


5-2012

Genome-wide Identification of Conditionally Essential Genes in Salmonella Typhimurium using Tn-Seq Method

Anita Khatiwara

University of Arkansas, Fayetteville

Follow this and additional works at: <http://scholarworks.uark.edu/etd>

 Part of the [Cell Biology Commons](#), [Molecular Biology Commons](#), and the [Pathogenic Microbiology Commons](#)

Recommended Citation

Khatiwara, Anita, "Genome-wide Identification of Conditionally Essential Genes in Salmonella Typhimurium using Tn-Seq Method" (2012). *Theses and Dissertations*. 274.

<http://scholarworks.uark.edu/etd/274>

This Dissertation is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu, ccmiddle@uark.edu.

**GENOME-WIDE IDENTIFICATION OF CONDITIONALLY ESSENTIAL GENES IN
SALMONELLA TYPHIMURIUM USING Tn-Seq METHOD**

GENOME-WIDE IDENTIFICATION OF CONDITIONALLY ESSENTIAL GENES IN
SALMONELLA TYPHIMURIUM USING Tn-Seq METHOD

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

By

Anita Khatiwara

Acharya N. G. Ranga Agricultural University
Bachelor of Veterinary Science and Animal Husbandry, 2001
University of Arkansas
Master of Science in Cell & Molecular Biology, 2007

May 2012
University of Arkansas

ABSTRACT

As more whole genome sequences become available, there is an increasing need for high-throughput methods that link genes to phenotypes and facilitate discovery of new gene functions. The objective of this study was to develop a high-throughput method to study gene functions in bacteria and use this method to study gene functions of *S. enterica* serotype Typhimurium (*S. Typhimurium*) under various environmental conditions encountered during its life cycle. Chapter I of this dissertation reviews the history and evolution of functional genomics in bacteria with focus on Salmonella, along with the recent techniques available. Chapter II, deals with the development of new version of Tn-seq (Transposon-sequencing) method involving a modified EZ:Tn5 transposon for genome-wide and quantitative mapping of all insertions in a complex mutant library utilizing massively parallel Illumina sequencing. The new version of Tn-seq method was applied to a genome-saturating *S. Typhimurium* mutant library recovered from selection under 3 different *in vitro* growth conditions (diluted LB medium, LB medium + bile acid, and LB medium at 42°C), mimicking some aspects of host stressors. We identified an overlapping set of 105 protein-coding genes in *S. Typhimurium* that are conditionally essential in at least one of the above selective conditions. Phenotypic study of deletion mutants (*pyrD*, *glnL*, *recD* and STM14_5307) confirmed the phenotypes predicted by Tn-seq data, validating the utility of this approach in discovering new gene functions. The functional relevance of the genes identified was also studied. In chapter 3, Tn-seq method was applied from food safety perspective to a genome-saturating *S. Typhimurium* mutant library recovered from selection under 2 different *in vitro* growth conditions (4°C and -20°C), mimicking storage temperature of chicken meat. We identified an overlapping 42 genes conditionally essential in the selective conditions. The overall study demonstrated the utility and efficiency of the Tn-seq method in

comprehensive identification of conditionally essential genes in *Salmonella*. The genes identified here could be an important resource for better understanding or control of *Salmonella*, including development of novel antimicrobials and vaccines. With continuously increasing sequencing capacity of next generation sequencing technologies, this robust Tn-seq method will aid in revealing unexplored genetic determinants and the underlying mechanisms of various biological processes in *Salmonella* and the other approximately 70 bacterial species for which EZ:Tn5 mutagenesis has been established.

This dissertation is approved for
Recommendation to the
Graduate Council

Dissertation Director:

Dr. Young Min Kwon, PhD

Dissertation Committee:

Dr. Billy M. Hargis, DVM, PhD

Dr. Narayan C. Rath, PhD

Dr. Mack Ivey, PhD

DISSERTATION DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this dissertation when needed for research and/or scholarship.

Agreed _____
Anita Khatiwara

Refused _____
Anita Khatiwara

ACKNOWLEDGEMENTS

I am thankful to many persons for their contributions towards the completion of this dissertation.

First and foremost, I would like to thank my advisor Dr. Young Min Kwon for his advice, encouragement, guidance and invaluable contribution for the completion of this dissertation and my PhD degree. I would also like to thank my committee members, Dr. Billy M. Hargis, Dr. Mack Ivey and Dr. Narayan C. Rath for their advice, constructive criticism and support throughout my PhD program. I also thank them for their time and agreeing to serve on my doctoral dissertation committee.

I would like to offer my deepest appreciation to my family whose unwavering belief in me, unconditional love and support has been the driving force for the completion of my PhD program. My special appreciation goes to my loving parents' father Mr. Jai Narayan Sharma Khatiwara and mother Mrs. Puspa Sharma Khatiwara who taught me the value of education, perseverance, dedication and humbleness. I would like to thank my husband Dr. Krishna Raj Hamal, who has been my best friend and partner since we met a decade ago and my daughter Miss Shreya Hamal, my sunshine. They are my pillars of strength and have helped me become a better person. I would also like to thank my siblings and my husband's family for their love and support.

