amino acid protein in the reference PT4 strain. The predicted protein is not only truncated at amino acid residue 108, but there is also a frame-shift is due to the deletion. The encoded wild-type protein is believed to belong to the GntR-family of transcriptional regulators and, more specifically, contains an aminotransferase domain of the MocR sub-family. Based on this homology, it is predicted that *sen3898* is likely involved in amino acid metabolism (49). The deletion observed in the PT13a 21046 isolate could affect growth in an environment limited in specific amino acid availability.

Isolate 21046 contains the largest identified polymorphism between the two subpopulations. The 135 bp upstream of *sen4316*, as well as the first 80 bp of gene are deleted from the isolate resulting in a 215 bp deletion. The gene *sen4316* contains a glycine-glycine-aspartic acid-glutamic acid-phenylalanine (GGDEF) domain, which denotes diguanylate cyclase activity. Diguanylate cyclase synthesizes 3 -5 -cyclic diguanylic acid (c-di-GMP), which is a regulatory signal involved in virulence, motility, long-term survival, and cellulose and fimbriae synthesis needed for biofilm formation (42). Genes encoding diguanylate cyclase are redundant in many bacteria species and *S.* Enteritidis alone has 12 genes with GGDEF domains (41).

Sonano *et al*. found diguanylate cyclase activity to be necessary for cellulose synthesis and the complete absence c-di-GMP resulted in a smooth colonial morphology. Activity specifically from *sen4316* was shown influence expression of *spvA* and *csgD*, regulators of virulence and cellulose and fimbriae expression, respectively (41). The disruption of this gene in PT13a 21046 could explain its inability to form a biofilm on agarose media.

S. *Enteritidis PT13a isolate 21027 ORF-disrupting polymorphisms*. The PT13a 21027 subpopulation isolate contained a total of 9 polymorphisms that appeared to disrupt an open reading frame. It contains a single base pair deletion 1,559 base pairs into *foxA*, resulting in a frame-shift three quarters into the protein. Similar to *fhuA*, *foxA* is predicted to be a ferrichrome outer membrane transporter or, more specifically, a ferrioxamine B receptor. The ability to utilize ferrioxamine as the sole source of iron by possession of the *foxA* gene is unique to *Salmonella enterica* subspecies I, II, and IIIb. A *foxA* knockout mutant introduced in *S.* Typhimurium reduced its ability to colonize rabbit ileal loops and reduced its virulence in mice compared to the parent strain (23). The observed reduced virulence in mice is intriguing since the rodent is a natural carrier for *S.* Enteritidis (19). It is possible that the natural deletion in *foxA* mitigates illness in the mouse host, thus, becoming allowing it to be good vector *S.* Enteritidis spread in the farm house. It is possible that the frame-shift deletion could have a similar effect in its pathogenicity of the egg-laying hen.

Several open reading frame disruptions observed in isolated 21027 occur in genes involved in specific substrate metabolism. The gene *putP* encodes a 498 amino acid proline permease, a sodium/proline symporter. Isolate 21027 contains a SNP that introduces an early termination codon at amino acid residue 120, thereby truncating a majority of the protein. *Salmonella* spp., including *S*. Enteritidis, contain additional low-affinity proline permease

systems, but PutP is a high-affinity proline transport system and is required when proline is the sole source of nitrogen or carbon (9, 24). The disruption observed in subpopulation isolate 21027 may limit its growth potential under specific nutrient availability conditions. The gene *sen0992* is predicted to encode KdgM, an oligogalacturonate-specific outer membrane porin. Deletion mutants in the gram negative *Erwinia chrysanthemi* resulted in the organism being unable to grow on oligogalacturonides longer than trimers (3). Subpopulation PT13a isolate 21027 contains a 1 bp deletion in the *kdgM* gene, resulting in a frame-shift mutation and truncated protein. The ORF *sen1576* is a poorly characterized protein with a conserved nucleotide sequence within *Salmonella* species. Amino acid sequence alignment analysis against the Pfam protein database (http://pfam.sanger.ac.uk/) predicts the protein to be a metal-dependent aminohydrolase, likely involved in metabolic processes. The *S.* Enteritidis isolate 21027 contains a SNP in *sen1567* that introduces an early termination codon at the amino acid residue 409 near the carboxyl-terminus and, thus, potentially limiting the subpopulation's substrate utilization for growth.

Another disrupted ORF predicted to be involved with substrate acquisition, but poorly characterized, is *ydjN*. The gene is a putative sodium dicarboxylate symporter but few experimental studies on the gene or its product exist. One study suggested that its expression was up-regulated in the presence of lactoperoxidase, but then it was subsequently found not to be involved with the oxidizing agent (36). Subpopulation isolate 20127 contains a SNP that introduces an early termination codon at amino acid residue 225 in *ydjN*. The gene *cysN* is part of the cystein regulon which participates in the assimilation of sulfur from inorganic sulfate through the cystein biosynthetic pathway (38). CysN is a subunit of sulfate adenylyltransferase required for selenocysteine tRNA synthesis (32). Isolate 20127 contains an early termination

codon in the carboxyl-terminal end of CysN, though it is unclear whether the truncation affects the catalytic domain of the protein. The subpopulation isolate also contains a 10 bp deletion in *dsdA*, which encodes a D-serine dehydratase. The deletion results in a frame-shift mutation. D-serine dehydratase catalyzes the breakdown of D-serine to pyruvate and ammonia and is necessary when D-serine is provided as the sole source of nitrogen or carbon. The deletion likely explains the results of a previous study where isolate 21027 could not grow in minimal media with D-serine as the nitrogen source (28). Other research has shown that D-serine is present in mammalian urine and that uropathogenic *E*. coli utilizes the *dsd* operon to modulate infection (34). An 11 bp deletion at the 3' end of *yjfK* removes the termination codon and the ORF stays in frame to theoretically fuse to the downstream ORF, *yjfL*. The two ORFs are highly conserved and, thus far, only found in *Salmonella* species. However, there is no known homology or predicted functions for these ORFs.

A gene disruption of particular interest in the *S*. Enteritidis PT13a subpopulation 21027 is a single base pair deletion in *sefD* after amino acid residue 67, resulting in a frame-shift mutation. SefD is predicted to be the minor adhesin subunit of *S*. Enteritidis-specific fimbriae SEF14. The major protein subunit of SEF14, SefA, is believed to be translocated outside the outermembrane only in the presence of SefD. Therefore, the absence of the SefD adhesin prevents the assembly of SEF14 fimbriae, effectively knocking out the *sef* operon (10). Considering the potential role of fimbriae in cell invasion and adherence, several studies have evaluated SEF14 virulence (8, 10, 37, 46). Systemic spread of SEF14 mutants appeared no different than the wild type strain, though the LD_{50} in some experimental setups was significantly increased in the mutants. It has been shown that SEF14 may be required for

macrophage uptake and, thus, systemic spread, though that seemingly contradicts the results of previous studies regarding systemic spread of SEF14 mutants (10).

Polymorphisms resulting in amino acid substitutions. Non-synonymous polymorphisms result in an amino acid substitution within an open reading frame but otherwise do not disrupt the ORF. The effects of such a change in a protein can range from no functional change to deleterious due to a change in the structure (33). Confirmatory sequencing found a total of 102 non-synonymous polymorphisms between the two *S.* Enteritidis subpopulations relative to the reference strain. Details of the polymorphisms for isolates 21046 and 21027, including predicted amino acid chemical property changes, can be found in Tables 3.10 and 3.11, respectively. The Clusters of Orthologous Groups (COGs) function categorization scheme for putative ORFs (44) showed that genes involved in metabolism had 37 amino acid substitutions, the highest number of non-synonymous polymorphisms between the two subpopulations. Isolates 21027 and 21046 contained amino acid substitutions in 19 and 18 different genes, respectively. A total of 16 non-synonymous polymorphisms were found in genes involved in regulation, with 21027 containing substitutions in 7 different genes and 21046 containing substitutions in 9 different genes. Isolate 21027 contained an additional 10 genes involved cell wall and membrane biogenesis, 8 genes involved in energy production and conversion, and 2 genes with roles in cell-defense mechanisms with non-synonymous polymorphisms. Isolate 21046 contained non-synonymous polymorphisms in 5, 3, and 3 genes in the same respective categories. There were a total of 18 genes of unknown functions, 12 in isolate 21027 and 6 in isolate 21046, with amino acid substitutions.

Other classes of polymorphism. Additional classes of identified polymorphisms between the two *S.* Enteritidis subpopulations include those found in non-coding regions, in

RNA genes, and synonymous polymorphisms found within open reading frames. While seemingly minor in comparison to ORF-disrupting polymorphisms, such classes of polymorphisms may very well have a profound effect on the phenotype and physiology of an organism by altering potential regulatory elements or affecting secondary structure. However, prediction and verification can be difficult and involve complex experimental methods (48).

Intergenic or non-coding regions are found between open reading frames and, therefore, are not translated into amino acids or proteins. However, there are circumstances, such as the 5' untranslated regions, where non-coding DNA is transcribed and acts as a regulatory element (43). Additionally, intergenic regions also may contain transcriptional terminators, promoters, operators, and other regulatory signals (35). A total of 38 intergenic polymorphisms were identified between the two *S.* Enteritidis subpopulations. Isolate 21046 was found to contain 16 intergenic regions containing a polymorphism and isolate 21027 contained 22 intergenic regions with polymorphisms. Specific details of the intergenic polymorphisms found in 21046 and 21027 can be found in Tables 3.4 and 3.5, respectively.

The discovery of small or other regulatory RNA elements located in the genome is a relatively new field and the published *S.* Enteritidis PT4 genome does not included annotations for such elements (39). However, other well-known RNA genes, such as transfer RNA (tRNA) and ribosomal RNA (rRNA) are readily identifiable. Polymorphisms in such RNA genes have the potential to alter secondary structure conformations and have also been shown to act as a mechanism to avoid antimicrobial target sites (7). The two subpopulations contained a total of 12 polymorphisms in RNA genes. Isolate 21046 contained 7 polymorphisms in 23S rRNA genes and 1 polymorphism in a tRNA gene, while isolate 21027 had 4 polymorphisms in 23S ribosomal genes. All but one of the polymorphisms were found in the *rrlA, rrlC,* and *rrlE* genes,

which is consistent with a previous study describing the nucleotide basis for different ribotype patterns observed between the two subpopulation isolates (27). Details of the RNA gene polymorphisms found in isolates 21046 and 21027 can be found in Tables 3.6 and 3.7, respectively.

Synonymous polymorphisms were the next most abundant polymorphism class identified by confirmatory sequencing, with a total of 82 polymorphisms between the two *S.* Enteritidis subpopulations. Synonymous polymorphisms do not affect the amino acid sequence and, therefore, presumably do not alter any structure function relationships of the encoded protein. However, it is possible that a nucleotide polymorphism is located in the target sequence of a DNA binding protein and could have a potential regulatory impact. Subpopulation isolate 21046 contained 39 synonymous polymorphisms, while 43 synonymous polymorphisms were found in isolate 21027. DNA substitution details of the synonymous polymorphisms identified in isolates 21046 and 21027 can be found in Tables 3.8 and 3.9, respectively.

Discussion

In summary, this study was able to compare the whole genomes of two *S.* Enteritidis PT13a subpopulations that vary in their ability to internally contaminate an intact egg. By comparing two organisms within the same phage type, but which nonetheless varied in phenotype, genetic background noise was reduced to a minimum so that a higher portion of the polymorphisms observed were more likely to be biologically relevant. The results are consistent with the conclusion of a previous study that suggested any genomic differences between the subpopulations are likely to be small-scale nucleotide polymorphisms (28). The identified genomic differences comprised less than 0.01% of their respective genomes. However, the potential phenotype and physiological impact of the polymorphisms is significant and has been

previously characterized by LPS structural analysis, Phenotype Microarray, and animal infection studies (16, 28). An organism considered highly clonal such as *S.* Enteritidis can exhibit great heterogeneity even when there is seemingly insignificant genomic diversity.

A total of 247 polymorphism events were found that differentiated the *S.* Enteritidis PT13a subpopulations from each other. Isolates 21046 and 21027 contained 111 and 136 polymorphisms, respectively, that differentiated them from the pandemic *S.* Enteritidis PT4 reference strain. Of the polymorphisms that are predicted to disrupted open reading frames, 75% (9 out of 12) are believed to be involved in metabolic pathways as well as over 37% (37 out of 99) of the non-synonymous polymorphisms. This result suggests that the two PT13a subpopulations may be evolving towards a more specific ecological niche compared to *S.* Enteritidis PT4. It has been previously shown that host-restricted *Salmonella* serotypes carry a substantial amount of pseudogenes compared broad host-range serotypes (21). Most notable is that subpopulation isolate 21027 contains more ORF-disrupting polymorphisms than isolate 21046 (9 versus 3 ORF-disrupting polymorphisms) and that over 77% (7 out of 9) are predicted to be involved with metabolic pathways. This pattern of potential metabolic restriction in isolate 21027 also correlates well with previous Phenotype Microarray studies where it was observed that 21027 did not contain the metabolically-diverse toolset of isolate 21046 or of the reference PT4 strain (28). Therefore, it is conceivable that subpopulation isolate 21027 has evolved in such a way that it cannot utilize specific substrates necessary to internally contaminate or survive in the hen egg. A potential example of nice specialization is the 10 bp deletion observed in *dsdA* in isolate 21027. A previous Phenotype Microarray study found that not only could 21027 not utilize D-serine as a sole source of nitrogen or carbon, but D-serine also acted as a growth inhibitor even when other suitable carbon and nitrogen sources present. If D-serine were present

in the avian reproductive tract, then isolate 21027 would have difficulty in successfully contaminating the egg. Perhaps strains like 21027 facilitate a part of the infection pathway that occurs external to the hen. For example, it could perpetuate especially well in rodents.

Polymorphisms resulting in an amino acid change (non-synonymous) were the most abundant class of polymorphism that differentiates the two *S.* Enteritidis PT13a subpopulations, comprising of 45% of total polymorphisms (111 out of 247 polymorphisms). Synonymous polymorphisms made up over 33% of the total polymorphisms (82 out of 247 polymorphisms) and polymorphisms in non-coding regions only comprised of 15% of the total polymorphisms (38 out of 247) between the subpopulations. This may suggest a non-neutral selection pressure occurring in the PT13a subpopulations when compared to the PT4 reference strain. For non-synonymous polymorphisms, it is believed that essentiality of a gene and its level of expression applies the most pressure for amino acid substitutions. High levels of expression can be costly to the cell, especially depending on the supply of available tRNA. Therefore, there can be selection pressure towards optimizing DNA substitutions that result in codons of abundant tRNA (33). The function or role of a gene may also influence its DNA substitution rate. This process could facilitate processes such as variability in outer membrane proteins, which facilitates escape from host defense mechanisms or aids survival and growth in diverse environments. Examples of non-synonymous polymorphisms in genes previously reported to be involved in albumen survival are amino acid substitutions found in *rfbX* and *rfbB* in the 21027 non-egg contaminator isolate. These genes are involved in LPS O-antigen biosynthesis and another gene within the same operon, *rfbH*, was found to be necessary for survival in the egg-white (14). Whether or not these non-synonymous polymorphisms truly have any effect on the final encoded product

requires further analysis. However, specific LPS structure, such as HMM LPS, has been established as a risk factor for egg contamination (16).

The Comparative Genomic Sequencing technology was found to have both benefits and limitations. At the time of the study, *de novo* sequencing was still prohibitively expensive and, thus, the Nimblegen service was an affordable option that offered whole genome sequence analysis. While mutation mapping was critical for identifying regions of potential polymorphisms between the two subpopulations, the technique had an overall false-positive rate of 56.25% when confirmed with Sanger method sequencing. However, the false-positive rates can be acceptable because they are inclusive, rather than exclusive, for true polymorpshims. False-negative rates of mutation mapping cannot be determined without sequencing the entire genome of both isolates using another method. Furthermore, the mutation mapping technology suffers from the same limitations of any microarray technology in that it is not a *de novo* analysis. The probes on the chip are only representative of the reference strain and do not survey any novel sequences in the test samples. This is exemplified by a previous study that showed the same two *S.* Enteritidis PT13a subpopulations to contain the Fels-2 prophage, which was not detected by mutation mapping since the reference strain does not contain the prophage (28). Therefore, it is possible for additional polymorphisms to exist or even the absence or presence of discrete genes between the two subpopulations that were not surveyed by the mutation mapping array used in this study. Once affordable and time efficient, *de novo* whole genome sequencing technologies will be the preferred method for whole genomic DNA analysis between any desired samples, but downstream processing of detailed information using bioinformatics will remain the major rate limiting step.

Comparative Genome Sequencing has revealed an unexpected amount of genetic diversity between subpopulations of the same parent strain that were isolated from the same host animal (30). Identifying all the polymorphisms between the two *S.* Enteritidis PT13a subpopulations on a genome-wide level is only the beginning of characterizing their biology. It would be ideal, but perhaps unrealistic, to experimentally analyze every identified polymorphism for its effect subpopulation phenotype, physiology, or even virulence. A more practical approach is to prioritize further analysis on polymorphisms predicted to have the most dramatic effect on the organism, such as ORF-disrupting polymorphisms, with additional emphasis on those with a predictable phenotype or expectation to play a role in the egg contamination pathway. Once significant genetic determinants are identified, nucleic acid-based epidemiological tools can be developed and used to gather more meaningful biological information from field isolates.

Acknowledgements

Funding for this research was provided by USDA-ARS CRIS project no. 6612-32000-004-00. Special thanks to Carolyn Cates, William Sago, and Gus Darnell for their technical assistance in this research.

References

1. **Albert, T. J., D. Dailidiene, G. Dailide, J. E. Norton, A. Kalia, T. A. Richmond, M. Molla, J. Singh, R. D. Green, and D. E. Berg.** 2005. Mutation discovery in bacterial genomes: metronidazole resistance in *Helicobacter pylori*. Nat Methods **2:**951-3.

2. **Anonymous.** 2010. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food - 10 states, 2009. MMWR Morb Mortal Wkly Rep **59:**418-22.

3. **Blot, N., C. Berrier, N. Hugouvieux-Cotte-Pattat, A. Ghazi, and G. Condemine.** 2002. The oligogalacturonate-specific porin KdgM of *Erwinia chrysanthemi* belongs to a new porin family. J Biol Chem **277:**7936-44.

4. **Bochner, B. R., P. Gadzinski, and E. Panomitros.** 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. **11:**1246-55.

5. **Braun, V.** 2009. FhuA (TonA), the career of a protein. J Bacteriol **191:**3431-6.

6. **Carver, T., N. Thomson, A. Bleasby, M. Berriman, and J. Parkhill.** 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics **25:**119-20.

7. **Cochella, L., and R. Green.** 2004. Isolation of antibiotic resistance mutations in the rRNA by using an *in vitro* selection system. Proc Natl Acad Sci U S A **101:**3786-91.