My humble appreciation goes to my lab members: Nicole Calhoun, Jiang Tieshan, Pallavi Singh, Jeong Nam Kim, Turki Dawoud and Mandy Cox for being wonderful lab mates. I thank Mrs. Donna Deolozzer, for her advice, life lessons and her culinary expertise. I will always treasure our time together.

Finally, I would like to thank my friends in Fayetteville, Bentonville, Rogers and Fort Smith: Silu, Srijana, Mahalaxmi, Rajeev, Samrat, Samjhana, Sujala, Satya, Samata, Pankaj, Basana, Bhooshan, Manita, Samjhana, Krishna, Prabhudha, Pratap, Sandeep and many more. I will always cherish our friendships and times together.

Thank you so much, I love you all.

DEDICATION

This dissertation is dedicated to my loving maternal grandmother late Mrs. Bhagiratha Sharma and my parents father Mr. Jai Narayan Sharma Khatiwara, mother Mrs. Puspa Sharma Khatiwara. These persons have been the greatest influence in my life and inspire me every day.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
CHAPTER I. LITERATURE REVIEW	4
DNA Sequencing.....	5
Next-generation Sequencing.....	6
454 Pyrosequencing	6
Illumina Sequencing.....	7
Sequencing by Oligo Ligation and Detection (SOLiD).....	8
Third generation sequencing.....	9
<i>Salmonella</i>	~ 1
<i>Salmonella</i> genome sequencing.....	13
Functional genomics of <i>Salmonella</i>	13
Genetic tools for <i>Salmonella</i> mutant construction	14
Transposon for random mutagenesis	14
Targeted mutagenesis: red recombinase system.....	16
Transposon-based functional genomics approaches.....	17
Signature Tagged Mutagenesis (STM).....	17
Microarray approaches.....	18
Transposon Site Hybridization (TraSH).....	19
Microarray Tracking of Transposon Mutants (MATT).....	20
Transposon-Mediated Differential Hybridization (TMDH).....	21
Next-generation sequencing (NGS)-based approaches.....	21
Insertion Sequencing (INSeq).....	22
Transposon Sequencing (Tn-seq).....	23
High-throughput Insertion Tracking by Deep Sequencing (HITS).....	24
Transposon-Directed Insertion-site Sequencing (TraDIS).....	24
Tn-seq Circle.....	25
An additional NGS-based strategy.....	26
References.....	28

CHAPTER II. GENOME SCANNING FOR CONDITIONALLY ESSENTIAL GENES IN <i>SALMONELLA</i>	42
Abstract	44
Introduction.....	45
Materials and Methods.....	47
Results.....	52
Discussion.....	62
References.....	67
References for supplementary material.....	71
CHAPTER III. IDENTIFICATION OF <i>SALMONELLA</i> TYPHIMURIUM GENES CONDITIONALLY ESSENTIAL DURING COLD TEMPERATURE STORAGE USING Tn-seq METHOD	95
Abstract	96
Introduction	97
Materials and Methods.....	101
Results.....	106
Discussion.....	111
References.....	114
CHAPTER IV. CONCLUSION	132
Conclusion.....	133
APPENDICES	136

LIST OF FIGURES

	Page
CHAPTER I	
Figure 1. 454 pyrosequencing workflow	39
Figure 2. Illumina sequencing workflow.....	40
Figure 3. SOLiD sequencing work flow.....	41
CHAPTER II	
Figure 1. Schematic diagrams for Tn-seq method used in this study.	79
Figure 2. Identification of the genes conditionally essential for fitness under 3 different selective conditions: dLB, LB-bile, and LB-42°C.	81
Figure 3. Conditionally essential genes. (A)The genes identified as conditionally essential for fitness under each of the three selective conditions ...	82
Figure 4. Competition between the wild type and each of $\Delta pyrD$, $\Delta glnL$, $\Delta recD$, and $\Delta STM14_5307$ mutants	84
Figure 5. Comparison of the fitness indices obtained by Tn-seq data (in Table S3) and a competition assay (CA) for $\Delta pyrD$, $\Delta glnL$, $\Delta recD$, and $\Delta STM14_5307$ mutants	88
Figure 6. <i>In vivo</i> functions of the gene products during animal infection	89
Figure S1. Structure of the 129 bp PCR amplicon containing Tn-5 junction sequence (12bp), ME sequence (19bp) and barcode (6bp).....	90

Figure S2.	The percentage of transposon-junction sequences of different lengths uniquely mapping to the genome	91
Figure S3.	Percentage of the sequence reads for Tn5-junction sequences of different lengths	92
Figure S4.	Reproducibility of biological replicates.....	93
Figure S5.	Growth curves of the wild type, $\Delta pyrD$ and $\Delta glnL$ strains in LB or 10X diluted LB medium (dLB)	94

CHAPTER III

Figure 1	Gel picture. (A) Genomic DNA from input pool, 4° C output pool and -20°C output pool in duplicates along with DNA marker in lane1.(B) PAGE gel picture showing PCR amplicon of 129bp size	123
Figure 2.	PAGE gel picture showing pooled sample of 129bp that was purified and sent for Illumina sequencing	124
Figure 3.	Reproducibility of biological replicates	125
Figure 4.	Genome-wide view of the fitness index	126
Figure 5.	Identification of the genes conditionally essential for fitness under 2 different selective conditions: 4°C and - 20°C	127
Figure 6.	Functional trends associated with two selective conditions	128

Figure 7.	Individual Survival Assay for $\Delta trkA$ and wild type	129
Figure 8.	Individual Survival Assay for $\Delta sapD$ and wild type	130
Figure 9.	Comparison of the fitness indices obtained by Tn-seq data and competition assay (CA) for $\Delta trkA$, $\Delta sapD$, $\Delta rfbF$, and $\Delta rfbI$...	131

LIST OF TABLES

	Page	
CHAPTER I		
Table 1.	Summary of functional genomics techniques based on Microarray and Transposon Mutagenesis	37
Table 2.	Summary of various functional genomics techniques base on Transposon mutagenesis and next-generation sequencing (Illumina)	38
CHAPTER II		
Table S1.	Oligonucleotides used in this study	75
Table S2.	Summary of Illumina sequencing data analysis.....	76
Table S3.	List of all conditionally essential genes identified in this study (fitness index ≤ 0.2 under at least one of the conditions)	CD
Table S4.	The genes conditionally essential for fitness in dLB medium.....	CD
Table S5.	The genes conditionally essential for fitness in LB-bile medium...	CD
Table S6.	The genes conditionally essential for fitness in LB-42°C medium...	CD

CHAPTER III

Table 1.	Oligonucleotides used in this study.....	119
Table 2.	Summary of Illumina sequencing data analysis	120
Table 3.	List of all conditionally essential genes identified in this study (fitness index ≤ 0.5 under at least one of the conditions).....	121
Table 4.	List of genes with transposon insertions identified in this study, previously recognized as cold stress protein encoding genes.....	122

INTRODUCTION

The term “genome” was first used in 1920 and refers to a complete set of genes and chromosomes of an organism. In 1986 Thomas Roderick coined the term “genomics” to name a new journal “*Genomics*” and to describe the science of sequencing, mapping, and genome analysis (McKusick, 1997). Since then the term has been universally accepted and this scientific discipline has undergone expansion with the rapid development of the sequencing technologies. Now genomics may be classified into two major field namely structural and functional genomics (Hieter and Boguski, 1997). Structural genomics deals with initial stages of genome analysis where high resolution genetic, physical and transcript maps are constructed to obtain the ultimate physical map of an organism, the complete DNA sequence. Functional genomics deals with the application of experimental approaches to assess the function of the genes utilizing information provided by structural genomics. It involves high through-put experimental approaches combined with statistical and computational analysis of the results. With the rapid development of next generation sequencing technologies (NGS), complete genome sequence information is available for increasing number of organisms. To keep up with these technologies and ever increasing sequence information, there is a need for high-throughput methods that link genes to phenotypes. This would lead to discovery of new gene functions and help us gain new insights into the behavior of biological systems.

Food borne pathogens are bound to encounter variety of stress conditions such as fluctuating temperature, osmotic pressures, and different food production environment, nevertheless they are able to survive and cause illnesses in the infected host. Bacterial pathogens have an ability to cope with the adverse conditions in the environment and survive to reach host microenvironments that are usually not accessible to nonpathogenic bacteria. The survival

strategy depends upon the expression of bacterial factors under different conditions in timely manner to evade various host immune mechanisms *in vivo* (Slauch *et al.*, 1997). The identification and characterization of the genetic factors responsible for the survival strategies has been an area of major interest to the researchers (Kwon *et al.*, 2009). The identification and characterization of genetic elements responsible for expression of bacterial factors has led to development of effective control measures like antibiotics and vaccines. Antibiotics have proved to be highly effective in treating and controlling bacterial infections; however, antibiotic resistance has steadily increased in the last three decades. Bacteria have a potential to develop resistance to almost every antibiotic irrespective of its chemical class or molecular target (Miesel *et al.*, 2003). The ever growing concern about antibiotic resistance has coincided with revolutionary progress in the availability of genome sequences and high throughput methods to study bacteria (Raskin *et al.*, 2006). Development of the new technologies and strategies combined with genome sequences and bioinformatics tools has revolutionized the way pathogens are studied today. Significant progress has been made in the development of various experimental approaches to identify bacterial essential genes or virulence genes for *in vivo* survival on a genome-wide scale which could be potential drug targets (Sasseti *et al.*, 2001; Hughes, 2003). These drug targets would lead to development of a narrow therapeutic spectrum antimicrobials directed specifically to the biochemical target and decrease the likelihood of development of broad antibiotic resistance.

This doctoral research project is focused on development of high through-put transposon-sequencing (Tn-seq) method and its application along with next generation sequencing (Illumina) to conduct genome-wide identification of *S. enterica* sub sp. *enterica* serovar Typhimurium (*S. Typhimurium*) genes that are conditionally essential for growth or survival

under the conditions mimicking the conditions *Salmonella* cells encounter inside a host and its surroundings during its lifecycle. The genes identified in this study could be an important resource for better understanding of *Salmonella* genetic mechanism for fitness under specific selective conditions, and also promising targets for development of novel antimicrobials and vaccines.