8. **Dibb-Fuller, M. P., and M. J. Woodward.** 2000. Contribution of fimbriae and flagella of *Salmonella* Enteritidis to colonization and invasion of chicks. Avian Pathol **29:**295-304.

9. **Dila, D. K., and S. R. Maloy.** 1986. Proline transport in *Salmonella* Typhimurium: putP permease mutants with altered substrate specificity. J Bacteriol **168:**590-4.

10. **Edwards, R. A., D. M. Schifferli, and S. R. Maloy.** 2000. A role for *Salmonella* fimbriae in intraperitoneal infections. Proc Natl Acad Sci U S A **97:**1258-62.

11. **Ferguson, A. D., J. Kodding, G. Walker, C. Bos, J. W. Coulton, K. Diederichs, V. Braun, and W. Welte.** 2001. Active transport of an antibiotic rifamycin derivative by the outer-membrane protein FhuA. Structure **9:**707-16.

12. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, T. J. Humphrey, and F. Van Immerseel.** 2009. Mechanisms of egg contamination by *Salmonella* Enteritidis. FEMS Microbiol Rev **33:**718-38.

13. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, and F. Van Immerseel.** 2008. *Salmonella enterica* serovar Enteritidis genes induced during oviduct colonization and egg contamination in laying hens. Appl Environ Microbiol **74:**6616-22.

14. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, and F. Van Immerseel.** 2009. The *Salmonella* Enteritidis lipopolysaccharide biosynthesis gene *rfbH* is required for survival in egg albumen. Zoonoses Public Health **56:**145-9.

15. **Gast, R. K., J. Guard-Petter, and P. S. Holt.** 2002. Characteristics of *Salmonella* Enteritidis contamination in eggs after oral, aerosol, and intravenous inoculation of laying hens. Avian Dis **46:**629-35.

16. **Guard-Bouldin, J., R. K. Gast, T. J. Humphrey, D. J. Henzler, C. Morales, and K. Coles.** 2004. Subpopulation characteristics of egg-contaminating *Salmonella enterica* serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl. Environ. Microbiol. **70:**2756-63.

17. **Guard-Petter, J.** 2001. The chicken, the egg and *Salmonella* Enteritidis. Environ. Microbiol. **3:**421-30.

18. **Guard-Petter, J.** 1999. Phage type and other outer-membrane characteristics of *Salmonella enterica* serovar Enteritidis associated with virulence, p. 221-232. *In* A.M. Saeed, R.K Gast, M.E Potter, P.G. Wall (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control., 1st ed. Iowa State University, Ames.

19. **Guard-Petter, J., D. J. Henzler, M. M. Rahman, and R. W. Carlson.** 1997. On-farm monitoring of mouse-invasive *Salmonella enterica* serovar Enteritidis and a model for its association with the production of contaminated eggs. Appl. Environ. Microbiol. **63:**1588-93.

20. **Guard-Petter, J., L. H. Keller, M. M. Rahman, R. W. Carlson, and S. Silvers.** 1996. A novel relationship between O-antigen variation, matrix formation, and invasiveness of *Salmonella* Enteritidis. Epidemiol. Infect. **117:**219-31.

21. **Holt, K. E., N. R. Thomson, J. Wain, G. C. Langridge, R. Hasan, Z. A. Bhutta, M. A. Quail, H. Norbertczak, D. Walker, M. Simmonds, B. White, N. Bason, K. Mungall, G. Dougan, and J. Parkhill.** 2009. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. BMC Genomics **10:**36.

22. **Humphrey, T. J.** 1999. Contamination of eggs and poultry meat with *Salmonella enterica* serovar Enteritidis, p. 183-192. *In* A.M. Saeed, Gast, R.K., Potter, M.E., Wall, P.G. (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control., 1st ed. Iowa State University Press, Ames.

23. **Kingsley, R. A., R. Reissbrodt, W. Rabsch, J. M. Ketley, R. M. Tsolis, P. Everest, G. Dougan, A. J. Baumler, M. Roberts, and P. H. Williams.** 1999. Ferrioxamine-mediated Iron(III) utilization by *Salmonella enterica*. Appl Environ Microbiol **65:**1610-8.

24. **Liao, M. K., S. Gort, and S. Maloy.** 1997. A cryptic proline permease in *Salmonella* Typhimurium. Microbiology **143 (Pt 9):**2903-11.

25. **Lu, S., P. B. Killoran, and L. W. Riley.** 2003. Association of *Salmonella enterica* serovar Enteritidis *yafD* with resistance to chicken egg albumen. Infect Immun **71:**6734-41.

26. **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:**607-25.

27. **Morales, C. A., M. Musgrove, T. J. Humphrey, C. Cates, R. Gast, and J. Guard-Bouldin.** 2007. Pathotyping of *Salmonella enterica* by analysis of single-nucleotide polymorphisms in cyaA and flanking 23S ribosomal sequences. Environ Microbiol **9:**1047-59.

28. **Morales, C. A., S. Porwollik, J. G. Frye, H. Kinde, M. McClelland, and J. Guard-Bouldin.** 2005. Correlation of phenotype with the genotype of egg-contaminating *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **71:**4388-99.

29. **Patrick, M. E., P. M. Adcock, T. M. Gomez, S. F. Altekruse, B. H. Holland, R. V. Tauxe, and D. L. Swerdlow.** 2004. *Salmonella* Enteritidis infections, United States, 1985-1999. Emerg. Infect. Dis. **10:**1-7.

30. **Petter, J. G.** 1993. Detection of two smooth colony phenotypes in a *Salmonella* Enteritidis isolate which vary in their ability to contaminate eggs. Appl Environ Microbiol **59:**2884-90.

31. **Rank, D. L., M. A. Saeed, and P. M. Muriana.** 2009. Cloning of *Salmonella enterica* serovar Enteritidis fimbrial protein SefA as a surface protein in Escherichia coli confers the ability to attach to eukaryotic cell lines. Appl Environ Microbiol **75:**6622-5.

32. **Riva, A., M. O. Delorme, T. Chevalier, N. Guilhot, C. Henaut, and A. Henaut.** 2004. The difficult interpretation of transcriptome data: the case of the GATC regulatory network. Comput Biol Chem **28:**109-18.

33. **Rocha, E. P., and A. Danchin.** 2004. An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol Biol Evol **21:**108-16.

34. **Roesch, P. L., P. Redford, S. Batchelet, R. L. Moritz, S. Pellett, B. J. Haugen, F. R. Blattner, and R. A. Welch.** 2003. Uropathogenic *Escherichia coli* use d-serine deaminase to modulate infection of the murine urinary tract. Mol Microbiol **49:**55-67.

35. **Rogozin, I. B., K. S. Makarova, D. A. Natale, A. N. Spiridonov, R. L. Tatusov, Y. I. Wolf, J. Yin, and E. V. Koonin.** 2002. Congruent evolution of different classes of non-coding DNA in prokaryotic genomes. Nucleic Acids Res **30:**4264-71.

36. **Sermon, J., E. M. Wevers, L. Jansen, P. De Spiegeleer, K. Vanoirbeek, A. Aertsen, and C. W. Michiels.** 2005. CorA affects tolerance of *Escherichia coli* and *Salmonella enterica* serovar Typhimurium to the lactoperoxidase enzyme system but not to other forms of oxidative stress. Appl Environ Microbiol **71:**6515-23.

37. **Shah, D. H., M. J. Lee, J. H. Park, J. H. Lee, S. K. Eo, J. T. Kwon, and J. S. Chae.** 2005. Identification of *Salmonella* Gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. Microbiology **151:**3957-68.

38. **Sheng, X., X. Huang, J. Li, X. Xie, S. Xu, H. Zhang, and H. Xu.** 2009. Regulation of sulfur assimilation pathways in *Salmonella enterica* serovar Typhi upon up-shift high osmotic treatment: the role of UhpA revealed through transcriptome profiling. Curr Microbiol **59:**628-35.

39. **Sittka, A., C. M. Sharma, K. Rolle, and J. Vogel.** 2009. Deep sequencing of *Salmonella* RNA associated with heterologous Hfq proteins in vivo reveals small RNAs as a major target class and identifies RNA processing phenotypes. RNA Biol **6:**266-75.

40. **Smith, L. M., J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. B. Kent, and L. E. Hood.** 1986. Fluorescence detection in automated DNA sequence analysis. Nature **321:**674-9.

41. **Solano, C., B. Garcia, C. Latasa, A. Toledo-Arana, V. Zorraquino, J. Valle, J. Casals, E. Pedroso, and I. Lasa.** 2009. Genetic reductionist approach for dissecting individual roles of GGDEF proteins within the c-di-GMP signaling network in *Salmonella*. Proc Natl Acad Sci U S A **106:**7997-8002.

42. **Solano, C., B. Garcia, J. Valle, C. Berasain, J. M. Ghigo, C. Gamazo, and I. Lasa.** 2002. Genetic analysis of *Salmonella* Enteritidis biofilm formation: critical role of cellulose. Mol Microbiol **43:**793-808.

43. **Sorek, R., and P. Cossart.** Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat Rev Genet **11:**9-16.

44. **Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin.** 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res **29:**22-8.

45. **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:**1624-37.

46. **Thorns, C. J., C. Turcotte, C. G. Gemmell, and M. J. Woodward.** 1996. Studies into the role of the SEF14 fimbrial antigen in the pathogenesis of *Salmonella* Enteritidis. Microb Pathog **20:**235-46.

47. **Vugia, D. J., M. Samuel, M. M. Farley, R. Marcus, B. Shiferaw, S. Shallow, K. Smith, and F. J. Angulo.** 2004. Invasive *Salmonella* infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. Clin. Infect. Dis. **38 Suppl 3:**S149-56.

48. **Waters, L. S., and G. Storz.** 2009. Regulatory RNAs in bacteria. Cell **136:**615-28.

49. **Winefield, C. S., K. J. Farnden, P. H. Reynolds, and C. J. Marshall.** 1995. Evolutionary analysis of aspartate aminotransferases. J Mol Evol **40:**455-63.

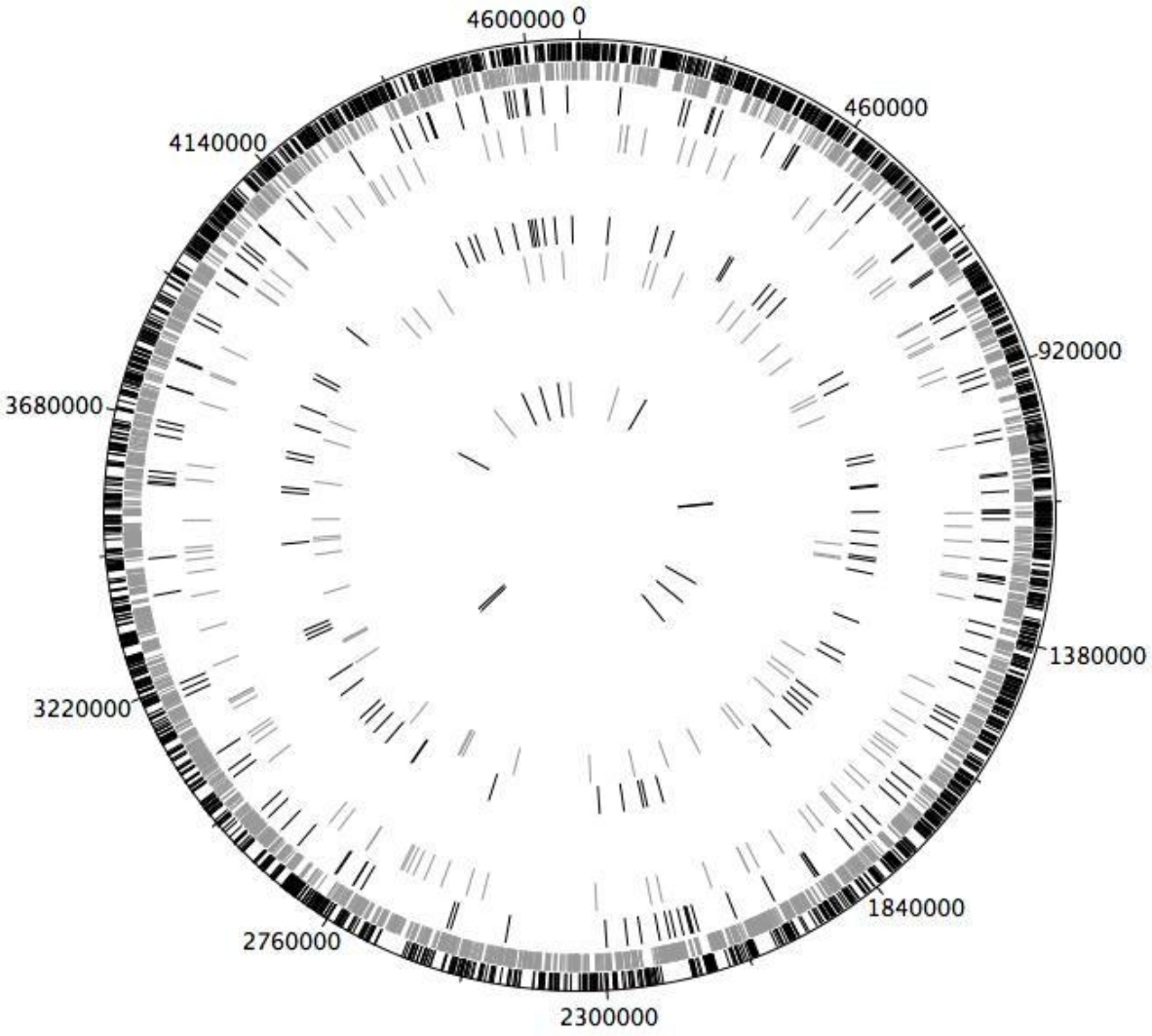

Figure 3.1 Visual representation of the identified polymorphisms between two subpopulations of *S.* Enteritidis PT13a. All positions are based on the complete genome sequence of *S.* Entertidis strain P125109 (GenBank AM933172). From the outside going in, tracks 1 and 2 are open reading frames on the forward strand (track 1, black) and the reverse strand (track 2, grey). Track 3 (black) contains all the polymorphisms in 21027 and track 4 (grey) contains all 21046 polymorphisms. Tracks 5 (black) and 6 (gray) are non-synonymous polymorphism for 21027 and 21046, respectively. Track 7 contains all ORF-disrupting polymorphisms with black ticks representing those found in 21027 and grey ticks representing those in 21046.

Table 3.1. Summary of 247 confirmed polymorphisms that differentiate subpopulations of *Salmonella* Enteritidis PT13a[a]

Type of genetic change	PT13a 21027	PT13a 21046
Deletion	8	2
Change in termination	4	2
Amino acid substitution	58	44
RNA gene polymorphism	4	8
Intergenic polymorphism	22	16
Synonymous	43	39
Total:	136	111

[a]Polymorphism is defined as contiguous base pair differences relative to the reference strain *S.* Enteritidis PT4 22079, ranging from a single nulceotide polymorphism to a 215 base pair deletion

Table 3.2. Polymorphisms that disrupt open reading frames in *S.* Enteritidis PT13a subpopulation 21046[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
226922	GCGGACAAAGGTCATTATCTGGCT CCGCTTCATACTGCTTACCTTTCG	153929762	*fhuA*	C/T	TER at gln 470	Ferrichrome outer membrane transporter
4189388	GCGTTAAGGGTCTGTTGTTCGTCA TATTCGGGTGGATAAACTGCCGC	153929844	*sen3898*	Deletion 4189388-9479	92 bp deletion	Putative GntR-family transcriptional regulatory amino acid transport protein.
4642252	CCCAAAGAGTAGGGAGATAAAGGG ACGCCAGTCCAACGATAATCACTC	153929850	*sen4316*	Deletion 4642252-2466	215 bp deletion	Diguanylate cyclase, synthesizes 3'-5'-cyclic diguanylic acid, a transcriptional regulatory signal

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046, or deletion with specified locus

Table 3.3. Polymorphisms that disrupt open reading frames in *S.* Enteritidis PT13a subpopulation 21027[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
393529	TGCGGCAGACCGGATACTATTTAC TTTCACATCGTAGCTTAGCCAGGCG	153929111	*foxA*	G/-	1 bp deletion	Ferrioxamine B receptor precursor
1095448	AATTTATCCGCCTGCTGCGTCA GGGTCAGTGTATTTGCTACTACGTTA	153929129	*putP*	G/T	TER at cys120	Putative sodium/glucose symporter
1099079	ACGCTATAGAACTGGGCGTAAGT GGACGTGAATACTCAAGACAAAGTC	153929131	*kdgM*	A/-	1 bp deletion	Oligogalacturonate-specific porin protein
1681327	AATGCGTAAGCTGGGTCTGGCATA ATCAACGCCCACCGGATCTCAATA	153929153	*sen1576*	G/A	TER at trp409	Predicted metal-dependent hydrolase with a TIM-barrel fold
1828544	ACTGACGGCGGAAGGTCTGGTTCA GCGCTTCGACATTAAGTGGAATGGAG	153929161	*ydjN*	G/A	TER at trp225	Predicted Na+/dicarboxylate symporter
2957802	GCACATTATCGCCATCCAACAGGT AGCTACGACGTTAAACTTGCGGGT	153929185	*cysN*	C/A	TER at glu447	Sulfate adenylyltransferase subunit 1
3880104	ACGCTATTTCTGGGCTATGCGGTA GTTCCATTTCATCTTCCGGCACCA	153929213	*dsdA*	Deletion 3880104-0113	10 bp deletion	D-serine dehydratase
4472579	ATGACGCTCAACTGGCAAGAACG TAGTTGAATCACCAGCGCAATGCC	153929223	*yjfK* and *yjfL*	Deletion 4472579-2589	11 bp deletion, fusion	Fusion of YjfK and YjfL. Proteins are of unknown function

Table 3.3 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
4574262	GGGAATAAGCCCTCCGATACTTGT CGCCAATTTGCGCACCCTTTAT	153929231	*sefD*	A/-	1 bp deletion	*S*. Enteritidis fimbrial protein subunit

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046, or deletion with specified locus

Table 3.4. Polymorphisms located in non-coding, intergenic regions in *S.* Enteritidis PT13a subpopulation 21046[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Intergenic region	Nucleotide substituion[d]
95090	GCCTTATTTGGTGTGACCGTCT CTGTACGAGGAAGCGTGGATATGT	142459090	*sen0083-0084*	A/G
95192	GCCTTATTTGGTGTGACCGTCT CTGTACGAGGAAGCGTGGATATGT	142459091	*sen0083-0084*	G/A
838275	AATTCATAGTGGGTAGTGTCCGGC TGGTAGTCAGCTCTTCCAGCACAT	142459092	*ybhP-ybhQ*	G/A
1035267	GATTGCCAATAACAATCTGTGGGCCAG TGTGCAATCACACTACACGAATACC	142459093	*ymbA-rmf*	A/G
1218767	CTGTAATCACATGACGTAGCGCCT AACACCTCTTTATGTTCCGCCAGC	142459094	*purT-yebG*	C/T
1541966	TGCTTTGTCCTGAACCGTCAACA AGCCCATCCGATAACGAAACGGTA	142459095	*sifB-ugtL*	G/T
1647065	GCGCCTTTGTAGAAGAACTGCCAA TGCGTCACTTGCGTTATTCGTCTG	142459096	*rspA-ynfA*	T/C
1679797	TCAACCTGCTGGATGATGACGA CCTGTAAATCGATGGTACGGGTATGG	142459097	*ompN-sen1576*	G/A
2008113	TGTTGGTGTGTGTTATGCCGCC CGTAGCTTGCAGAGATGGCTTTGT	142459098	*csgB-csgD*	G/A