CHAPTER I
LITERATURE REVIEW

I. DNA Sequencing

Since the identification of DNA as a genetic material by Hersey and Chase in 1952 and discovery of double helix structure by Watson and Crick in 1953, there has been increasing demand for DNA sequence information. This demand has led to the development of innovative methodologies and technologies to obtain DNA sequence information. DNA sequencing entails these methods and technologies used to determine the order of the nucleotide bases namely adenine, thymine, guanine and cytosine in a DNA molecule. The advancement of DNA sequencing technologies has been characterized by introduction of a method, its adoption, improvement, and use for a period of time then its replacement by a new technology. The very earliest sequencing techniques involved laborious technique of cleaving short polynucleotides and subsequent identification of the bases by their migration characteristics. The DNA was sequenced by a chemical procedure that broke a terminally labeled DNA molecule partially at each repetition of a base. The length of the labeled fragment was used to identify the positions of that base. The products obtained were resolved by size, by electrophoresis on a polyacrylamide gel and the DNA sequence was read from the pattern of radioactive bands (Maxam and Gilbert, 1977). Maxam-Gilbert sequencing was not adopted further due to its technical complexity prohibiting its use in standard molecular biology kits, extensive use of hazardous chemicals, and difficulties with scale-up (Pesole and Saccone, 2003). This was followed by a significant milestone in mid 1970s by Sanger group when they introduced their idea of using primed template replication by polymerase and separation of the products by gel electrophoresis to obtain DNA sequence information (Sanger and Coulson, 1975). Sanger's group further modified this approach by introducing the base specific chain termination using dideoxynucleotides and this served as the foundation of sequencing for more than three decades (Sanger *et al.*, 1977).

During these 30 years, significant improvements were made that included the use of fluorescent labeled terminators instead of radiolabeled, separation of the products on acrylamide matrices in capillaries instead of gel slabs, introduction of mechanized template preparation and devices for automated production and reading of sequence ladders (Holt and Jones, 2008). This mechanization and industrialization of the sequencing approaches has led to the dawn of modern era of genomics with introduction of next generation and upcoming third generation technologies. DNA sequencing has revolutionized the biological research; today it has become the indispensable part of biotechnology, diagnostic biology, molecular genetics and forensic biology.

2. Next-generation Sequencing (NGS)

The basic characteristic of the past sequencing platforms was the electrophoretic separation of terminated DNA chains whereas present NGS platforms are characterized by flow cell sequencing. Flow cell sequencing involves stepwise determination of DNA sequences by repetitive cycles of nucleotide extension conducted in parallel on massive number of clonally amplified template molecules (Holt and Jones, 2008). In other words, NGS technologies allow massively parallel sequencing of a large number of template fragments. There are different platforms of NGS technologies based on the same flow cell sequencing, yet utilizing different sequencing chemistries. Various NGS platforms and their workflow are described briefly as follows

2.1. 454 Pyrosequencing

Roche 454 Life Sciences was the first company to commercialize a next generation sequencing platform in 2005 (Margulies *et al.*, 2005). The 454 flow cell supports a “picotiter”

plate, with wells of 55 μm depth and 44 μm widths on 75 mm \times 75 mm fiber optic slide (Rothberg and Leamon, 2008). Each individual molecule of sheared template DNA is captured on a separate bead, and each bead is compartmentalized in a private droplet of aqueous PCR reaction mixture within an oil emulsion. Figure 1 shows the 454 flow chart; figure has been reproduced from *Trends Genet* (Mardis, 2008a) with permission. Template is clonally amplified on the bead surface by emulsion PCR, and the template-loaded beads are then distributed into the wells of the picotiter plate. Sequence is obtained by pyrosequencing, the wells are loaded with bead-tethered sequencing enzymes (polymerase, sulfurylase, and luciferase), and buffer containing one of four dNTPs is flowed across the plate wells. If there is a match to the primed template, polymerase incorporates the nucleotide and releases a pyrophosphate molecule which, when converted to ATP by sulfurylase, generates a luciferase-catalyzed chemiluminescence signal that is imaged and recorded (Holt and Jones 2008). The residual nucleotides are washed out and the cycle is repeated with the next dNTP. The current Genome Sequencer (GS) FLX Titanium chemistry platform features long reads of 300-500bp, exceptional accuracy and high throughput (<http://uagc.arl.arizona.edu/index.php/next-gen-sequencing-services/roche454.html>).

2.2. *Illumina sequencing*

The second next sequencing platform called 1G Analyzer, today known as Illumina Genome Analyzer was developed by Solexa which, was later owned and marketed in 2006 by Illumina Inc. (San Diego, CA). This technology is the first of the massive parallel short-read platforms and today is the most successful and widely-adopted next-generation sequencing platform (www.illumina.com). The key features of this system are the in-situ template amplification and use of four-color Sanger-like but reversible terminators. The Illumina flow cell is a planar optically transparent surface comparable to a microscope slide containing a lawn of

oligonucleotide anchors attached to its surface (Holt and Jones, 2008). Figure 2 shows workflow of Illumina genome analyzer and this figure was reproduced from *Trends Genet* (Mardis, 2008a) with permission. Adapters complementary to oligos on the flow cell surface are ligated to the ends of size-selected DNA and these adapted single-stranded DNAs are bound to the flow cell and amplified by solid-phase bridge PCR. In each bridge PCR cycle, priming occurs by arching of the template molecule such that the adapter at its free end hybridizes to and is primed by a free oligo on the flow cell surface resulting in a raindrop pattern of clonally amplified templates. Sequencing proceeds by synthesis using reversible four-color fluorescence where a mix of the four bases each labeled with a different cleavable fluorophore is used simultaneously to find out a given nucleotide position in the template. Labeled terminators, primer, and polymerase are applied to the flow cell. Each base incorporation step is followed by imaging and recording of the fluorescent signal at each cluster, the sequencing reagents are washed away, labels are cleaved, and the 3' end of the incorporated base is unblocked in preparation for the next nucleotide addition. The end result is highly accurate base-by-base sequencing data for broad range of applications.

2.3. Sequencing by Oligo Ligation and Detection (SOLiD)

The third next generation platform called SOLiD system was developed by Applied Biosystems in 2008. It combines elements of various approaches to sequence clonally amplified DNA fragments linked to beads. The amplification and the attachment to the beads are similar to Roche and Illumina but it relies on unique sequence by ligation approach using dye-labeled oligonucleotides. The method provides two base redundancies in sequence reads that enables extra quality check of read accuracy. Figure 3 shows the SOLiD work flow; figure has been reproduced from *Trends Genet* (Mardis, 2008a) and *Annu Rev Genomics Hum Genet* (Mardis,

2008b) with permission. Briefly, oligo adaptor-linked DNA fragments are coupled with 1-mm magnetic beads displaying complementary oligos and each bead–DNA complex is amplified by emulsion PCR. The beads are then covalently attached to the surface of a glass slide that is placed into a fluidics cassette within sequencer. The annealing of a universal sequencing primer that is complementary to the SOLiD specific adapters on the library fragments initiates the ligation based sequencing process. The addition of a limited set of semi-degenerate 8mer oligonucleotides and DNA ligase is automated by the instrument. When a matching 8mer hybridizes to the DNA fragment sequence adjacent to the universal primer 3' end, DNA ligase seals the phosphate backbone. After the ligation step, a fluorescent readout identifies the fixed base of the 8mer, which corresponds to either the fifth position or the second position, depending on the cycle number. A subsequent chemical cleavage step removes the sixth through eighth base of the ligated 8mer by attacking the linkage between bases 5 and 6, thereby removing the fluorescent group and enabling a subsequent round of ligation. The process occurs in steps that identify the sequence of each fragment at five nucleotide intervals, and the synthesized fragments that end at base 25 (or 35 if more cycles are performed) are removed by denaturation and washed away. A second round of sequencing initiates with the hybridization of an n-1 positioned universal primer, and subsequent rounds of ligation-mediated sequencing, and so on. The SOLiD system applications include mutation discovery, metagenomic characterization, non-coding RNA and DNA–protein interaction discovery.

3. Third generation sequencing

The third generation sequencing platforms based on single molecule sequencing have already been envisioned and all set to emerge in the market soon. The advantage of single molecule sequencing is the ability to sequence without amplifying the template which permits

the accurate quantification of specific RNA or DNA molecules rapidly from very small starting template amount (Podolak, 2010). There are three major companies namely Helicos Biosciences (Cambridge, MA), Pacific Biosciences (Menlo Park, CA), and Complete Genomics (Mountain View, CA) racing not only to create novel technology platform but also to provide whole genome sequencing services at \$1000 per genome (Podolak, 2010).

With the aid of these sequencing technologies, it is now possible to acquire gigabases of sequence information in just few days. These technologies have applications in genomics, metagenomics, transcriptomics, epigenomics, mutation mapping, chromatin immunoprecipitation, and discovery of noncoding RNAs (Mardis, 2008a; MacLean *et al.*, 2009). The sequence based characterization of genome has revolutionized the biological research, the fundamental knowledge has been enhanced by the ability to gather genome wide sequence information and it has helped better understand the functional genome.

4. Salmonella

Salmonella is a gram negative, facultative anaerobe, flagellated rod shaped bacterium. Salmonellae are non-fastidious as they can grow and survive under various environmental conditions outside a host. Salmonella is responsible for causing 16 million annual cases of typhoid, 1.3 billion cases of gastroenteritis and 3 million deaths worldwide (Bhunia, 2008). It is one of the most common food borne bacterial pathogen affecting about 1.4 million people annually in the U.S. causing 20000 hospitalizations and 400 deaths resulting in total cost of estimated 3 billion dollars annually in the U.S (Economic Research service 2011; Scallan *et al.*, 2011).