Table 3.4 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Intergenic region	Nucleotide substituion[d]
2607540	ATGGCTCGCTGTACAATTTGCCG TGATTATTCAGGCGCTGGATGGCA	142459099	*cdaR-perM*	T/C
2764932	TGCTCCGGCGTCATTAATACCTGT ACGCGTTGAGCATACAGAGGCAA	142459100	*aroF-yfiR*	G/A
2764986	TGCTCCGGCGTCATTAATACCTGT ACGCGTTGAGCATACAGAGGCAA	142459101	*aroF-yfiR*	T/C
3058806	GGCGTTAACCAGCGCATATACCAT CACACGAATTACACCAACCTGCGT	142459102	*aas-galR*	C/T
3995713	AGAACCCAGGTACTGCGCCTAAAT AATGGTGTGAGTTCTTAAACTTGGG	142459103	*trxA-rho*	A/T
4223137	AAGTCCTGCCTGAAGGGACGAGGTTA GAATTACGTGTTCACTCTTGAGACTTGGT	142459104	*murI-sen_r017*	G/A
4503965	ACAGACTTCCCGGTAGGTTGGATT TTCAGACGAACATTTGCGGCTACC	142459105	*msrA-ytfM*	G/T

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.5. Polymorphisms located in non-coding, intergenic regions in *S.* Enteritidis PT13a subpopulation 21027[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Intergenic region	Nucleotide substituion[d]
186989	CAATTTGCGCAGTGGTAAAGGTCG CTCTCATTGTCATAATGCGGCGCT	142459106	*yacH-acnB*	T/G
664020	AGGATGGCGCCTTACTGATT ATGCCAGTTGCCTTTATCGTCCCA	142459107	*dcuC-pagP*	C/T
716869	ACCACATGCTTAAAGCTGACCTCG AGCTGCAACGCCCTACCCA	142459108	*nagB-nagE*	T/C
721398	GTAGCGAGTCAAACCTCACGTACT TTGCGCCTGCGGAACAACGAT	142459109	*glnS-ybfM*	C/T
920824	TGAGTACACGCAGCAGCGAACTTT ATCAACCATTCCGAACCGAGATGC	142459110	*artP-ybjP*	G/A
1126845	TGGGATGGTCTTCCCAATGAACGTC AAATCCCGGCCCTACGGGT	142459111	*yedP-yodD*	C/T
1159508	GCTGCACGTGGAGATTTGCTTGAT TCGCGGTCGGCATACAAATTGAAG	142459112	*cysT-glyW*	C/T
1164733	TAAGCCACTGTCCGTAACGTCCAA TTTATTGCCGCCTCCAGCGAGGTAT	142459113	*tyrP-yecH*	C/T
1363969	AGGTAAAGCTTGAAGATGGTGGTGGG AGGCAAAGGAATGCTTTACAGCGG	142459114	*purU-tyrT*	C/T

Table 3.5 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Intergenic region	Nucleotide substituion[d]
1383947	CGACCAGCGGAAGGATATAGATGA GTCTATGCGCAGGTTTCGGTCAAA	142459115	*cls-yciI*	A/G
1542105	TGCTTTGTCCTGAACCGTCAACA AGCCCATCCGATAACGAAACGGTA	142459116	*sifB-ugtL*	T/C
1575469	CATCAAGCAGGGTGCGGATCAAAT CCCAGCATCAGCAGGCGTTTA	142459117	*TetR-smvA*	G/T
1903278	TCAAGCTCTTCCTCTTTCCAGCCA ACTGACGAATACGACGACCGTGAA	142459118	*sen1800-1801*	G/A
1908163	AATTTAGCGAGTACGTCTACCTCCGC CAGTTGTCCGAATTTGTTCCGCGT	142459119	*sen1805-1806*	G/A
1990183	GCCGCATTCACGTCATCGTAAAGA TCCTGGAGAATCTCGCTAATGCGT	142459120	*sen1885-1886*	C/T
2140726	CCCGTAGTGTAATCCATCAACACCCA AGTTCAATCCACTGCAACAACTGGC	142459121	*dacD-phsC*	G/A
3942784	ATTGCGTGACCTGTTCCCAACTGT AAACCAGCGGCCTCTGTCG	142459122	*glmU-SEN3677*	G/A
3970231	TCCAGAAAGTTCCAGTTTCCCTGC TGTACCAACTACAGGAGCCGCAAA	142459123	*yieP-rrsC*	T/A

Table 3.5 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Intergenic region	Nucleotide substituion[d]
4066450	TATTCCGTTAACCTGGTGGCGCGTAA AGAAGAGTTGGATCTGGCATGGCT	142459124	*rrsA-ileT*	CT/GC
4066454	TATTCCGTTAACCTGGTGGCGCGTAA AGAAGAGTTGGATCTGGCATGGCT	142459125	*rrsA-ileT*	T/A
4066460	TATTCCGTTAACCTGGTGGCGCGTAA AGAAGAGTTGGATCTGGCATGGCT	142459126	*rrsA-ileT*	A/C
4596231	TTCTGTATCAAGCGGCATGG CGCTGCCTCTCAAATCATCACA	142459127	*yjiH-yjiJ*	A/G

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.6. Polymorphisms located with RNA genes in *S.* Enteritidis PT13a subpopulation 21046[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Ribosomal region	Nucleotide substituion[d]	Product Description
3973593	TCGTGAAGAGGGAAACAACCCAGA TCGTCGTCACGCCTCAGTGTTAAA	159695620	*rrlC*	C/T	23S
3973601	TCGTGAAGAGGGAAACAACCCAGA TCGTCGTCACGCCTCAGTGTTAAA	159695621	*rrlC*	GCCT/TGTC	23S
3973611	TCGTGAAGAGGGAAACAACCCAGA TCGTCGTCACGCCTCAGTGTTAAA	159695622	*rrlC*	GT/CA	23S
4066568	GTGGATCAGAATGCCACGGTGAAT TCGCAGTGAACCCTTTCAGGTACA	159695623	*ileT*	G/A	tRNA
4068166	AATGTTATCACGGGAGACACACGG ACCTACACGCTTAAACCGGGACAA	159695624	*rrlA*	C/T	23S
4068174	AATGTTATCACGGGAGACACACGG ACCTACACGCTTAAACCGGGACAA	159695625	*rrlA*	GCCT/TGTC	23S
4068184	AATGTTATCACGGGAGACACACGG ACCTACACGCTTAAACCGGGACAA	159695626	*rrlA*	GT/CA	23S
4268912	ACTGACTTACGAGTCACGTTTGAG TATTTAGCCTTGGAGGATGGTCCC	159695627	*rrlE*	T/C	23S

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.7. Polymorphisms located with RNA genes in *S.* Enteritidis PT13a subpopulation 21027[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Ribosomal region	Nucleotide substituion[d]	Product Description
2751287	TCCATCCGTTAGGTGAGCAACACA GATGCTGGCGTTTCCGTTTCAGAT	142460370	rrlG	AA/GG	23S
2751875	AGATGAATTCACGAGGCGCTACCT CAGCATGTGTGTTAGTGGAAGCGT	142460371	rrlG	T/C	23S
3972440	TCCAGAAAGTTCCAGTTTCCCTGC TGTACCAACTACAGGAGCCGCAAA	142460372	rrlC	A/G	23S
4269421	TGTGTGTTAGTGGAAGCGTCTGGA CGAGGCGCTACCTAAATAGCTTTC	142460373	rrlE	TT/CC	23S

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.8. Synonymous polymorphisms in *S.* Enteritidis PT13a subpopulation isolate 21046[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
129592	GGCCGGTATGCTGGGTAAATTTCT ATTTCCGTCACCGACATCATCCC	141096998	*leuC*	C/T	Valine, leucine and isoleucine biosynthesis
267017	AGCAGGCGACAGGATCGTTAAAGT ACTAAAGTCTTGAACCCGCTCGGA	141096999	*yaeL*	G/A	Zinc metallopeptidase
633202	AATTACGCTCTTTGACGAGCAGGG AGATAGCCCATCTGCGCC	141097001	*cstA*	C/T	Carbon starvation protein
684875	AAACCGCAGCAAGCACCGGTTAGC CCTGCATCATTTCCCACGTTGCTT	141097004	*uxaA*	C/A	Pentose and glucuronate interconversions
789332	CTGGCGTTTCTGTTCCTGCTGATT TTCATGGCATCGAAGACGCGGAA	141097005	*sen0712*	C/T	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789335	CTGGCGTTTCTGTTCCTGCTGATT TTCATGGCATCGAAGACGCGGAA	141097007	*sen0712*	T/G	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789347	CTGGCGTTTCTGTTCCTGCTGATT TTCATGGCATCGAAGACGCGGAA	141097010	*sen0712*	T/C	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789359	CTGGCGTTTCTGTTCCTGCTGATT TTCATGGCATCGAAGACGCGGAA	141097012	*sen0712*	C/T	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789362	CTGGCGTTTCTGTTCCTGCTGATT TTCATGGCATCGAAGACGCGGAA	141097014	*sen0712*	T/C	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2

Table 3.8 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
900607	CGCCTTCGACATTAATCGCCACTT GCTTAGGATCATTGGCGCGTCTTA	141097015	*deoR*	G/A	Putative transport protein/regulator
1164068	CCATCACAAATCCACGTGGGTAGA TAACCGCTTCCTGTTTAGCGCA	141097017	*tyrP*	G/A	Tryptophan/tyrosine permase family
1189967	GGTTCGGGTCGTAGCTGGGCATAGA CCCGATGATATCGCCGACTTTCG	141097020	*ftsI*	C/T	Penicillin-binding protein
1328395	TGTGCAGCGTCCTCAGGATTTCTA TACGCATATCCAGCAGCTCAACGA	141097021	*sen1250*	C/T	Ni,Fe-hydrogenase I large subunit; fermentative growth
1503343	TCTGGATCCGTTTCTCGTAAGGCA CGCTGTCTGCTTCGATTAGAAGAACC	141097024	*sen1418*	C/T	ABC-type polar amino acid transport system
1610591	TTTATCTGATGATGGCCGGGCTAC ATAGCACCAGCCAGATAGAACGCA	141097025	*sen1509*	C/A	Sugar metabolism/transportprotein
1738692	CAGCAGGCTGCATCAAGATACT GCTGGATAACATGGCATGAAGGT	141097027	*ssaJ*	C/T	Type III secretion, lipoprotein EscJ
1775109	CTTTACCCTTGAACGTAACCGAGC AACAGCTACCGTGGGCTGGTAAA	141097030	*sufB*	G/A	Activates cysteine desulfurase SufS with SufCD
1800374	ACGTTCATCATGACCTTCAGCGGT GAAACAAGCGGTGCAGATTGAGCA	141097032	*pps*	T/C	Forms phosphoenolpyruvate from pyruvate

Table 3.8 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
1806066	ATCACCGGTTAAACACCGGCAAAC GGCGTTGTTAGAAACCTGTCAGCA	141097033	*ydiV*	G/T	Putative diguanylate cyclase/phosphodiesterase
1934221	GAACCACCGCCAGAATACGGTTAT CGTGGCGATGCCATTATTGATTGG	141097035	*lolD*	G/A	ABC-type antimicrobial peptide transport; lipoprotein transport
2008553	GGCAGCTGTCAGATGTGCGATTAAA TCATTATTAGAGGCACCAATACGTAACT	141097037	*csgD*	T/C	DNA-binding transcriptional regulator; activates csg operons involved in biofilm formation
2203153	AAATGTCGTAATCCTGCGCCGCCT AGCATGACAACAGACAATCCGCTG	141097039	*wcaA*	G/A	Colanic acid:glycosyl transferase
2557022	CAATGCTGTCTTGCCCGTGAGTTT GGGCGTACCAAAGAAGCCAATGAA	141097041	*sen2427*	G/T	Iron-dependent peroxidase
2642024	AAATCACGTCCACGCTGGCTTTCA CTTGATCCACAACCCGTCAGCAAT	141097044	*ratB*	G/A	Pseudogene; unknown function; putative outer membrane protein; many repeats
2664521	GTCCGGATCGCGAATATTGTCTTC AGACGACGCCAGCAAAGTGA	141097045	*sen2510*	G/A	Anaerobic DMSO reductase
2829865	TCAGAGGCGGAACTCCTCGATATT CGAGGCAAATAAACCGATGCTG	141097047	*ygaE*	G/A	Regulator of gab gene expression; DNA-binding trnascriptional regulator CsiR; GntR family
3014949	GCAATCATGTAGTGAATCGCCGGA GCATTTATCGGCGAGAGAAACGGT	141097049	*fucA*	C/T	Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases

Table 3.8 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
3073028	GGTATGTGACATTGCAACACCTCAC TGACGTTGCACCTTCCGGATGTAT	141097051	*rcnA*	C/G	Nickel/cobalt efflux protein RcnA ; membrane protein conferring nickel and cobalt resistance
3135538	GTTTAACTGCTTCCAGCGCGGATT GCTGTCTACATAGGTATGTGCATGGAAC	141097054	*fadR*	G/A	Similar to Escherichia coli regulator for uxu operon (AAC77280.1); GntR family
3225510	TCTACCGGGAACATCAACATCCGT AAGAGCTACAGGCAGATTCCGATG	141097055	*par C*	G/A	Decatenates newly replicated chromosomal DNA and relaxes positive and negative DNA supercoiling
3363009	TTGACTTTCAGCAGGCAGGCAAAG AGGATGCGACAGAGCGAAATCCT	141097057	*yrbH*	G/A	D-arabinose 5-phosphate isomerase: polysialic acid capsule expression protein
3432239	TTTCCGGTGATGCAGCAGGCTTAT CGCTAATGATAGCCGGGATGTTCA	141097060	*accB*	G/A	AcetylCoA carboxylase, BCCP subunit:biotin
3586071	GTTGTGATAATCGCCATCGTCCCT AAAGATCATCCGCTGTTGCACGG	141097062	*glgX*	C/A	Glycogen debranching enzyme
3793026	CTCTCAAGTATGTCATAACGCTGC ATCGTGTTCGGCGAGAAGGGTTTA	141097063	*rfaY*	T/C	LPS: modification of heptose region of core
3846103	ACCGGCGATAGCTCTGTTATGATG TTGGATTGCCGTGAGTCACCAGTA	141097066	*yicL*	G/A	Permease
4018775	AGTGCGTACTCTGCACTTTAACGG ATACGTGAAGCGATCGTGCGATGA	141097067	*cyaA*	C/T	Adenylate cyclase

Table 3.8 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
4296365	TTTATTGCACACCAGCGTACCGCC GCCCTGACCGATGAACTGGT	141097070	*lysC*	T/C	Lysine sensitive aspartokinase III
4331502	CGATCACTTTATCGAAATGTTCCAG(GGGTTAGTCGGCGTGATGTATGT	141097072	*uvrA*	C/T	DNA excision repair enzyme
4365146	CGCAAACAGGGAACGCATGGAAAT TACCGCCGATAGCGACCTTAAA	141097074	*yjcE*	C/T	Sodium/hydrogen exchanger: pH maintenance

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.9. Synonymous polymorphisms in *S.* Enteritidis PT13a subpopulation isolate 21027[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
241034	TTACTGGCGATGGTCGGCTACTTT CCGCCATGAAAGCGCTGAAACATA	141100738	*yadQ*	G/A	Chloride channel protein important for extreme acid resistance
255011	TGTCAACCATGGTGCCTTCGTCAA ATTCATAATCGCCACAGCGGATCG	141100740	*dapD*	A/G	2,3,4,5-tetrahydropyridine-2-carboxylate N-succinyltransferase
351933	TCTGCTCATTCCACCGTATTAGCC CCGCGTTCTCTTAACCCGCTAAAT	141100742	*sen0311*	G/A	Permease
399525	TTTCGGCGGTCTCCTACAACATGA GGCGGCAAGCCAGGTATCATTAAA	141100744	*prpD*	C/T	2-methylcitrate dehydratase; propionate catabolism
661305	CAGGCGCAGCAGCAACTTATTGAT TATATTCTGCGTGCATCTCCGCCA	141100746	*dpiB*	G/T	Sensory histidine kinase; regulation of cit operon and plasmid inheritance genes
789389	GTTCGGTGCTATTAGGTGAAGGCT TTCCACCACGTCATCGGCGTAAT	141100748	*sen0712*	T/C	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789488	GTTCGGTGCTATTAGGTGAAGGCT TTCCACCACGTCATCGGCGTAAT	141100750	*sen0712*	G/C	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789503	GTTCGGTGCTATTAGGTGAAGGCT TTCCACCACGTCATCGGCGTAAT	141100752	*sen0712*	C/G	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2
789512	GTTCGGTGCTATTAGGTGAAGGCT TTCCACCACGTCATCGGCGTAAT	141101249	*sen0712*	A/G	Biotin requiring enzyme; forms oxaloacetate from pyruvate and CO_2

Table 3.9 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
909178	TGCAGGAGTTTGCGAAACGTAAGC AGATACGCAATATGCGCCACTTCG	141101251	*potG*	G/T	Putrescine transport ATP-binding protein
937016	TATTGCGTCGGTGGTGTCGATTGT AGCCATGACTGACCCATTATGCGA	141101253	*yjbZ*	C/T	Macrolide export
1175680	TTGCGCATCCAGATTATCGACAGC TCCATATCAGCCAACAGCTCGTCA	141101255	*motB*	G/A	Couples rotation with proton/sodium motive force
1276695	AATCAACGTTAGCGACTGTCCCTGG CTGCCAGCTCTATCTGGATTGTCA	141101257	*sen1194*	G/T	Multidrug efflux system protein domain PRK10504
1281189	AATCTACCGGGTATTAACCTGCGG GGATACGGTGTCCATCGAGAGTAA	141101259	*ftsI*	C/T	Penicillin-binding protein; peptidoglycan synthetase
1317293	TGCCGCATATATTAGGCGATGTCC GCATATTGCATCACCTGCGGATCT	141101261	*emtA*	C/A	LysM/invasin domains; lytic transglycosylase; Goose Egg White Lysozyme (GEWL) domain.
1416687	CACGGTTGTAAATTACCGTGGCGT GACTATTTACCTGTTGTAACGAGCG	141101263	*osmB*	G/A	Osmotically inducible lipoprotein; may aid stationary-phase survival
1474943	CTTCTGGTGCTTTATATCCTTGAGC CATAAAGGCAGTCGAGCCAGCAAA	141101265	*sen1386*	T/C	Degenerate lysogenic bacteriophage in Enteritidis, E. coli and Typhi CT18;
1785781	TATGTGGTTTGGCTGGTCCTTCCT CGCTGTAATAAATGCCGGTCGCTT	141101267	*ydiM*	C/A	Major facilitator superfamily