Epidemiologically, *Salmonella* classification is based on host preference (Pui *et al.*, 2011). There are 3 groups, first includes host restricted serovars that infect only humans e.g., *S. Typhi*. The second group includes host adapted serovars that are associated with one host but can cause infection in other host. Host adapted serovars Dublin and Choleraesuis, for example, are generally associated with severe systemic disease in cattle and pigs, respectively, but may also infrequently cause disease in other mammalian hosts including humans. Gallinarum, Abortusovis, and Typhisuis are, respectively, avian, ovine, and porcine *Salmonella* serovars, these host-adapted serovars cannot grow on minimal medium without growth factors, contrary to the ubiquitous *Salmonella* serovars. The third group comprises of remaining serovars that are ubiquitous including *S. Enteritidis*, *S. Typhimurium* and *S. Heidelberg*, these are the common serotypes recovered from humans (Boyen *et al.*, 2008). The important characteristic of these serovars is that they can be harbored sub-clinically in livestock, poultry and persists in environment for long period of time and are thus difficult to control in absence of a detailed knowledge of the genetic requirement of these organism in that particular niche (Andrews-Polymeris *et al.*, 2009).

Kauffmann-White Scheme of classification was adopted by International Association of Microbiologist in 1934 where *Salmonellae* are classified according to three major antigenic determinants (Pui *et al.*, 2011). These determinants are composed of flagellar H antigens, somatic O antigens, and virulence capsular K antigens. World Health Organization Collaborating Center (WHOCC) and Centers for Disease Control and Prevention (CDC) use widely accepted classification system based on phylogenetic tree derived from comparison of 16S rRNA gene sequences. According to this system there are 2,463 *Salmonella* serotypes which are placed under two species based on the differences in 16S rRNA gene sequence analysis. *S. enterica* has

2,443 serotypes while *S. bongori* has 20 serotypes. *Salmonella enterica* is further divided into six subspecies designated by roman numerals namely: *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI) (Janda and Abbott, 2006). Historically, subspecies (V) was *bongori*, which is now recognized as a species. These days the subspecies information is often omitted to bring uniformity in reporting (Bhunia, 2008; Pui *et al.*, 2011).

Salmonella are widely distributed in nature as they survive well in a variety of environments. Poultry, eggs and dairy products are the most common mode of transportation of salmonellosis. Fresh produce like fruits and vegetables have gained concern as mode of transmission in recent years. *Salmonella* can enter into food chain at any stages from livestock feed, through food manufacturing, processing, retailing, catering and food preparation at home (Lo Fo Wong *et al.*, 2002; Bouchrif *et al.*, 2009). Clinical manifestation of the disease in human is characterized by enteric fever or typhoid caused by *S. Typhi* and *Paratyphi*, while non-typhoidal salmonellosis is caused by *S. enteric* especially *S. Typhimurium* and *S. Enteritidis*.

Salmonella are able to survive under various niches ranging from diverse animal hosts, fruits, vegetables to various livestock and food production environments; this ability has made the control of *Salmonella* transmission difficult (Reynolds *et al.*, 2011). Genetic factors required for survival in these niches have been studied traditionally with gene expression and forward genetic studies. The availability of complete genome sequences, microarray technology and cost effective sequencing technologies has enabled high-throughput functional genomics-based tools to understand the genetic requirements for *Salmonella* survival under various conditions. The development and application of such novel functional genomics tools may reveal previously unknown vulnerabilities that can be explored to develop novel intervention and break *Salmonella* transmission chain (Andrews-Polymenis *et al.*, 2009).

5. *Salmonella* genome sequencing

Sequencing and annotation are the initial steps in understanding *Salmonella* genome. The first complete annotated *Salmonella* genome sequence was published in 2001 for *S. Typhi* CT18 (Parkhill *et al.*, 2001) and *S. Typhimurium* LT2 (McClelland *et al.*, 2001). Since then the rapid development and cost effectiveness of the sequencing technologies has made sequencing possible for more than 100 *Salmonella* genomes. These include 23 genomes representing 15 different subspecies I serovars, one subspecies IIIa and one *S. bongori* isolate (Reynolds *et al.*, 2011). The next generation technologies such as 454 pyrosequencing and Illumina (Holt *et al.*, 2008) are now being applied for extensive sequence comparison of non-typhoidal Salmonellae to understand the genetic diversity within and between serovars. Complete genome sequences have led to the development of open reading frame microarrays (Porwollik *et al.*, 2002) and complete tiling arrays (Chan *et al.*, 2003). Several classification methods, such as multilocus enzyme electrophoresis (MLEE) (Boyd *et al.*, 1996) and multilocus sequence typing (MLST) (Arrach *et al.*, 2008) has been developed and are used at the serovar and genovar level. Genome-wide single nucleotide polymorphism analysis has been developed and used for genomic typing (Octavia and Lan, 2007). Therefore, complete genome sequencing of Salmonellae has helped understand the organism better and develop novel tools to understand the complex biology of these important organisms of public health importance.