Table 3.9 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
1801472	TCTGGCATATCTTCCGCCGT TATTCAGGGAATGCCGTGCTATGC	141101269	*pps*	A/G	Forms phosphoenolpyruvate from pyruvate
1901744	TAGCGGTAGCCATGATTCGCCATT CCAGCTCCGTTGGTTAACGATGAT	141101271	*sen1799*	G/A	Predicted neuraminidase (sialidase)
2062613	TAACAAGGCGCTACGTAAGATGGG ACGCTGACTGATCCTAATGCCTCA	141101273	*sen1970*	G/A	DNA breaking rejoining enzyme superfamily. Phage integrase domain.
2142697	GCCACCTTGCCCTGAATGAAGTAA TGGTGGTCAGCTGCGATGTCTATT	141101275	*phsA*	T/C	Thiosulfate reductase precursor
2144992	GTTGAGCATCGAAGCGGGTTTACT TCTTGCCTGCATTATTTGTGTCTT	141101277	*sopA*	A/G	Secreted effector protein
2158117	ACATGGGCTGGAATCGCGTGTAT TTGAACCGATGACCACACGGGCAA	141101279	*hisA*	C/T	Amino acid transport and metabolism
2200415	TTAATGAGACCAGCGCCAGTGGAA ACGGCATTTAACTTCCTGTTCGGC	141101281	*wcaD*	G/A	Putative colanic acid polymerase
2256571	GAAGATAGCCGCTAGCAAAGAGTC CGGACCATAAGAACTTCCGCGAAT	141101283	*sen2150*	T/C	Putative protein of unknown function; possible lipoprotein
2468813	ACCGAACATTCAAACTGCTGG TCTACAGCGCTCGTGTGAATATGG	141101285	*purF*	G/A	Nucleotide:amidophosphoribosyltransferase

Table 3.9 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
2566232	AGCCCTGTCCGGTTTCAAAGTACA CGCGAATACGTACTGAGCGATGAA	141101287	*eutB*	G/T	Adenosylcobalamin: ethanolamine ammonia-lyase
2736265	TGCGTCTGGCAGGGATCAACAAT GGGTTAACGAACCCGGCATTATCA	141101289	*srmB*	C/T	ATP-dependent RNA helicase
3046877	AAGTTCATGTCTTGCAGCGCGT AAACGACCCGGACTCTCGTCGTATTA	141101291	*thyA*	G/A	Catalyzes formation of dTMP and 7,8-dihydrofolate from 5,10-methylenetetrahydrofolate and dUMP
3377504	GATCTGCGCGCCTTCTTTGACTTT ACTGCTGGCGGAAGTAGTCGTAAA	141101293	*gltB*	G/A	Catalyzes the formation of glutamate from glutamine and alpha-ketoglutarate
3576418	TCATGCCTCAGTGGTGCCTTATGA TCCGATCAAACAGTTCGGGCTTCA	141101295	*sen3355*	G/A	Dihydroxyacid dehydratase
3594297	ATCTGCCGGGAGCAGATTTGAA AGAACACCAGGACAGTATGGGCAA	141101297	*yhhW*	G/A	Hypothetical protein with a cupin domain
3683137	CAACGAATACGCTGCCCGATATGC GCGGGCGTGAATCATTTCGTCATT	141101299	*bcsA*	C/A	Polymerizes uridine 5'-diphosphate glucose to cellulose; acts with BcsB, BcsZ and BcsC in cellulose biosynthesis
3744397	AATCATCGTCGCCGACTGGAAAGT CGCTCTGGCCAACTGAGATCTTT	141101301	*sgbH*	G/T	3-hexulose-6-phosphate isomerase
3795869	CGTGCCGGGAATAATATCATGCTG GGGCGTCTTAAATTTGAAGGACAGAA	141101303	*rfaB*	C/A	LPS: 1,6-D-galactosyltransferase

Table 3.9 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Product information
4004604	TCGATATTCGTCCGGATACGATG AATAGAAGGCAGCTCGATACGCCC	141101305	*wecE*	C/T	Transaminase: DegT, DnrJ, EryC1
4027489	CGCTTCTTTGAAGACGAAGACGGA ATGCGGTTCTGCTCGATGTTGATG	141101307	*corA*	C/T	Magnesium/nickel/cobalt transporter
4133134	CGATAATTTCGACGCTTAGCGACCA CCACGGCGAAGTCGCCCGTTAT	141101309	*sen3843*	T/G	Pseudogene in *S.* Enteritidis; missing 5' 28 aa of *S.* Typhimurium gene
4152430	TGAAATTATGCGCGGCTTGATGCG CTGTTGGGTTGTCAGGGCGTTAAT	141101311	*ego*	A/G	ABC-type sugar transport system
4351538	AAACGGGAAATCCAGCACACTGAC GCTCTGCAGTACCTTCGAATTTGG	141101313	*siiE*	T/C	Fusion protein; sulfite oxidase (SO); molybdopterin binding
4418467	GAATACTCTTGCTACGCCTGTTTC TCGCCACCTGATACCTTAAGTTCG	141101315	*hilD*	C/T	Arabinose operon control protein: DNA binding
4420944	CCGGTCTGGCCTCGGATAATACTAA TTATTGCCTGATCCGGAGTGAGTCTT	141101317	*phoN*	A/G	Non-specific acid phosphatase

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046

Table 3.10. Nonsynonymous polymorphisms in *S.* Enteritidis PT13a subpopulation isolate 21046[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
82917	TGATCGTCACGGATTTCGGCTTCA TTATCCACTGACGATACGGCGGAA	153929758	*caiC*	C/T	AA val to ile 258 nn to nn	AMP-binding enzyme;L-carnitine production
204450	TTGGACTGGTGCCACAGGAGTTTA TAACACTTCCACCTCCAGAGTCGAG	153929760	*yadG*	G/A	AA gly to ser 166 nn to np	ABC-type multidrug transport system
304365	AAGGGCACTACCATGACAACGCAT TCAGCCATTCGCCACTGGAATAGT	153929764	*yafE*	G/A	AA ala to thr 101 nn to np	Menaquinone/biotin biosynthesis
466363	AGCAGACAATCGTCAGGTTGCCAT CGAGCAAGACGGCGATTACAGTTA	153929766	*apbA*	C/T	AA ser to asn 237 np to np	Pantothenate and CoA biosynthesis
505777	TGTCGTGTTGTCACAGTGATGTCC TGCGTTTGACTATCTGGCAATCCG	153929768	*rpmE2*	C/A	AA ser to arg 18 np to bp+	50S ribosomal protein L31 type B
562921	TGGTGATTAACGCCGTAGTCG TCATAATATTTGTAATCACCGCTGGC	153929770	*allP*	C/T	AA ala to val 268 nn to nn	Cytosine/uracil/thiamine/ allantoin permeases
643964	TGTCAACGGTAAAGAGTTCGGCCA CCGCCACACGCTTCCCTTTAAATA	153929772	*ahpF*	C/A	AA asp to glu 281 ap to ap	Protects DNA from alkyl hydroperoxides
696522	TTCTGGGCCAGTAAACGAAACGC CGCTGGCCTGGTATATGAATGACA	153929774	*sen0629*	G/A	AA arg to cys 288 bp to np	Pseudogene in SEN after AA 171
818905	TCCAGAATCACGGCGGCTATTTCA TGGCATCCTTACACTTCCATGACC	153929776	*bioA*	C/T	AA asp to asn 110 ap to np	Biotin metabolism

Table 3.10 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
833026	TGTTCTGTCGGGACTGGTGCATAA AATGGCAAGCGCAGGAAA	153929778	*ybhL*	T/C	AA phe to leu 178 nn to nn	Putative membrane protein. BAX inhibitor (BI)-1 like protein family
883330	TAGACGTAGTGCAGACGTGTCCAGA TTATTGACAGCGCGGTACAAGAGTCG	153929780	*yliG*	C/A	AA arg to leu 139 bp+ to nn	2-methylthioadenine synthetase: RNA binding; catalytic activity; iron-sulfur cluster binding
1249431	AATGGACGGCGAACTGAAGAACCT GCTGCGGAGCCATATTTATTCCC	153929782	*sen1164*	A/G	AA asp to gly 68 ap to nn	Remnant of putative phage terminase
1284118	CCCAATTGACGGCGCAACATTAAC CACTCTGCTTTCTATAGCCCTGCT	153929784	*sen1204*	C/T	AA arg to gln 129 bp+ to np	Putative membrane protein
1289642	TAATGGGCATCGGCTTCGCATAGA TCCTATCGCCGAGGAACATAACCT	153929786	*sen1210*	C/T	AA gly to ser 421 nn to np	Diguanylate cyclase/ phosphodiesterase
1579797	TGCAGTACGTCGCCCATATAAACG TTTGCGCGGTATTCAGGACAGTCA	153929788	*yddG*	C/T	AA cys to tyr 78 np to nn	Unknown function
1638463	ATAGCGCTGAACGGCAAACAAACC GCACTGAGCTTTATGCGTGGCATT	153929790	*dcp*	T/C	AA asp to gly 386 ap to nn	Dipeptidyl carboxypeptidase II; amino acid transport
1659052	GGCTTCTTTCCAAGTTACGCTGTC GCGGCAAACTATCAAGCGCAGTAT	153929792	*sen1558*	T/G	AA leu to arg 22 nn to bp+	Proline/glycine betaine transport system
1723077	GCCCTTACTGGGAAGCTGGATACT CGTGGCTTGAGGGAAAGGGCGTAA	153929794	*ydhC*	G/A	AA gly to ser 241 nn to np	Putative integral membrane transport protein; major facilitator superfamily

Table 3.10 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
1843306	TAAATTCGCCACCGTCAGGTGG AGTACCGCGCCGTTTGGTGGTAT	153929796	*astB*	G/A	AA arg to cys 39 bp to np-	Succinylarginine dihydrolase
1860952	GAGGTGTAATGTATTGCTTATGGA TTGGCGTGCTCAATGCTCTGAAAG	153929798	*ydjG*	G/T	AA arg to ser 245 bp to np	Hypothetical oxidoreductase
1987706	ATGGGCTTTAGCCGTCTGGGATTA ACCGTTAACTTTGCGCAACCACTC	153929800	*grxB*	A/G	AA ile to val 5 nn to nn	Glutaredoxin 2 [Posttranslational modification, protein turnover, chaperones]
2084629	TTGAGCGACTACCGGGAGCATAAT TTGGTTCCTATGGGCGCTTCACA	153929802	*sen1995*	T/C	AA arg to gly 149 bp to nn	Uknown
2182436	GTTCTATCCAACTCTGATACGCCC GCATCCAGCGATCATTTAGTCCGT	153929804	*rfbB*	G/T	AA his to asn 261 bp- to np	dTDP-glucose 4,6-dehydratase
2307838	CAAACGCATACGCCGCTTTGTTCA TGTTGACGGGCGTTTCTTACATGC	153929806	*fruA*	C/T	AA ala to thr 321 nn to np	Fructose-specific metabolism/transportprotein
2526004	GCGTTTCAACAAACTCAGCCACGA TATGACGAGATTCTGACGCGGCTT	153929808	*yfeA*	A/G	AA his to tyr 600 bp- to nn	Diguanylate cyclase/phosphodiesterase
2677674	CCGCATCGGTATTCATGACTTCGA ATGCCCAACGTCAGGAGCTGGAT	153929810	*pepB*	G/A	AA pro to ser 187 nn to np	Cytosol aminopeptidase family
2686717	TGCCAGAACGCAACGTAAACCTGT AGCGGATAGCGCTGTCTGAATA	153929812	*sen2527*	C/T	AA thr to met 191 np to nn	Putative membrane protein

Table 3.10 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
2690699	CATCGCCGCTATGCAATTGTTCCT GGCGGTCTGTGGATATGGCA	153929814	*nicO*	C/A	AA phe to leu 128 nn to nn	High-affinity nickel protein
2851388	ACTGGAACCGCTCACCAATAAGAC TCGATCATACCCGGATGCAAGC	153929816	*luxS*	C/T	AA ala to thr 109 nn to np	Produces autoinducer; AI II quorum sensing
3081698	TTTCTGCTTGCGCAGCCAGTTGTT CAGCGCTATTCCGATGAAAGCAGT	153929818	*sen2874*	C/T	AA met to ile 6 nn to nn	Putative periplasmic protein
3148475	ATCCACTAATCTTTCCACCGCCTG GCAACAATATCAGTGGCGAACGCA	153929820	*yggR*	C/T	AA ala to thr 138 nn to np	Twitching motility protein; pilus retraction
3153640	AGTGTTGGTTGTGCGAGAGGTCTT TAAGACCAATACGACTGACGTGGC	153929822	*ansB*	G/A	AA gly to asp 237 nn to ap	Catalyzes the formation of aspartate from asparagine CRP
3153778	AGTGTTGGTTGTGCGAGAGGTCTT TAAGACCAATACGACTGACGTGGC	153929824	*ansB*	C/T	AA ala to val 283 nn to nn	Catalyzes the formation of aspartate from asparagine CRP
3296285	AACTCCTCACGCAGGCGAAT GTGATGGCGGGAATACGTATCGA	153929826	*tdcE*	T/C	AA asp to gly 135 np to nn	Pyruvate formate lyasea:anaerobic glycolysis
3407874	ACGCTGCATAACTTGCCCAATCTC AAGATCCAAAGAACGCAAACCCGC	153929828	*sen3191*	C/G	AA val to leu 140 nn to nn	GntR transcriptional regulator
3446987	AGCTGGTCAATTCCGTTCTCGGTA GGCCGACATGAGGAATCATCTCAA	153929830	*acrF*	G/A	AA arg to his 992 bp+ to bp-	Multidrug transport;acriflavin resistance

Table 3.10 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
3455766	TTAACTCATCCGAACGCGCAAACG AAAGCGGTATTTGATTGCTGGCCC	153929832	*yrdC*	C/G	AA gly to ala 171 nn to nn	Required for maturation of 16s RNA; binds preferentially double stranded RNA
3507751	ACGAACATCTGGAAGAAGAGATG AGGCGTTGCGCCTCCCAAA	153929834	*nirD*	G/A	AA val to ile 28 nn to nn	Nitrite reductase
3613736	TACGGACCGCTGGGATCTTTCTT AGGTATTCTGATGATCTCTCCGGG	153929836	*livK*	C/A	AA ala to ser 207 nn to np	Branched chain amino acid binding
3734249	AACATCAATACTACCGCCTGACCC CGCCGAAACGAAGAGGTCGT	153929838	*avtA*	G/A	AA ser to asn 60 np to np	Valine--pyruvate transaminase
3788343	TGAATAGCGAGCAGCAGGCAT TGAAAGGCTGTGCGTTCCGCTTTA	153929840	*rfaF*	C/G	AA thr to arg 331 np to bp+	LPS heptosyltransferase II
4146218	AAATTGCTGGAACCATGCCCTGCT TCTGTACCGACTGTTTCTCGCCTT	153929842	*sen3857*	A/G	AA thr to ala 42 np to nn	Putative ADP-ribosylglycohydrolase
4274880	AAGCGACCCTGTTGATTGAAACGC ATGTACTGCACTGCCACGCGAATA	153929846	*aceB*	G/C	AA val to leu 345 nn to nn	Malate synthase A
4532612	GCATATCGATAGCGAGCGTTGTCT GCAGATGCATCCGGCGCTAAA	153929848	*treB*	A/G	AA leu to pro 235 nn to nn	Trehalose-specific IIBC part of phospotransferase system
4579004	TGAAGCCGGACGAACCGTTAAAGA TCTGGTTAAGCGCTTCCGGTGT	141097075	*sen4256*	C/T	AA TER at gln 25 np to TER	Putative transposase (pseudogene)

Table 3.10 continued

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046, or deletion with specified locus

[e]Abbreviations: TER, terminating codon; bp, base pair; standard 3-letter amino acid codes; nn, neutral non-polar; np, neutral polar; ap, acidic polar; bp, basic polar; np-, neutral slightly polar; bp+, basic strongly polar

Table 3.11. Nonsynonymous polymorphisms in *S.* Enteritidis PT13a subpopulation isolate 21027[a]

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
74552	ATAAAGTGGATGCACCGTCGGGTA CAGTCTTATCAGTCTGGTCGAGGT	153929104	*dapB*	G/A	AA asp to asn 263 ap to np	Reduces 2,3-dihydrodipicolinate to 2,3,4,5-tetrahydrodipicolinate
196607	TATCGGACTCTAGCTGGTGGATCA CCGGCCTCAAATGTCGCTATTACA	153929107	*speD*	T/C	AA thr to ala 55 np to nn	Spermidine biosynthesis from putrescine
237853	GACGGCAACATTACAGGTGGAT CAAAGGTAACTAGTGCGCTTGCGA	153929109	*stfG*	A/C	AA thr to pro 98 np to nn	High-affiinity transport of Fe (3+) ferrichrome
401171	AGCAACGCATTCTGGATGTCTCCCT AAACCGCCAAAGACCACCGAAT	153929113	*prpE*	C/G	AA arg to gly 60 bp+ to nn	Forms propionyl-CoA from propionate
513535	GCGGCGCAATCTATTGGCATTTCA CACGCCTGTAACAAACAGAATAGCG	153929115	*acrR*	A/G	AA ser to gly 137 np to nn	Regulates acrAB operon; TetR/AcrR family
533246	TCTGCGGTTGATATTGAGTCAGGG TTGCGCCACCTCTTTACGGATACT	153929117	*ushA*	T/G	AA asp to glu 41 ap to ap	UDP-sugar hydrolase. Nicotinate/nicoinamide/purine/ pyrimidine metabolism;
565589	CCGATACCAGCGCCAGAATCAAAT GCCAGTGGAATAGCGCAAA	153929119	*ybbY*	T/C	AA pro to leu 238 nn to nn	Purine permease
799987	CGATGCGGATGATCGATGGGATTA GAACAGATCCTCACCGATCTGCAA	153929121	*galE*	T/A	AA lys to asn 253 bp to np	UDP-glucose 4-epimerase; galactose and nucleotide sugars metabolism
831768	AGCTGATCCCGCTGTGCCA TAAACGGCGCAGGGCCGACTACAAT	153929123	*moaD*	C/A	AA phe to leu 9 nn to nn	Biosynthesis of the molybdenum cofactor

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
1019733	CGCAGGTATGACGCGGTATGTTAT ACAGACCCGTGAAGAATATCTGGC	153929125	*sen0917*	T/G	AA glu to ala 109 ap to nn	Hypothetical phage protein
1035179	GATTGCCAATAACAATCTGTGGGCC TGTGCAATCACACTACACGAATACC	153929127	*ymbA*	C/T	AA ala to val 169 nn to nn	Putatuve lipoprotein
1163256	GCCCGGCTTGAAGATTATCCAGAA GCTGGTGTTTATCTGCGGCATTGT	153929133	*yecA*	T/G	AA trp to gly 182 nn to nn	Metal-binding protein
1216457	CACTGGATAACGATGTCATTGCCTC AGCGCGCCAAACAACTCA	153929135	*edd*	A/G	AA lys to arg 503 bp to bp+	Pentose phosphate pathway
1248118	AAGGGTGTCTGAAATATTCCGGCG CCAGCATAATGTTTAGCTTCTTCTGC	153929137	*sen1162*	T/A	AA val to asp 52 nn to ap	Phage-tail assembly-like protein
1279901	TGATGCACTGGAAGAAGCGAAGT TCGAACTCAACGTTCTGGCCTTCA	153929139	*yobF*	A/G	AA lys to arg 12 bp to bp+	Unknown function
1287804	TATAGGGCAGCAGCCCGTTGTAT AGAAGCTCGCATAGCCTTTACCGT	153929141	*yoaE*	G/A	AA val to ile 29 nn to nn	TerC, tellurium resistance
1319180	GCGCGATTATGGTGCTGGTGATTT TATGTAAACGCCGCCGCGAAAGAT	153929143	*treA*	T/C	AA thr to ala 512 np to nn	Hydrolysis of trehalose to glucose
1445558	ACTACCTTTGCCGACCACATCGTA ATAGACCTGCATCGCTTCCAGCTT	153929145	*sen1355*	C/T	AA asp to asn 49 ap to np	2'-hydroxyisoflavone reductase

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
1533003	ACAGGCAGAGTATCTTAACGGCGA GAACCATTCGCGGATTGTCCATCA	153929147	*ydcK*	A/G	AA val to ala 250 nn to nn	UDP-3-O-[3-hydroxymyristoyl]: LPS
1554930	GCCAACGTTCTGATACCCGCTAAT ACTCGGTTTATGTTCATCCGGCG	153929149	*ydcZ*	Del1554930-935	AA 11&12 deleted in-frame deletion	Unknown function
1656696	CCGCACCTTTCGGGATTGTTTGAA AGAGTCAGACGCGCCAGTT	153929151	*ynfI*	T/G	AA asp to glu 93 ap to ap	TorD; cytoplasmic chaperone TorD
1697323	AAAGAGAAACAGGCGCAGGCCA GCAGCAACCGCGGCTTTA	153929155	*rnfC*	G/A	AA glu to asp 591 ap to ap	NADH ubiquinone dehydrogenase; RnfC; electron transport to nitrogen
1715322	TTGCCACAGGCGCAGGTACTTTAT AAGAAGTGCGAGTTCCCGCATTGA	153929157	*nemA*	C/T	AA pro to ser 38 nn to np	Old yellow enzyme (OYE)-like FMN; N-ethylmaleimide reductase;
1764145	GCAGCTCGAAACGAACATTGAAGC TGTGCCAGGATTCGCCAATAA	153929159	*orf 408*	C/A	AA his to asn 327 bp- to np	Ribokinase; DeoR family
2131226	GCTTTACCACCTTCACTATCGCCA GCGCATTGTTGAACAGCGGATACT	153929163	*pduQ*	A/C	AA asp to ala 64 ap to nn	Propanol dehydrogenase
2172706	AAATACCATTGAGTCACACCACCA CTATCACGTTCATCAGGGTTTCTG	153929165	*rfbX*	T/C	AA arg to gly 300 bp+ to nn	O-antigen transporter: in SG & STY, not in STM
2182397	GTTCTATCCAACTCTGATACGCCC GCATCCAGCGATCATTTAGTCCGT	153929167	*rfbB*	A/T	AA cys to ser 274 np to np	dTDP-glucose 4,6-dehydratase

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
2231955	GGTTGAGGATATCTTACGCCGCT TCTGGTTGTAGCGTTCGTCCAGTT	153929169	*sen2129*	G/A	AA gly to glu 207 nn to ap	Putative inner membrane protein
2293090	ATCGTTATCTGAGCGATCCTGACG AATAGTTTACTGGCGCTCGGC	153929171	*yeiB*	G/A	AA pro to ser 290 nn to np	Putative inner membrane protein
2573093	CGCGAATGGCATGAATGGCAAG ATTGCCTGATGGTGCGCAGAT	153929173	*eutN*	G/A	AA his to tyr 71 bp- to nn	Ethanolamine utilization
2786647	TAGCTGGAGCTTTACACCGTCTAC TTTCATAGAGCGTGATCGTGGCG	153929175	*bapA*	G/T	AA trp to cys 2319 nn to np	Proline/threonine-rich protein
2788796	GCAAGGGATAACATCACCAGCCTTC TACTGACGCCGTTGGCGTTGGTGAT	153929177	*bapA*	A/C	AA ile to leu 3036 nn to nn	Proline/threonine-rich protein
2869327	GCAGCATGTTCTGGCTTTCCAGTT GATGCTGTCAATATGACACTATA	153929179	*ygaA*	C/T	AA gly to arg 72 nn to bp+	Anaerobic nitric oxide reductase transcription regulator
2908407	GTTAGTGCATGGAGAAATGTCTGT GCTAAAGATGATATCTGGTTATTG	153929181	*sicP*	G/A	AA ser to leu 112 np to nn	Chaperone related to virulence
2939833	GCGTGAATTTACTCAACGCAGCGA ATTAACGTTCCATCCTCCAGTGCG	153929183	*pgk*	Del2939833-844	AA 167-170 deleted in-frame deletion	LysR; transcriptional regulator
3032200	TCGTCAGCGGCAATTTCACCATTG AGTGGGCGGATAAGCTCCAGAATA	153929187	*argA*	C/G	AA leu to val 225 nn to nn	N-acetylglutamate synthase

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
3081785	TTTCTGCTTGCGCAGCCAGTTGTT CAGCGCTATTCCGATGAAAGCAGT	153929189	*sen2875*	G/A	AA leu to phe 12 nn to nn	Ail/OmpX-like protein; possible surface antigen
3195186	TCTACCGGGAACATCAACATCCGT AAGAGCTACAGGCAGATTCCGATG	153929191	*hybG*	C/A	AA arg to leu 74 bp+ to nn	Hydrogenase 2 accessory protein HypG
3206865	ACACCGGCGAATATCCTAAGCA AATTGAGGATTGCCGCCTTGGTTG	153929193	*yghA*	C/A	AA ser to arg 110 np to bp+	Oxidoreductase
3218705	TCTGATCGATGGTTTCGCCCATTG ACGTGACGAGATGGATCAACTTGG	153929195	*ygiR*	A/C	AA leu to arg 121 nn to bp+	Fe-S oxidoreductase family 2
3441850	GCTGGCGAAAGCGTTAAATCTGCAA AGTGCCTCGTCAGCAGGCATTA	153929197	*envR*	A/G	AA val to ala 194 nn to nn	Regulator for envCD; acrEF; TetR/AcrR family
3572250	TGGCCTCGGCCAATTGCTTT AGGTGGTGCAAACGCTTTGGTTAG	153929199	*sen3353*	C/T	AA gly to glu 241 nn to np	ABC-type phosphate transport system
3584158	AGTACGGTTCGGCTTGCGGAT GCTGGTCACTGTTTAGCGAAGAGA	153929201	*glgC*	C/A	AA glu to asp 174 ap to np	ADP-glucose pyrophosphorylase
3658575	ACGTTGTGATAAATCGTCGGCAGG TGAATCCCGCTTCGTTTCTGCT	153929203	*sen3427*	C/T	AA gly to glu 369 nn to np	Putative inner membrane protein
3673351	AAAGATTGACGGAGAGCGGCAATG TCAGCGGGTATGACTTCGCCATTA	153929205	*yhjK*	G/A	AA arg to cys 398 bp+ to np	Putative phosphodiesterase; may inhibit biofilm formation

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
3744227	AATCATCGTCGCCGACTGGAAAGT CGCTCTGGCCAACTGAGATCTTT	153929207	*sgbH*	G/T	AA gly to trp 145 nn to nn	3-hexulose-6-phosphate isomerase
3792668	GATCATCCACTGAAGCCTGTTCGT TCTTGCGTCGTATTAAGGGTAAGG	153929209	*rfaZ*	G/T	AA ala to glu 17 nn to ap	LPS core biosynthetic protein
3867458	GCCAAACAGAACACTGACCAGCA ATGAATGCCGAGCAAGTCCACAAC	153929211	*sen3608*	G/A	AA arg to cys 254 bp+ to np	Putative L-fucose permease
4018861	AGTGCGTACTCTGCACTTTAACGG ATACGTGAAGCGATCGTGCGATGA	153929215	*cyaA*	A/G	AA asn to ser 702 np to np	Adenylate cyclase
4368126	GCTGTTATTCTTTGTTCCCGCCGT CAAACGGCGTAGTAACGGAACGTA	153929217	*lrgB*	Del4368126-8137	AA 43-46 deleted in-frame deletion	LrgB-like: murein: penicillin: seritonin
4402066	ATCAGCTCTCTTGCCTGTTGGGAA AACAGTACCGGCAGGACAATCA	153929219	*melB*	G/A	AA gly to asp 331 nn to ap	GPH family melibiose permease II
4418965	TGCGAACTTAAGGTATCAGGTGGC TGGCATTCGCTGTGAGTGATTCCT	153929221	*sen4087*	C/A	AA val to phe 34 nn to nn	Putative cytoplasmic protein
4517227	CACAAGTTGAAGCGCAGCGTAAGA ATGACGCAGCTTGAGAGCGAAT	153929225	*pmbA*	C/T	AA thr to met 103 np to nn	Putative modulator of DNA gyrase
4558130	ATGAAGTACTGGCAGAGGCGAAAG GTCCAGTTTGCCTTTACCGCCATT	153929227	*yjgP*	T/C	AA phe to leu 227 nn to nn	Permease

Table 3.11 continued

Polymorphism Location[b]	Primers (forward and reverse, 5' to3')	dbSNP ID[c]	Gene	Nucleotide substituion[d]	Disruption Type[e]	Product information
4565438	ATGGAAGTTGCCGCCATACGTCAG GCGGCACAGGATAAGGGACTAAA	153929229	*yjgB*	C/T	AA gly to ser 154 nn to np	Alcohol dehydrogenase
4592540	ATGATCAGCATCCTGGACGCCAC TGCCGACGAACTGGTTATTCCGAT	153929233	*sen4271*	C/T	AA thr to ile 215 np to nn	Aspartate racemase
4622524	AATGGCGCGCATCCCATAAAGT CGTCCTGATAGCTTTGAGTCGGTT	153929235	*tsr*	A/G	AA met to val 82 nn to nn	Serine sensor receptor
4666711	AACAGATTCGGCATCCAGGTGGT AAGTGGTTCCGCCGAAACGTCATA	153929237	*yjjK*	T/C	AA asn to ser 99 np to ap	ABC transporter

[a]Relative to reference strain *S.* Enteritidis PT4 22079

[b]Locations are according to *S.* Enteritidis PT4 genome (GenBank accession: AM933172)

[c]dbSNP IDs are NCBI's Entrez dbSNP unique identifiers

[d]Reference Strain 22079 / subpopulation 21046, or deletion with specified locus

[e]Abbreviations: TER, terminating codon; bp, base pair; standard 3-letter amino acid codes; nn, neutral non-polar; np, neutral polar; ap, acidic polar; bp, basic polar; np-, neutral slightly polar; bp+, basic strongly polar

CHAPTER 4

MUTAGENIC CHARACTERIZATION OF POLYMORPHISMS OBSERVED BETWEEN TWO SUBPOPULATIONS OF *SALMONELLA* ENTERITIDIS[2]

[2] Morales, C.A., J.Y. Guard and M.A. Harrison. To be submitted to *Applied and Environmental Microbiology*

Abstract

The genomes of two subpopulations of *Salmonella* Enteritidis PT13a that vary in phenotype and their ability to contaminate the egg were previously re-sequenced, revealing nearly 250 polymorphisms between the two subpopulations when compared to the *S.* Enteritidis PT4 reference genome. Twelve of the polymorphisms were predicted to disrupt open reading frames by naturally occurring deletions. Of these, the genes *sen4316*, *dsdA*, and *sefD* were selected for characterization via deletion knockout mutagenesis in the reference strain *S.* Enteritidis PT4 strain 22079, which was originally used to identify the polymorphisms. The gene *sen4316*, encoding diguanylate cyclase, was confirmed to be necessary for *S.* Enteritidis to produce a biofilm phenotype on agar media under ambient conditions. The gene *dsdA* was confirmed to be necessary for *S.* Enteritidis survival in the presence of D-serine, as well as for the ability to utilize D-serine as the sole nitrogen or carbon source. Phenotype Microarray also revealed a possible role for *dsdA* in resistance to osmotic stress. The product of *sefD* is predicted to be a minor subunit of SEF14, a fimbria unique to *S.* Enteritidis. It was found to be deficient in the non-egg-contaminating *S.* Enteritidis PT13a subpopulation isolate. A *ΔsefD* mutant strain was created and used to challenge egg-laying. Egg production and blood calcium levels in hens infected with the *ΔsefD* mutant were not significantly different from the results observed in the parent strain, indicating that *sefD* is unlikely necessary for egg contamination. Interestingly, the *sefD* complement strain containing a plasmid-borne *sefD* was attenuated, suggesting that expression levels of *sefD* may play a role in mitigating signs of illness in the chicken.

Introduction

Salmonella enterica subspecies *enterica* serotype Enteritidis (*S.* Enteritidis) is the leading cause of salmonellosis world-wide and has been the top reported *Salmonella* serotype in the United States since 2007 according to the Center for Disease Control and Prevention's Foodborne Diseases Active Surveillance Network (4, 19). Food vehicles for which *S.* Enteritidis has been determined to be the etiological agent of an illness or outbreak include fruits, vegetables, nuts, poultry, raw meat, and dairy products. However, infection from the ingestion of raw or undercooked eggs, egg products or foods containing such eggs, is the primary reason *S.* Enteritidis remains a major human health concern (24, 28). The definitive mechanism for its unique ability to internally contaminate an intact egg while mitigating signs of illness in the hen remains unknown.

The role of subpopulation heterogeneity has been used to study the egg contamination pathway of *S.* Enteritidis (18, 27). Two subpopulations of *S.* Enteritis PT13a isolated from different organs of the same naturally infected farm mouse were found to vary in their ability to contaminate the egg. Chemotyping revealed cell-surface lipopolysaccharide (LPS) heterogeneity between the two subpopulations. The egg-contaminating isolate was able to produce high-molecular-mass (HMM) LPS, defined as over 50% LPS molecules containing ten or more repeating O-chain units, and the non-egg-contaminating isolate produced low-molecular-mass (LMM) LPS, defined as 90% of LPS molecules containing an average of 5 repeating O-chain units (21). The ability to produce HMM LPS correlated with high cell density growth and high-incidence of egg contamination when hens were challenged intravenously. Other investigators later confirmed that LPS synthesis of a capsular-like O-antigen was necessary for survival in the egg albumen (16, 18). However, the HMM LPS-producing subpopulation that made the

capsular-like O-antigen was unable to infect eggs at high incidence when hens were challenged orally. It was also unable to produce a biofilm phenotype on agar media under ambient conditions. In contrast, the LMM LPS-producing strain showed strong systemic invasion when hens were dosed orally and it displayed a biofilm phenotype on agar media. Phenotype Microarray analysis comparing the two subpopulations found growth differences in over 20% of the test conditions. When combined into an inoculum for low-dose contact infection, the subpopulations appeared to complement each other and resulted in high-incidence egg contamination (19, 22).

In contrast to studies evaluating genes unique to *S.* Enteritidis by comparing it to other *Salmonella* serotypes or ever other genera, the genomes of subpopulations within the same phage type were nearly clonal, thus genetic noise was minimized and the strains were theoretically ideal for obtaining meaningful data from genomic comparisons (7, 26, 36). Recently, the genomes of the two subpopulations were re-sequenced and a total of 247 DNA polymorphisms differentiating the two isolates were found (Chapter 3). A majority of the polymorphisms resulted in either single amino acid substitutions, were synonymous, or were found in intergenic regions. Changes in protein function or expression due to such polymorphisms are not straight-forward to predict. However, twelve polymorphisms were predicted to disrupt 13 open reading frames (ORFs) and, thus, these were naturally occurring gene knockouts. Nine of the identified ORF-disrupting polymorphisms were found in the non-egg-contaminating subpopulation. Two, in genes *sen4316* and *dsdA*, were readily identified as disruptions likely to be associated with two previously observed phenotypes differentiating the *S.* Enteritidis subpopulations based on references associated with other gram negative pathogens (17, 30, 33-34). The gene *sen4316* encodes a protein with diguanylate cyclase activity, which synthesizes a regulatory signal

molecule, cyclic-di-GMP. It has been associated with biofilm formation (33-34) and was absent in one of the *S.* Enteritidis subpopulations that cannot produce biofilm. The gene *dsdA* encodes D-serine dehydratase, which is necessary for metabolizing D-serine, the enantiomer of L-serine. The *S.* Enteritidis PT13a subpopulation with the *dsdA* deletion was unable to utilize D-serine as a carbon or nitrogen source as determined by Phenotype Microarray (27). The non-egg-contaminating subpopulation contained a notable ORF-disrupting, single base pair deletion in *sefD*, a gene predicted to be minor subunit of an fimbria unique to *S.* Enteritidis.

Characterizing the observed nucleotide polymorphisms between two *S.* Enteritidis subpopulations that naturally vary in phenotype and the ability to contaminate the egg has the potential to identify genetic determinants for egg contamination in a background with minimal genetic noise. The objective of this study was to characterize three selected polymorphisms out of the 247 previously identified that differentiate the two subpopulations. Using mutagenesis to create deletion knockout mutants that mimic the observed natural ORF-disruptions, the genes *sen4316*, *dsdA*, and *sefD* were characterized by Phenotype Microarray and other analyses appropriate for their predicted phenotype changes. The gene *sefD* was further analyzed in a hen infection study to determine if its disruption in the non-egg-contamination subpopulation isolate is responsible for its inability to contaminate the egg. This study also established the methodological template for characterizing the remaining ORF-disrupting polymorphisms between the two subpopulations.

Materials and methods

Strains used for analysis. All strains used and generated in this study are listed in Table 4.1. *Salmonella* Enteritidis PT4 isolate 22079 was selected to be the parent strain for all subsequent mutagenic analyses, because it was previously used as the reference strain to which

subpopulation isolates were compared. Reference strain *S.* Enteritidis PT4 22079 is effectively the wild type strain as compared to both *S.* Enteritidis PT13a subpopulation isolates (Chapter 3). It is a field isolate obtained from a farm environment in California from an index outbreak that appeared to introduce this phage type into chicken flocks in the United States. It is capable of forming a biofilm under ambient conditions, produces high molecular mass LPS, and results in high-incidence of egg contamination in hen experiments (27). The ability to form biofilm was determined as follows. Culture was spread on brilliant green agar (Acumedia-Neogen, Lansing, MI) at a density of between 25 and 50 CFU and then incubated for 72 h at temperatures between 23 and 25°C. Colonies that produced a distinctive organic matrix was indicative of biofilm formation.