6. Functional genomics of *Salmonella*

Complete genome sequences provide us the information of an organism's entire gene complement and an opportunity to explore the function of each gene. There are two basic approaches to functional genomics of *Salmonella*, one is expression-based analyses and the other

one is function-based analyses. Expression based analysis determines genes that are expressed in a particular condition or environment and genes identified have potentials to be important for growth and survival. Microarray Technology, In vivo expression technology (IVET), recombinase-based IVET (RIVET) system has been used for gene expression studies under particular conditions relevant to food safety (Frye *et al.*, 2006; Ledebouer *et al.*, 2006; Huang *et al.*, 2007) but has not been directly used to study *Salmonella* genes expressed on the food surface or inside the foods. Expression analyses can identify genes expressed under certain condition but cannot define genes that are required for survival in that particular niche. Therefore, a more direct way of finding the required genes is the use of function-based approaches. The large scale genomic studies provide us with a broad view of the bacterial genome and also create a platform to conduct focused studies by using classical technique like forward genetic screening (Reynolds *et al.*, 2011). The forward genetic screening examines the function of a gene by disrupting the gene and studying the change in the phenotype. This dissertation focuses mainly on the function-based approach using transposon mutagenesis in combination with NGS to conduct genome-wide identification of genes that are conditionally essential under certain growth conditions mimicking environments *Salmonella* cells encounter inside a host and its surroundings.

7. Genetic tools for *Salmonella* mutant construction

7.1. Transposon for random mutagenesis

Transposons (Tn) are discrete transposable DNA segment that can relocate from one genomic site to another. These DNA elements are ubiquitous and have been reported in bacteria, archaea, and Eukarya, including humans. Transposons in bacteria range from simple insertion sequence (IS) elements that consist of a gene(s) for transposition bounded by inverted repeat

sequences, to composite transposons composed of a pair of IS elements that bracket additional genetic information for antibiotic resistance or other properties, to more complex conjugative transposons that exhibit hybrid properties of transposons, plasmids and bacteriophages (Hayes, 2003). Transposon mutagenesis is a biological process that allows a transposon to transfer to a host chromosome leading to a mutation. *In vitro* transposon mutagenesis was first developed to facilitate detailed biochemical studies of transposition mechanisms (Clubb *et al.*, 1996). The findings revealed not only details of transposition but also made researchers realize its potential as tool for genomic studies (Hayes, 2003). Since then *in vitro* transposition reaction has been used to generate genome wide insertions in numerous bacteria and yeast (Goryshin *et al.*, 2000; Hendrixson *et al.*, 2001; Antão *et al.*, 2009; Oh and Nislow, 2011; Yeung *et al.*, 2011).

Transposons used for microbial genomics and proteomics are based on well characterized elements from Gram-negative bacteria. The most favored for genetic tools are those that can insert randomly or can be manipulated to insert randomly. Transposons' Tn3, Tn5, Tn7, Tn10, Tn5 52, and IS 911 are well characterized and extensively used as genetic tools. Tn5, a composite transposon was further modified and used in our study. Tn5 transpose by cut and paste mechanism in which the element is excised from its location and inserted at a new location. Various experimental approaches have been under taken to identify essential and nonessential genes in an organism. Most of the transposon-based high-throughput experimental approaches involve generation of a mutant library either by a random or systematic method (input pool) followed by selection of the mutant library under different conditions (output pool) and then comparison of the original library with that after selection to identify the mutants that are underrepresented or lost from the pool (Kwon *et al.*, 2009). The mutants that are not recovered from output pool are assumed to harbor insertion in genes that are essential for survival under a

specific condition; such genes are also termed as conditionally essential genes. All of the transposon-based approaches make use of different transposon insertion sites or the transposon tagged with a unique marker as a means to identify unique insertions.

7.2. Targeted mutagenesis: red recombinase system

The precise deletion of a gene is an important step towards understanding the function of that particular gene. A targeted systematic inactivation of *Bacillus subtilis* (Kobayashi et al., 2003) and *Saccharomyces cerevisiae* (Winzeler et al., 1999) genes were carried out to estimate minimal set of essential genes and for functional characterization of the genes, respectively. Although both studies provide insights into the respective bacterial genomes, the method is laborious and time consuming and is practical only for small genome sized organisms. There are varieties of techniques to introduce site -specific chromosomal mutations that could be used to delete a specific gene. The technique we adopted to study the function of the genes of interest in this dissertation is the Red Recombinase system developed by Datsenko and Wanner (Datsenko and Warner, 2000). Briefly the target gene was replaced with selectable antibiotic resistance gene sequence generated by PCR using primers with approximately 20bp homology extensions. This gene replacement was accomplished by Red-mediated recombination in the flanking homologies. The Red system includes products Gam, Bet and Exo produced by 3 genes: γ , β and *exo* respectively. Gam inhibits the host RecBCD exonuclease V that would otherwise compete with the Red function (Murphy, 2007). Gene *exo* and β initiate recombination event, *exo* protein is 5'-3' dsDNA exonuclease that binds to the dsDNA ends and degrades the 5' strand leaving 3' ssDNA tails. The Beta protein binds to ssDNA produced by *exo* and promotes annealing of ssDNA stimulating the DNA recombination process (Murphy, 2011). Furthermore Datsenko and Wanner (2000) also created antibiotic resistant cassettes flanked by FLP recognition target (FRT)

sites which could be used for subsequent removal of antibiotic cassette by FLP recombinase produced from a conditionally replicating plasmid. Red recombinase system, developed in *E. coli* has proven to be very efficient and been successfully used in other bacteria including *Salmonella* (Husseiny and Hansel, 2005), *Klebsiella* (Janes et al., 2001), *Pseudomonas* (Lesic and Rahme, 2008) and *Vibrio* (Yamamoto et al., 2009). We were able to use the Red recombinase system to construct the deletion mutants with ease.