Plasmids. Plasmids used for mutagenesis were originally described by Datsenko and Wanner (11). Plasmid pKD4 was used as template DNA for generating site-specific kanamycin resistance gene cassette PCR products to be used in the gene knockout protocols. It contains a polar kanamycin resistance gene flanked by flippase recognition target (FRT) sites and a specific DNA sequence that when excised by flippase, creates a non-polar mutation by introducing stop codons in all six reading frames and a ribosomal binding site for downstream translation. Plasmid pKD46 was used as a helper plasmid to facilitate the recombination of the PCR-generated site-specific kanamycin cassette with its homologous gene target on the *S.* Enteritidis parent strain 22079 genome. It contains the phage lambda Red locus *bet*, *exo*, and *gam* with expression under the control of an arabinose-inducible promoter, as well as the beta-lactamase gene, *bla*, for ampicillin selection and a temperature sensitive replicon for simple removal from the host. Plasmid pCP20 was used as a helper plasmid to excise the polar kanamycin cassette resulting in a non-polar knockout mutation. It contains a flippase gene under the control of a

temperature sensitive promoter for simple induction. Plasmid pCP20 also contains an ampicillin resistance cassette for selection and a temperature sensitive replicon for quick removal from the host. Plasmid pCR2.1-TOPO (Invitrogen, Carlsbad, CA) was used as a vector for complementation of the mutant strains. It contains the α-fragment of β-galactosidase with the upstream *lac* promoter but without the *lacI* repressor so that inserts may be constitutively expressed. It also contains ampicillin and kanamycin resistance cassettes for selection. Derivative plasmids of pCR2.1-TOPO are noted in Table 4.1 and contain full, intact genes of *sefD*, *dsdA*, and *sen4316*.

Nucleic acid isolation. Single colonies of *S.* Enteritidis were grown in 10 ml of brain heart infusion (BHI) broth (Difco BD, Franklin Lakes, NJ) at 37°C for 16 h. Broth media was supplemented with the appropriate antibiotic (50 µg/ml kanamycin or 200 µg/ml ampicillin) when needed. Approximately 10^8 bacterial cells (1 ml of culture) were pelleted in an Eppendorf microcentrifuge (Centrifuge 5417C) at 5000 x *g* for 10 min. Total DNA was extracted using a Qiagen DNeasy Tissue kit following the protocol designated for bacteria (Qiagen, Valencia, CA). DNA was eluted in 200 µl of AE buffer (10 mM Tris-HCl, 0.5 mM EDTA [pH 9]) and stored at -20°C. Concentration and purity of DNA samples were measured using a NanoDrop 1000 (Wilmington, DE) spectrophotometer measuring optical densities at wavelengths of 230, 260, and 280 nm. $OD_{260/280}$ ratios were ensured to be greater than 1.70 before further processing.

Plasmids were isolated using a Qiagen Spin QIAprep Miniprep kit. Single colonies were inoculated in 10 ml Luria-Bertani (LB) broth (10 g tryptone, 5 g yeast extract, and 10 g NaCl per L) supplemented with 50 µg/ml kanamycin or 200 µg/ml ampicillin when appropriate and incubated at 37°C with agitation for 16 h. One milliliter of culture was processed as per

manufacturer's instructions for plasmid isolation. Plasmid DNA was eluted in 50 µl EB (10 mM Tris·Cl [pH 8.5]) and stored at -20°C.

Polymerase chain reactions. All primers used for polymerase chain reactions (PCR) are listed in Table 4.2. Synthetic oligonucleotide primers were designed using PrimerQuest (Integrated DNA Technologies, http://www.idtdna.com) and all PCR reactions followed recommended parameters and annealing temperatures based PrimerQuest calculations. Primers for generating site-specific kanamycin knockout cassettes were designed to contain 40 bp homology to the *S.* Enteritidis 22079 target gene at the 5' end of the primer and 20 bp homology to the kanamycin cassette at the 3' end. Polymerase chain reactions included 100 nM of each primer and 1X concentration of Applied Biosystems (Carlsbad, CA) GeneAmp Fast PCR Master Mix in a final reaction volume of 20 µl. The PCR reactions were performed in an Applied Biosystems Veriti 96 Well Fast Thermal Cylcer system using annealing temperatures calculated from PimerQuest and default cycle conditions for Applied Biosystem's Fast PCR protocol. Single-band, amplified PCR products of expected size were confirmed by applying an unspecified, proprietary amount of voltage on Invitrogen precast 2% agarose E-Gel48 gels (Carlsbad, CA) for 20 min on an Invitrogen E-Gel electrophoresis system and visualizing under ultraviolet light of 302 nm wavelength. Amplicons created to be site-specific kanamycin resistance knockout cassettes were isolated and purified using Qiagen's QIAquick Gel Extraction kit. Confirmation of successful recombinations with the knockout cassette, excisions of the knockout cassette, and the generation of gene inserts for complementation were performed by PCR. The genotypes of *S.* Enteritidis isolates from the hen infection study were also confirmed by using PCR.

Mutagenesis of *sefD, dsdA*, and *sen4316*. Mutagenesis was performed using the λ Red system described by Datsenko and Wanner (11). *Salmonella* Enteritidis PT4 parent strain 22079 was made electrocompetent as previously described (34). In brief, a 1 ml starter culture in LB broth grown overnight at 37° C was used to inoculate 100 ml of LB broth and incubated at 37° C with aeration until an OD_{600} of 0.6 was reached. The suspension was chilled on ice for 20 min and then centrifuged at 4000 x g for 15 min at 4°C in a Sorval Ultracentrifuge (Asheville, NC). The cells were washed in subsequent steps with 80 ml, 40 ml, and 20 ml of ice-cold sterile double distilled water and resuspended in 500 µl of sterile ice-cold double distilled water. Plasmid pKD46 was electroporated into *S.* Enteritidis 22079 competent cells using a Bio-Red Gene Pulser Xcell electroporation System (Hercules, CA) according to the manufacturer's instructions (2,500 kV, 200 Ω, 25 µF). Cells carrying pKD46 were selected by plating on LB agar supplemented with 100 µg/ml ampicillin overnight at 30°C. Confirmed transformants were frozen in 20% glycerol at -80°C and assigned with accession number 28000.

Plasmid pKD4 was used as template plasmid to create knockout kanamycin resistance cassettes targeting *sefD, dsdA*, and *sen4316,* separately. The parent strain carrying pKD46 (28000) was made elecrocompetent as described above with the addition of 0.08% L-arabinose to induce the recombinase genes and 100 µg/ml ampicillin in the LB broth culture. Site-specific kanamycin knockout cassettes were electroporated as described above into the 28000 strain and recombinants were selected by growing on LB agar supplemented with 25 µg/ml kanamycin overnight at 37°C. Plasmid pKD46 was then cured from transformants by growing on LB/kanamycin (25 µg/ml) agar overnight at 42°C. Knockout mutations were confirmed by PCR looking for the correct expected insert size. Confirmed mutants for *sen4316::kan, dsdA::kan,*

and *sefD::kan*, were frozen down in 20% glycerol and assigned accession numbers 28017, 28025 and 28029, respectively.

Kanamycin cassettes were excised using plasmid pCP20 containing a gene for flippase. Mutant strains 28017, 28025, and 28029 were made electrocompetent as described above and transformed with pCP20. Transformants were selected and flippase was induced by growing on LB agar with100 µg/ml ampicillin overnight at 30°C. Plasmid pCP20 was cured from transformants by growing overnight on LB agar with no antibiotic supplementation at 42°C. Transformants were confirmed to not grow on LB agar supplemented with either ampicillin or kanamycin and cassette excisions were confirmed by PCR. Confirmed mutants for *ΔdsdA*, *ΔsefD*, and *Δsen4316* were frozen down in 20% glycerol and assigned accession numbers 29106, 29108, and 100426, respectively.

Complementation. Mutant strains 29108, 29106, and 100426 were complemented by inserting their respective intact genes into plasmid pCR2.1-TOPO. Polymerase chain reaction generated ORFs for *sefD*, *dsdA*, and *sen4316* were ligated with pCR2.1-TOPO as per manufacturer's instructions in-frame with the N-terminus of the *lacZα* fragment under control of the *lac* promoter. Complement plasmids were electroporated and transformed into their respective mutant strains as described above. Successful transformants were selected under high antibiotic selection pressure (200 µg/ml ampicillin) to result in high-copy plasmid numbers with the cells. Confirmation of the gene in the correct orientation was performed by submitting isolated plasmid for sequencing to Retrogen, Inc. (San Diego, California) using universal M13 sequencing primers. Confirmed complementation strains for *ΔdsdA*, *ΔsefD*, and *Δsen4316* were frozen in 20% glycerol and assigned accession numbers 29107, 100126, and 100627, respectively.

Phenotype Microarray. The three mutant strains containing *ΔsefD*, *ΔdsdA*, and *Δsen4316* mutations, respectively, were analyzed for phenotype changes compared to the parent strain using Phenotype Microarray (Biolog; Hayward, CA). The array measures cellular respiration by electron transfer in twenty 96 well plates of 1,920 different conditions representing carbon, nitrogen, phosphorous, and sulfur utilization, response to nutrient stimulation compounds, osmotic and pH stress, and sensitivity to antimicrobials and other inhibitory compounds (5). The Phenotype Microarrays (PM) were performed as previously described (40). Inoculated PM panels were incubated at 37°C for 24 h and performed in duplicate for each strain. Data was collected as an arbitrary respiratory activity unit measured every 15 min within the 24 h incubation period, effectively creating a growth curve for each of the 1,920 test conditions. Height of the growth curve (based on respiratory activity units) was used as the comparison parameter between the parent and mutant strain. Loss of a phenotype was defined as 80% or greater reduction in respiratory activity compared to the parent strain.

Infection of hens. Specific-pathogen-free mature leghorn hens between 30 to 35 wk of age housed singly in layer cages were used to compare egg production and calcium blood levels of the hens when challenged with parent strain 22079, the *ΔsefD* mutant strain 29108, and the *ΔsefD* pCR2.1-TOPO Φ(*lacZα'-sefD*$^{+}$) complement strain 100126. Four rooms were used with a minimum of 12 hens and three of the treatment groups were challenged with 10^{7} CFU suspended in 0.5 ml of phosphate buffered saline (PBS) (Sigma-Aldrich, 10mM phosphate, 138mM NaCl, 2.7mM KCl [pH 7.4]) of the parent, mutant, and complement strains, respectively, by intramuscular (IM) infection. The hens in the fourth room were IM injected with 0.5 ml of PBS. Egg production was recorded for 14 d pre-infection and 25 d post-infection for all four rooms, because production suppression was previously correlated with egg contamination under these

conditions (18). Blood serum was collected from three birds in each room starting on the day of infection for 21 days and samples were submitted to Athens Veterinary Diagnostic Labs (Athens, GA) for blood calcium level measurements. Statistical significance of differences in egg production and blood calcium levels were determined by the Student's T-test. The hens were euthanized 26 days post infection. Necropsies were performed to retrieve spleen, cecal contents, and intestinal contents from each surviving bird to isolate and confirm the genotype of the challenge strain. All procedures involving hens were approved by a facility Institutional Animal Care and Use Committee.

Culture of *S.* Enteritidis from spleen, cecal and intestinal contents. Methods for *Salmonella* Enteritidis isolation from spleen, cecal and intestinal contents were previously described (20). In brief, spleen samples were homogenized by stomaching and added to 20 ml of tryptic soy broth (Difco BD, Franklin Lakes, NJ) for enrichment. Cecal and intestinal contents were added directly to 20 ml of tryptic soy broth, separately, and incubated at 37°C for 48 h. One milliliter of each enrichment culture was transferred to 10 ml of Rappaport-Vassiliadis broth (Difco BD) for incubation at 37°C for 24 h. Single colonies were obtained using 10 µl inoculation loops to streak the cultures onto Brilliant Green agar (Acumedia; Lansing, MI) and incubated at 37°C for 24 h. Colonies were confirmed as *Salmonella* species using Enterotube II (Difco BD) and further processed for genotype confirmation by PCR using the appropriate primers listed in Table 4.2.

Results

Mutagenesis. Non-polar knock out mutations were created for the genes *sen4316*, *dsdA*, and *sefD* with the intention of mimicking the equivalent natural ORF-disruptions identified in two *S.* Enteritidis PT13a subpopulations (Chapter 3). All knockout genotypes were confirmed

by PCR based on their expected size at each step in the process. Kanamycin cassettes were initially introduced into the gene via homologous recombination to create the knockout but have the potential to create polar effects on downstream genes in the same operon (25). Therefore, to ensure non-polar mutations, the kanamycin cassettes were excised by the flippase activity provided by helper plasmid pCP20. The mutations were introduced into *S.* Enteritidis PT4 strain 22079 since it was originally used as the reference strain to which the two PT13a subpopulations were compared, representing the wild type genome. Thus, the natural ORF-disrupting polymorphisms found in either of the two subpopulations could be introduced into the one reference strain for analysis rather than introducing mutations into multiple strains. This strategy was additionally beneficial since one of the two subpopulations contained antibiotic resistances that would have made selection for transformants more difficult. Complementation was performed for all mutants by introducing a plasmid-encoded functional gene into the respective mutant strain to assess if the wild type phenotype was restored.

Analysis of Δsen4316. The open reading frame *sen4316* is predicted to encode a protein with diguanylate cyclase activity. Diguanylate cyclase synthesizes 3 -5 -cyclic diguanylic acid (c-di-GMP), which is a regulatory signal involved in virulence, motility, long-term survival, and cellulose and fimbriae synthesis needed for biofilm formation (33-34). A 215 bp deletion incorporating the first 80 bp of *sen4316* was found in the *S.* Enteritidis PT13a subpopulation that does not exhibit a biofilm morphotype on agar media. Therefore, it was predicted that the deletion resulted in the lack of biofilm formation (Chapter 3) for this subpopulation. Morphology analysis of the *Δsen4316* mutant showed inhibition of biofilm formation compared to the parent strain. The complement strain 100627 carrying a plasmid-borne *sen4316* displayed a restored biofilm phenotype on BG agar. However, no significant loss or gain of phenotype in

the 100426 *Δsen4316* mutant strain was identified in Phenotype Microarray analysis compared to the parent strain. The lack of a phenotype change may be explained by *Salmonella* Enteritidis containing 11 additional ORFs with diguanylate cyclase activity and, thereby, redundancy in c-di-GMP synthesis results in no change to cell function (33). It is also possible for a phenotype change to exist, in addition to the biofilm phenotype differences, that cannot be assayed by Phenotype Microarray.

Anaylsis of ΔdsdA. The biofilm-forming *S.* Enteritidis PT13a subpopulation was recently found to contain a 10 base pair frame shift deletion in *dsdA* (Chapter 3), presumably responsible for its inability to metabolize D-serine as previously observed in PM analysis (27). Introducing the mutation in parent strain 22079 resulted a similar phenotype. Like the biofilm-forming PT13a subpopulation the PM analysis performed showed that the *ΔdsdA* mutant strain 29106 was unable to utilize D-serine as the sole source of carbon or nitrogen, nor was it able to grow in the presence of D-serine as an inhibitor. An additional phenotype change that was observed was increased sensitivity to osmolytes, as well as to the antimicrobial polymyxin B. The change in respiratory activity compared to the parent strain can be seen in Table 4.3. The complement strain 29107 with a plasmid-borne *dsdA* gene was able to restore the original parent strain phenotypes for both D-serine utilization and osmotic sensitivity. While the mode of action for D-serine metabolism is known since *dsdA* encodes for D-serine dehydratase (3), the mechanism for the observed change in osmotic sensitivity is unclear.

Analyis of ΔsefD. The gene *sefD* is the fourth gene of an operon that encodes an *S.* Enteritidis-specific fimbria, SEF14. The exact function of SefD has not been agreed on as separate studies hypothesize that it is the major fimbrin unit of a distinct fimbria (10), the minor subunit adhesin tip of SEF14 (13), or a minor fimbrin subunit of SEF14 inserted at intervals and

acts as a branching molecule (6). It has been demonstrated, however, that SefD is required for proper assembly of SEF14. Thereby, *sefD* mutants have been previously shown to lack SEF14 fimbriae altogether (6, 13). The non-egg-contaminating *S.* Enteritidis PT13a subpopulation was previously found to have a single base pair frame-shift deletion in *sefD* (Chapter 3) and, thus, was predicted to not produce the SEF14 fimbria. Phenotype Microarray analysis of the *ΔsefD* mutant strain 29108 revealed no differences to the parent. This was not unexpected since SEF14 is presumed to have a role in attachment or invasion rather than in the class of conditions assayed in the Phenotype Microarray. Given that *S.* Enteritidis has a unique ability to contaminate the internal contents of the egg and that SEF14 is a fimbria unique to *S.* Enteritidis, the *ΔsefD* mutant strain was further characterized in an egg production animal experiment.

Hen infection. Egg production of hens over the course of 37 d in four rooms pre and post-challenged with the parent strain 22079, *ΔsefD* mutant strain 29108, complement strain 100126, and PBS control, separately, can be seen in Figure 4.1. Significant depressions in egg production ($P < 0.001$) after the challenge were observed for hens infected with the parent 22079 strain and the *ΔsefD* mutant strain 29108 compared to uninfected hens. Egg production between hens infected with the parent and mutant strain was not significantly different ($P = 0.10$), indicating no change in reproductive tract tropism between the two strains. Unexpectedly, egg production from hens challenged with the complement strain 100126 was not significantly different from the control hens ($P = 0.12$), suggesting *S.* Enteritidis attenuation when *sefD* is introduced on a plasmid. Preliminary real-time reverse-transcriptase PCR analysis on *sefD* expression of the complement strain compared to the parent strain suggested that *sefD* is significantly over-expressed in the complement strain (data not shown).

Blood calcium levels measured from serum samples throughout the challenge study correlated with the egg production results. Results can be seen in Figure 4.2. The birds infected with the parent strain 22079 and *ΔsefD* mutant strain 29108 exhibited significant drops in blood calcium levels compared to the control birds ($P < 0.005$). Hypocalcemia can be a symptom of sepsis, suggesting that hens infected with the parent and *ΔsefD* strains were bacteremic (39). Blood calcium levels between the parent and mutant strain were not significantly different after infection ($P = 0.36$). Similarly to what was observed in the egg production results, the blood calcium levels of hens infected with the complement strain 100126 were not significantly different from the control birds ($P = 0.33$), further suggesting a mechanism of attenuation associated with plasmid-borne *sefD*.

Salmonella was recovered from spleen, ceca, and intestinal contents of 67% (8 out of 12 hens) recovery of the hens challenged with the parent strain, 70% (7 out of 10 hens) from birds infected with the *ΔsefD* mutant strain, 31% (4 out of 12 hens) from birds challenged with the complement strain, and 0% from the control birds. No *Salmonella* isolates were found in the intestinal contents for any of the different challenges and only one ceca sample (parent strain challenged) was found to contain *Salmonella* Enteritidis. The rest of the *Salmonella* positive samples were spleen, indicative of a long-term persistence association with the organ, which has previously been demonstrated in *Salmonella* Pullorum and *S.* Enteritidis (23, 37). Genotyping of the recovered samples was performed by PCR to confirm isolation of the expected challenge strain. The isolates recovered from the birds challenged with the parent strain and *ΔsefD* strain contained the expected genotype. Isolates recovered from the birds challenged with the complement strain contained the *ΔsefD* mutant genotype, suggesting that only the cells that were successful at curing the plasmid carrying *sefD* were capable of resulting in a systemic infection.

Discussion

In summary, select mutants were created based on a previously established database of sequenced polymorphisms between two subpopulations of *S.* Enteritidis PT13a that differ in their ability to contaminate the egg (Chapter 3). Mutant knockout strains for *sen4316* and *dsdA* were initially constructed as a proof-of-concept model for mutagenesis in *S.* Enteritidis PT4 strain 22079 since their expected phenotype changes were relatively simple to screen in comparison to putative cell-surface proteins or ORFs of unknown function. All mutant strains were analyzed for changes in respiratory activity in 1,920 different conditions using Phenotype Microarray technology. Expected phenotype changes for the *Δsen4316* and *ΔdsdA* mutant strains were observed, though the *ΔdsdA* strain also exhibited unexpected osmolyte sensitivity. Mutant strain 29108, a *ΔsefD* knockout mutant, was further characterized in a hen infection study for a potential change in egg production compared to its parent strain. The removal of *sefD* had no significant effect on egg production compared to the parent strain; however, increased copy number of *sefD* appeared to attenuate the strain to control levels.

The mutagenesis analysis of *sen4316* confirmed its involvement in biofilm formation. Diguanylate cyclase activity observed from the GGDEF domain is responsible for synthesizing cyclic di-GMP, which has been shown to be required for cellulose synthesis, as well as involved in regulating developmental transitions, aggregative behavior, adhesion, biofilm formation, and the virulence of animal and plant pathogens (31, 33, 35). However, *sen4316* is one of 12 putative proteins containing a GGDEF domain in *S.* Enteritidis, so it is unclear if *sen4316* serves a role specific to biofilm synthesis or if the lack of *sen4316* reduces overall levels of cyclic-di-GMP to be below a threshold necessary to induce biofilm formation. One previous study showed that complementing a naturally biofilm-negative *S.* Typhimurium strain with *stm1987*, another gene

with diguanylate cyclase activity, restored the strain's ability to form biofilm (17). However, an alternate study that deleted all genes with diguanylate cyclase activity in *S.* Enteritidis found that a complemented *sen4316* to be the only gene that restored all phenotypes associated with cyclic-di-GMP regulation, including virulence in mice (33). Both studies also performed various permutations of different GGDEF gene knockouts and complementations that removed and restored similar phenotypes to demonstrate redundancy in the regulatory networks controlled by cyclic-di-GMP. Therefore, it is unlikely that *sen4316* has an exclusive role in biofilm formation in *S.* Enteritidis, but rather that the ability to form biofilm is influenced by the levels of cyclic-di-GMP. The levels of GMP, in turn, can be affected by ORF-disruptions in proteins containing GGDEF domains, such as the *sen4316* disruption observed in the original biofilm-negative *S.* Enteritidis PT13a subpopulation isolate 21046 (Chapter 3). The lack of any observable phenotype changes in the Phenotype Microarray could be explained by redundancy in diguanylate cyclase activity as a result of 12 genes with a GGDEF domain in the gnome or that any other cellular changes in addition to biofilm formation may not be assayed by the array. Any potential changes in virulence in the hen or egg contamination caused by the deletion of *sen4316* in *S.* Enteritidis has yet to be determined.

Mutagenesis analysis of *dsdA* in *S.* Enteritidis PT4 22079 produced the expected result of altering the utilization of D-serine as nitrogen or carbon, as well as the survival in the presence of D-serine. Additionally, the mutant phenotype displayed increased sensitivity to osmotic stress and multiple osmolytes. It is unclear if the osmotic sensitivity phenotype is an artifact of the *dsdA* deletion since it is possible for truncated proteins to alter member permeability (14). However, DsdA is expected to be a cytoplasmic protein and the complement strain 29107 with a plasmid-borne *dsdA* gene, which still contains the knockout *ΔdsdA* in the chromosome, restored

osmotic sensitivity to the parent strain levels, suggesting a possible role in resistance to osmotic stress. The ability to metabolize D-serine could also facilitate egg contamination. D-serine is known to be present in mammalian urine and the presence of *dsdA* in uropathogenic *Escherichia coli* has been shown to aid in its virulence (30). Levels of D-serine within the tissues of hens have not been reported, but if present, it is possible that D-serine in the vaginal region of the oviduct would select for microorganisms carrying *dsdA*. Colonization of the vagina is correlated with egg-contamination, though predominantly through exterior egg-shell contamination by which bacteria can still penetrate the shell via temperature and pressure differentials (15).

The fimbrial subunit protein SefD is necessary for the assembly of *S.* Enteritidis-specific fimbria SEF14 (13). Therefore, studies that created insertional mutants in sefD are functionally equivalent to studies that knockout the gene encoding the SEF14 major fimbrial unit, *sefA*. Previous animal infection studies by different research groups have produced conflicting conclusions on the role of SEF14 in virulence. Allen-Vorcoe and Woodward found no significant differences in a chicken *in vitro* gut adherence assay between a wild-type *S.* Enteritidis strain and its SEF14-deficient mutant (2). Similarly, Dibb-Fuller et al. observed that a SEF14-deficient mutant had no change in its ability to invade cultured epithelial cells compared to its wild-type strain (12). Furthermore, additional studies demonstrated that SEF14-deficient mutants invade Caco-2 and HT enterocytes at levels equal their wild-type parent, as well as exhibited no difference in survival in mouse peritoneal macrophages (29, 36). Multiple animal oral infection studies in chicken and mice resulted in no significant difference between SEF14 deficient mutants and the wild type strain (8, 36). In contrast, Rajashekara et al. observed low numbers of SEF14-deficient mutants recovered from the liver and spleens of chickens compared to the wild-type (29). Additional virulence studies found sefD knockout mutants to be attenuated

with a higher LD_{50} in mice and 1-day old chicks compared to the wild-type strains (13, 32). Edwards et al. concluded that SEF14 was required for uptake in peritoneal macrophages in chickens (13), although Thorns et al. concluded that SEF14 was not required in peritoneal macrophages in mice (36), suggesting that the role of SEF14 may be species-specific. The conflicting results of past studies may be explained by experimental variables such as different animal model species and age, infection method, and the specific strain used for analysis. It is possible that setting up a similar hen infection study with the *ΔsefD* mutant using an alternate infection route, such as oral, could indirectly result in different egg production results if SefD is necessary for S. Enteritidis to cross the intestinal barrier and, therefore, cannot become systemic. However, the results of this study reveal that the *ΔsefD* mutant strain alters egg production no differently than the parent strain, indicating that *sefD* is does not play a direct role in interacting with the reproductive tract.

The *sefD* complement strain 100126 had the unexpected result of producing attenuation in regards to blood calcium and egg production. It did not mirror levels equivalent to the parent strain. The plasmid vector carrying the complemented *sefD* gene contains a pUC origin of replication resulting in a high-copy number plasmid. Additionally, the *sefD* insert is expressed constitutively under the control of the *lac* promoter without the presence of the *lacI* inhibitor. Therefore, it is possible that over-expression of *sefD* compared to the parent strain resulted in the strain's attenuation. The exact mechanism of attenuation is unclear since there are at least three different proposed roles for SefD (6, 9, 13). Genomic comparisons of sequenced *Salmonella* serotypes in GenBank reveal that *sefD* is present in many serotypes as a non-functional pseudogene (9, 13). The results presented here suggest that *sefD* is not required for virulence and that over-expression of *sefD* suppresses virulence. It is possible that when expressed under

normal regulatory conditions, the presence of SefD mitigaes signs of illness, which is a notable characteristic of *S.* Enteritidis hen infections (18-19). Other *Salmonella* may have selected against carrying a functional *sefD* to not encumber virulence or to limit a host-immune response if SefD contains highly antigenic properties.

Another unexpected phenotype of the complement strain was that it was unable to produce a biofilm on agar medium under room temperatures between 20 to 25 °C (data not shown). In wild-type strains, SEF14 is up-regulated at 37 °C while SEF17 fimbriae, which are the "curli" fimbriae encoded by the *agf* operon and associated with biofilm formation, are up-regulated at 25 °C (1, 38). It is possible that plasmid-borne *sefD* in the complement strain, which would still be constitutively expressed at 25 °C, results in over-production of SEF14, interfering with SEF17 production and altering the biofilm phenotype. A study by Botten also noted altered fimbriae morphology when *sefD* was provided as plasmid-borne for complementation analysis (6). Further analysis is necessary to determine the effect of over-expression of *sefD* on biofilm formation.

While a direct correlation has yet to be found between a gene or genes and the differing abilities of the two *S.* Enteritidis PT13a subpopulations to contaminate the egg, insights were still gained on the three selected polymorphisms that disrupt open reading frames. Thus far, only *sefD* has been evaluated for its possible role in avian reproductive tract tropism. The genes *sen4316* and *dsdA* will undergo a similar study to evaluate whether or not they play a role in the egg contamination pathway, as will other mutations that disrupted ORFs differentially between the two subpopulations. The study presented herein represents the first stage in characterizing the effects of polymorphisms on subpopulations of *S.* Enteritidis.

Acknowledgements

Funding for this research was provided by USDA-ARS CRIS project no. 6612-32000-004-00. Robert Maier is thanked for providing strain 28006 harboring plasmid pKD4. Special thanks are given to Roxana Sanchez-Ingunza and Tod Stewart for their technical assistance in this research project.

References

1. **Allen-Vercoe, E., M. Dibb-Fuller, C. J. Thorns, and M. J. Woodward.** 1997. SEF17 fimbriae are essential for the convoluted colonial morphology of *Salmonella* Enteritidis. FEMS Microbiol Lett **153:**33-42.

2. **Allen-Vercoe, E., and M. J. Woodward.** 1999. The role of flagella, but not fimbriae, in the adherence of *Salmonella enterica* serotype Enteritidis to chick gut explant. J Med Microbiol **48:**771-80.

3. **Anfora, A. T., and R. A. Welch.** 2006. DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. J Bacteriol **188:**6622-8.

4. **Anonymous.** 2010. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food - 10 states, 2009. MMWR Morb Mortal Wkly Rep **59:**418-22.

5. **Bochner, B. R., P. Gadzinski, and E. Panomitros.** 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res. **11:**1246-55.

6. **Botten, J.** 2001. Role of *sefD* and *sefR* in the biogenesis of *Salmonella enterica* serovar Enteritidis SEF14 fimbriae. Ph.D. thesis. University of Adelaide, Australia.

7. **Clavijo, R. I., C. Loui, G. L. Andersen, L. W. Riley, and S. Lu.** 2006. Identification of genes associated with survival of *Salmonella enterica* serovar Enteritidis in chicken egg albumen. Appl Environ Microbiol **72:**1055-64.

8. **Clayton, D. J., A. J. Bowen, S. D. Hulme, A. M. Buckley, V. L. Deacon, N. R. Thomson, P. A. Barrow, E. Morgan, M. A. Jones, M. Watson, and M. P. Stevens.** 2008. Analysis of the role of 13 major fimbrial subunits in colonisation of the chicken intestines by *Salmonella enterica* serovar Enteritidis reveals a role for a novel locus. BMC Microbiol **8:**228.

9. **Clouthier, S. C., S. K. Collinson, and W. W. Kay.** 1994. Unique fimbriae-like structures encoded by *sefD* of the SEF14 fimbrial gene cluster of *Salmonella* Enteritidis. Mol Microbiol **12:**893-901.

10. **Collinson, S. K., S. L. Liu, S. C. Clouthier, P. A. Banser, J. L. Doran, K. E. Sanderson, and W. W. Kay.** 1996. The location of four fimbrin-encoding genes, *agfA*, *fimA*, *sefA* and *sefD*, on the *Salmonella* Enteritidis and/or *S.* Typhimurium XbaI-BlnI genomic restriction maps. Gene **169:**75-80.

11. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc Natl Acad Sci U S A **97:**6640-5.

12. **Dibb-Fuller, M. P., E. Allen-Vercoe, C. J. Thorns, and M. J. Woodward.** 1999. Fimbriae- and flagella-mediated association with and invasion of cultured epithelial cells by *Salmonella* Enteritidis. Microbiology **145 (Pt 5):**1023-31.

13. **Edwards, R. A., D. M. Schifferli, and S. R. Maloy.** 2000. A role for *Salmonella* fimbriae in intraperitoneal infections. Proc Natl Acad Sci U S A **97:**1258-62.

14. **Ferguson, A. D., J. Kodding, G. Walker, C. Bos, J. W. Coulton, K. Diederichs, V. Braun, and W. Welte.** 2001. Active transport of an antibiotic rifamycin derivative by the outer-membrane protein FhuA. Structure **9:**707-16.

15. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, R. Gast, T. J. Humphrey, and F. Van Immerseel.** 2009. Mechanisms of egg contamination by *Salmonella* Enteritidis. FEMS Microbiol Rev **33:**718-38.

16. **Gantois, I., R. Ducatelle, F. Pasmans, F. Haesebrouck, and F. Van Immerseel.** 2009. The *Salmonella* Enteritidis lipopolysaccharide biosynthesis gene *rfbH* is required for survival in egg albumen. Zoonoses Public Health **56:**145-9.

17. **Garcia, B., C. Latasa, C. Solano, F. Garcia-del Portillo, C. Gamazo, and I. Lasa.** 2004. Role of the GGDEF protein family in *Salmonella* cellulose biosynthesis and biofilm formation. Mol Microbiol **54:**264-77.

18. **Guard-Bouldin, J., R. K. Gast, T. J. Humphrey, D. J. Henzler, C. Morales, and K. Coles.** 2004. Subpopulation characteristics of egg-contaminating *Salmonella enterica* serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl. Environ. Microbiol. **70:**2756-63.

19. **Guard-Petter, J.** 2001. The chicken, the egg and *Salmonella* Enteritidis. Environ. Microbiol. **3:**421-30.

20. **Guard-Petter, J., D. J. Henzler, M. M. Rahman, and R. W. Carlson.** 1997. On-farm monitoring of mouse-invasive *Salmonella enterica* serovar Enteritidis and a model for its association with the production of contaminated eggs. Appl. Environ. Microbiol. **63:**1588-93.

21. **Guard-Petter, J., L. H. Keller, M. M. Rahman, R. W. Carlson, and S. Silvers.** 1996. A novel relationship between O-antigen variation, matrix formation, and invasiveness of *Salmonella* Enteritidis. Epidemiol. Infect. **117:**219-31.

22. **Guard-Petter, J., B. Lakshmi, R. Carlson, and K. Ingram.** 1995. Characterization of lipopolysaccharide heterogeneity in *Salmonella* Enteritidis by an improved gel electrophoresis method. Appl. Environ. Microbiol. **61:**2845-51.

23. **Guard, J., R. K. Gast, and R. Guraya.** 2010. Colonization of avian reproductive-tract tissues by variant subpopulations of *Salmonella* Enteritidis. Avian Dis **54:**857-61.

24. **Humphrey, T. J.** 1999. Contamination of eggs and poultry meat with *Salmonella enterica* serovar Enteritidis, p. 183-192. *In* A.M. Saeed, Gast, R.K., Potter, M.E., Wall, P.G. (ed.), *Salmonella enterica* serovar Enteritidis in humans and animals: epidemiology, pathogenesis, and control., 1st ed. Iowa State University Press, Ames.

25. **Ling, M. M., and B. H. Robinson.** 1997. Approaches to DNA mutagenesis: an overview. Anal Biochem **254:**157-78.

26. **Lu, S., P. B. Killoran, and L. W. Riley.** 2003. Association of *Salmonella enterica* serovar Enteritidis *yafD* with resistance to chicken egg albumen. Infect Immun **71:**6734-41.

27. **Morales, C. A., S. Porwollik, J. G. Frye, H. Kinde, M. McClelland, and J. Guard-Bouldin.** 2005. Correlation of phenotype with the genotype of egg-contaminating *Salmonella enterica* serovar Enteritidis. Appl. Environ. Microbiol. **71:**4388-99.

28. **Penteado, A. L., B. S. Eblen, and A. J. Miller.** 2004. Evidence of *Salmonella* internalization into fresh mangos during simulated postharvest insect disinfestation procedures. J. Food Prot. **67:**181-4.

29. **Rajashekara, G., S. Munir, M. F. Alexeyev, D. A. Halvorson, C. L. Wells, and K. V. Nagaraja.** 2000. Pathogenic role of SEF14, SEF17, and SEF21 fimbriae in *Salmonella enterica* serovar Enteritidis infection of chickens. Appl Environ Microbiol **66:**1759-63.

30. **Roesch, P. L., P. Redford, S. Batchelet, R. L. Moritz, S. Pellett, B. J. Haugen, F. R. Blattner, and R. A. Welch.** 2003. Uropathogenic *Escherichia coli* use d-serine deaminase to modulate infection of the murine urinary tract. Mol Microbiol **49:**55-67.

31. **Ryan, R. P., Y. Fouhy, J. F. Lucey, and J. M. Dow.** 2006. Cyclic di-GMP signaling in bacteria: recent advances and new puzzles. J Bacteriol **188:**8327-34.

32. **Shah, D. H., M. J. Lee, J. H. Park, J. H. Lee, S. K. Eo, J. T. Kwon, and J. S. Chae.** 2005. Identification of *Salmonella* Gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. Microbiology **151:**3957-68.

33. **Solano, C., B. Garcia, C. Latasa, A. Toledo-Arana, V. Zorraquino, J. Valle, J. Casals, E. Pedroso, and I. Lasa.** 2009. Genetic reductionist approach for dissecting individual roles of GGDEF proteins within the c-di-GMP signaling network in *Salmonella*. Proc Natl Acad Sci U S A **106:**7997-8002.

34. **Solano, C., B. Garcia, J. Valle, C. Berasain, J. M. Ghigo, C. Gamazo, and I. Lasa.** 2002. Genetic analysis of *Salmonella* Enteritidis biofilm formation: critical role of cellulose. Mol. Microbiol. **43:**793-808.

35. **Tamayo, R., J. T. Pratt, and A. Camilli.** 2007. Roles of cyclic diguanylate in the regulation of bacterial pathogenesis. Annu Rev Microbiol **61:**131-48.

36. **Thorns, C. J., C. Turcotte, C. G. Gemmell, and M. J. Woodward.** 1996. Studies into the role of the SEF14 fimbrial antigen in the pathogenesis of *Salmonella* Enteritidis. Microb Pathog **20:**235-46.

37. **Wigley, P., A. Berchieri, Jr., K. L. Page, A. L. Smith, and P. A. Barrow.** 2001. *Salmonella enterica* serovar Pullorum persists in splenic macrophages and in the reproductive tract during persistent, disease-free carriage in chickens. Infect Immun **69:**7873-9.

38. **Woodward, M. J., M. Sojka, K. A. Sprigings, and T. J. Humphrey.** 2000. The role of SEF14 and SEF17 fimbriae in the adherence of *Salmonella enterica* serotype Enteritidis to inanimate surfaces. J Med Microbiol **49:**481-7.

39. **Zaloga, G. P., and B. Chernow.** 1987. The multifactorial basis for hypocalcemia during sepsis. Studies of the parathyroid hormone-vitamin D axis. Ann Intern Med **107:**36-41.

40. **Zhou, L., X. H. Lei, B. R. Bochner, and B. L. Wanner.** 2003. Phenotype microarray analysis of *Escherichia coli* K-12 mutants with deletions of all two-component systems. J. Bacteriol. **185:**4956-72.

Figure Legend

Figure 4.1. Hen egg production analysis of the *ΔsefD Salmonella* Enteritidis mutant strain. Egg production measured over 39 days in mature White Leghorn hens challenged with the parent strain 22079, the *ΔsefD* mutant strain 29108, the complement strain 100126, and a negative control challenge. Trend lines were calculated using 4th polynomial analysis in Microsoft Office 2007 Excel spreadsheet software. The black bar in each strain's chart denotes the day of infection. The parent and mutant strain significantly altered egg production compared to the control challenge ($P < 0.001$), but not compared to each other ($P = 0.10$). The complement strain did not significantly alter egg production compared to the control challenge ($P = 0.12$).

Figure 4.2. Hen blood calcium levels analysis of the *ΔsefD Salmonella* Enteritidis mutant strain. Mature White Leghorn hens were challenged with the parent strain 22079, the *ΔsefD* mutant strain 29108, the complement strain 100126, and a negative control challenge. Blood calcium levels were measured post-infection over the course of 21 days from three birds in each room at the time of collection. The blood calcium levels of hens infected with the parent and mutant strains were significantly reduced ($P < 0.005$) compared to the control challenge, but not when compared to each other ($P = 0.38$). The blood calcium levels of hens infected with the complement strain were not significantly different from the control challenge ($P = 0.33$).

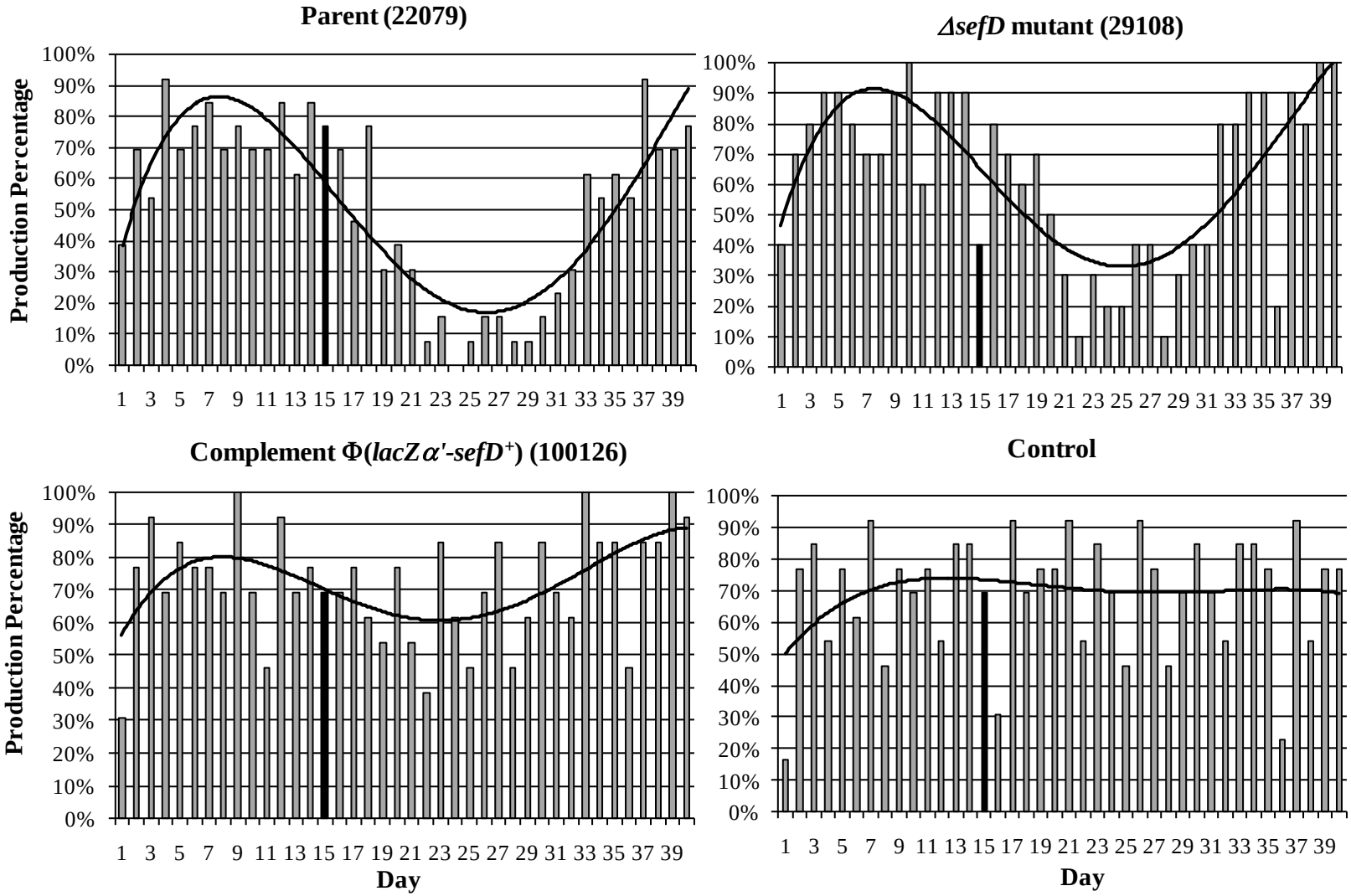
Parent (22079)
Δ*sefD* mutant (29108)
Complement Φ(*lacZα'-sefD*⁺) (100126)
Control
Production Percentage
Day
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39

Figure 4.1

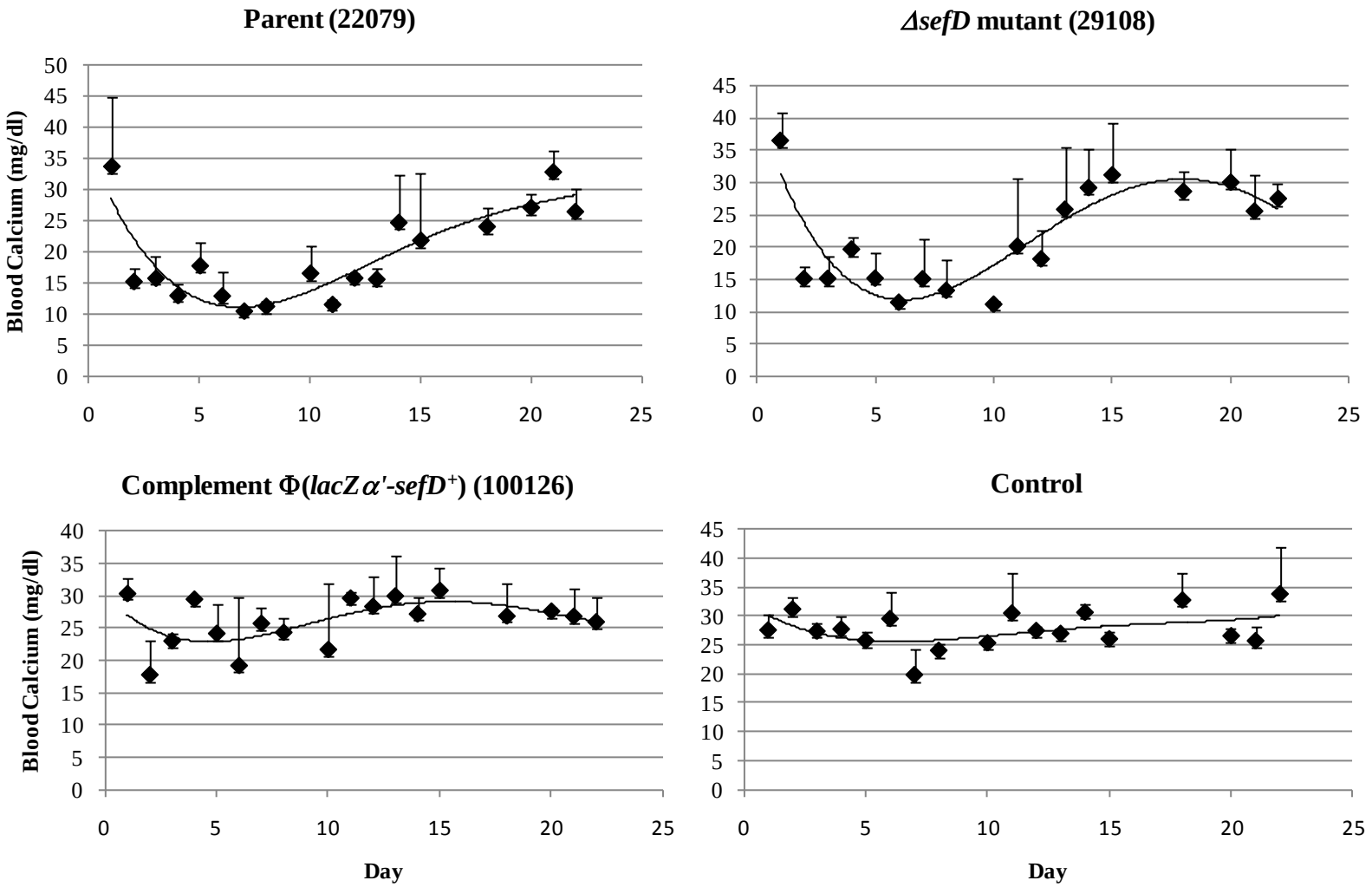
Parent (22079)
ΔsefD mutant (29108)
Complement Φ(lacZα'-sefD+) (100126)
Control
Blood Calcium (mg/dl)
Day

Figure 4.2

Table 4.1 Bacterial strains used and created for mutagenic characterization of *sen4316*, *dsdA*, and *sefD*

Accession number	Genotype	Description	Source
22079	Parent strain	*S.* Enteritidis phage type 4	Farm environment, California
27004	Carries pCP20	*E. coli* K-12 derivative	Salmonella Genetic Stock Centre
27005	Carries pKD46	*E. coli* K-12 derivative	Salmonella Genetic Stock Centre
28006	Carries pKD4	*E. coli* K-12 derivative	Robert Maier, Athens, GA
28000	Carries pKD46	22079 strain carrying pKD46	This study
28017	*sen4316::kan*	28000 with *sen4316::kan*, pKD46 removed	This study
28025	*dsdA::kan*	28000 with *dsdA::kan*, pKD46 removed	This study
28029	*sefD::kan*	28000 with *sefD::kan*, pKD46 removed	This study
29106	*Δ dsdA*	28025 with the kanamycin cassetted removed	This study
29107	pCR2.1-TOPO with *Φ(lacZ α'-dsdA$^+$)*	29106 complemented with functional *dsdA* inserted in pCR2.1-TOPO	This study
29108	*Δ sefD*	28029 with the kanamycin cassetted removed	This study
100126	pCR2.1-TOPO with *Φ(lacZ α'-sefD$^+$)*	29106 complemented with functional *sefD* inserted in pCR2.1-TOPO	This study
100426	*Δ sen4316*	28017 with the kanamycin cassetted removed	This study
100627	pCR2.1-TOPO with *Φ(lacZ α'-sen4316$^+$)*	100426 complemented with functional *sen4316* inserted in pCR2.1-TOPO	This study

Table 4.2 Polymer chain reaction primers designed to create and confirm knockout mutations for *sen4316* , *dsdA* , and *sefD*

Primers	Orientation	Primer sequence (5' to 3')
Knockouts		
sen4316	Forward	CTGACCAGAATGTGGTTACGTCCAAGATGCTTGGCTTCGTGTGTAGGCTGGAGCTGCTTC
	Reverse	TGATCGGGATAGTGGTCCTGAGCGTATTACTGTTGATATGCATATGAATATCCTCCTTAG
dsdA	Forward	ACCACCTGGTTTAACCCGGGCGCCACCTCTCTTGCACAAGGTGTAGGCTGGAGCTGCTTC
	Reverse	CCATACCAGATGCGTCGCGTTGCCGAGCTGCGTCTGACTGCATATGAATATCCTCCTTAG
sefD	Forward	TGAATCAGTATAATTCGTCAATACCTAAGTTCATTGTCTCGTGTAGGCTGGAGCTGCTTC
	Reverse	TATGCTTATTAAATATGTGTCAACAGGAATGTCTCCATTCCATATGAATATCCTCCTTAG
Confirmation		
sen4316	Forward	CTGACCAGAATGTGGTTACGTCCA
	Reverse	TGATCGGGATAGTGGTCCTGAGCG
sefD	Forward	TGAATCAGTATAATTCGTCAATACCTA
	Reverse	TATGCTTATTAAATATGTGTCAACAGG
dsdA	Forward	GTATCCTTTGGTCGAGGATCTGGT
	Reverse	TCCGGCACTGAGTGAGTATTAGC
Complementation		
sen4316	Forward	ATGACAACACCATCCTGGCGGTC
	Reverse	TCATAGGGCGCGCATGTCGTC
dsdA	Forward	ATGGAAAACATACAAAAGCTCATCGC
	Reverse	TTAGCGTCCTTTTGCCAGGTATTG
SefD	Forward	GTGCAAATGAATCAGTATAATTCGTC
	Reverse	TTATAATTCAATTTCTGTCGCATATATGC

Table 4.3. *Salmolnella* Enteritidis 29106 *ΔdsdA* mutant loss of phenotypes compared to the parent strain 22079 as determined by Phenotype Microarray

Compound	Difference[a]	Mode of action
D-Serine	-430.66	inhibits 3PGA dehydrogenase
Polymyxin B	-294.68	membrane, outer
Polymyxin B	-366.1	membrane, outer
pH 9.5 + Glycine	-164.7	pH, deaminase
D-serine	-188.51	nitrogen source
D-serine	-149.06	carbon source
Gly-D-ser	-129.84	nitrogen source
NaCl 3%	-150.53	osmotic sensitivity
NaCl 4%	-131.2	osmolyte
NaCl 5%	-74.45	osmolyte
NaCl 6% + Betaine	-109.31	osmolyte
NaCl 6% + N-N Dimethyl Glycine	-88.94	osmolyte
NaCl 6% + Sarcosine	-70.95	osmolyte
NaCl 6% + Dimethyl sulphonyl propionate	-78.35	osmolyte
NaCl 6% + MOPS	-76.63	osmolyte
NaCl 6% + Ectoine	-68.68	osmolyte
NaCl 6% + Choline	-43.36	osmolyte
NaCl 6% + L-proline	-72.08	osmolyte
NaCl 6% + N-Acethyl L-glutamine	-78.17	osmolyte
NaCl 6% + β-Glutamic acide	-68.14	osmolyte
Potassium chloride 4%	-142.75	osmotic sensitivity
Potassium chloride 5%	-120.16	osmotic sensitivity
Potassium chloride 6%	-118.3	osmotic sensitivity
Sodium sulfate 2%	-155.82	osmotic sensitivity
Sodium sulfate 3%	-160.58	osmotic sensitivity
Sodium sulfate 4%	-171.03	osmotic sensitivity
Sodium sulfate 5%	-177.49	osmotic sensitivity

[a]Difference values are measured as respiratory activity. Loss of phenoype is defined as an 80% or greater reduction in respiratory activity compared to the parent strain

CHAPTER 5

CONCLUSION

In summary, whole genome resequencing technology, combined with Sanger dye-terminator sequencing, was used to discover 247 distinct regions of chromosomal nucleotide polymorphisms between two subpopulations within *Salmonella enterica* serotype Enteritidis (*S.* Enteritidis) phage type (PT) 13a. The studied subpopulations were previously reported to vary in their lipopolysaccharide (LPS) structure, ability to produce biofilm, ability to achieve high-cell density growth, Phenotype Microarray results, and in their ability to contaminate table eggs. Presumably, the genetic determinants for the described phenotype differences are among the DNA polymorphisms reported in this study. Therefore, three polymorphisms that were predicted to disrupt their respective gene functions were selected for mutagenesis and characterization. The investigation of all polymorphisms is an objective for further research. The genes *sen4316* and *dsdA* were confirmed to be necessary for biofilm formation and D-serine metabolism, respectively. The *S.* Enteritidis unique fimbrial gene, *sefD*, was found not to be necessary for reproductive tract tropism in a hen egg production experiment; instead, it appears likely to play a role in mitigating disease depending on its expression.

Within the discovered 247 polymorphisms, twelve polymorphisms were predicted to disrupt open reading frames (ORFs), 99 were predicted to result in amino acid substitutions, 82 were predicted to occur within an open reading frame but were synonymous, 38 polymorphisms were found in non-coding intergenic regions, and 12 were found in RNA genes. Of the twelve ORF-disrupting polymorphisms, the biofilm-producing, non-egg contaminating *S.* Enteritidis

PT13a subpopulation contained nine and the non-biofilm-producing, egg-contaminating subpopulation contained three. Considering how the subpopulations complement each other to result in high-incidence of egg contamination, it is possible that each is evolving towards specific niches in the overall egg contamination pathway. This concept is further exemplified by how the *S.* Enteritidis PT4 reference strain used for analysis contains intact genes of those lost by either subpopulation and exhibits advantageous phenotypes from each subpopulation, such as biofilm production and the ability to contaminate the egg at high-incidence. It is thus a dimorphic organism, able to bridge multiple environments by having inherent plasticity encoded in its genome.

The study presented herein confirmed the genotype-phenotype relationship observed between the two *S.* Enteritidis 13a subpopulations for *sen4316* and *dsdA*. The gene *sefD* was found not to be directly involved in reproductive tract tropism but its over-expression was revealed to result in an avirulent phenotype, suggesting a possible role in modulating virulence to mitigate signs of illness in the chicken. Nonetheless, the conflicting conclusions of previous *sefD* studies warrant further study of the exact role of this fimbrial protein in *S.* Enteritidis. While the remaining polymorphisms will require additional characterization, this research confirms that small-scale nucleotide polymorphisms that are not detected by conventional typing methods can be responsible for diverse phenotypes between two subpopulations, including phenotypes that differ in their impact on human health.