8. Transposon-based functional genomics approaches

The availability of numerous genomes sequences of Salmonellae has initiated the comprehensive understanding of *Salmonella* genome and its biology (Reynolds *et al.*, 2011) Transposon-based approaches are very powerful for identification of essential and virulence gene in context of microbial genomics. This section presents an overview of various transposon-based functional genomics approaches that have been developed to identify *in vitro* and *in vivo* survival factors in various bacteria including *Salmonella*.

8.1. Signature Tagged Mutagenesis (STM)

STM is a negative selection strategy developed to identify virulence factors of *Salmonella enterica* serovar Typhimurium in murine infection model (Hensel *et al.*, 1995). STM employs the insertion of random sequence tags as a means to differentiate the individual mutants within a complex input pool of transposon mutants and uses a comparative hybridization strategy to identify transposon mutants that are missing in output pool (Hensel *et al.*, 1995). The STM variations have been used since its development for comprehensive identification and characterization of genes in many microorganisms (Shea *et al.*, 2000; Mecsas, 2002; Shah *et al.*, 2005; Thune *et al.*, 2007; Oh and Nislow, 2011). STM allows parallel comparison of input and

output pools but have limitation in the size of the pools that can be simultaneously screened due to the sensitivity of the membrane-based blotting system and hybridization kinetics. STM is proven to be an important resource that has aided in understanding the growth, infectivity, and virulence factors of many bacterial pathogens (Mecenas, 2002). STM has been extensively used to identify *Salmonella* genes necessary for colonization of calves, chicken and swine that are the primary sources of contaminated meat and poultry products (Bispham *et al.*, 2001; Morgan *et al.*, 2004; Shah *et al.*, 2005; Carnell *et al.*, 2007). STM combines the advantages of transposon mutagenesis with the ability to screen reasonably large number of mutants using fewer animals and is important when using livestock models that are cumbersome and expensive. The limitations of STM include relatively small size pool and the labor required to perform a comprehensive screening.

8.2. *Microarray approaches*

The availability of whole genome sequence provides a catalogue of all genes of an organism and this information has led to the construction of DNA microarrays. A DNA microarray is a collection of microscopic DNA spots attached to a solid surface. The core principle behind microarrays is hybridization between two DNA strands. A high number of complementary base pairs in a nucleotide sequence mean tighter non-covalent bonding between the two strands. After washing off of non-specific bonding sequences, only strongly paired strands will remain hybridized. So fluorescently labeled target sequences that bind to a probe sequence generate a signal that depends on the strength of the hybridization determined by the number of paired bases and the hybridization conditions. Microarrays use relative quantization in which the intensity of a feature is compared to the intensity of the same feature under a different condition, and the identity of the feature is known by its position. DNA microarray have been

used to measure changes in genome wide gene expression in response to various environmental conditions (Sasseti *et al.*, 2001) and has proven valuable in studying gene expression under various conditions. But expression studies have significant limitations in identifying genes that are constitutively or the genes whose expression is modified by posttranscriptional mechanisms. These limitations can be overcome by using techniques that allow the functional assessment on whether the gene is essential for survival under certain condition. The techniques best suited for this kind of analysis is transposon mutagenesis where complex mutant libraries can be constructed readily and each mutation is marked by an insertion (Hamer *et al.*, 2001). Therefore, researchers have developed approaches where they have combined transposon mutagenesis and microarray hybridization to identify genes important in various selective conditions. There are three main approaches that are described below and summarized in Table 1.

8.2A. Transposon Site Hybridization (TraSH)

Sasseti *et al.* (2001) combined microarray technology with the power and simplicity of transposon insertion mutagenesis to develop a microarray-based method known as transposon site hybridization (TraSH) for identifying conditionally essential genes genome wide. TraSH utilizes transposon harboring outward-facing T7 RNA polymerase promoters allowing the preparation of labeled RNA complementary to the transposon-flanking sequences in a complex pool of mutants. The labeled RNA probes prepared from a pool of mutants selected under different conditions for a microarray were used to determine the differential gene requirements of *Mycobacterium tuberculosis* (Sasseti *et al.*, 2001; Sasseti and Rubin 2003). In 2003, Sasseti and Rubin successfully expanded the use of TraSH to genome-wide screening for virulence genes in *M. tuberculosis* required for in vivo survival (Sasseti and Rubin, 2003). A variant of TRaSH was used by Chan *et al.* (2005) to study 50,000 pooled *S. Typhimurium* transposon

mutants during intraperitoneal infection of BALB/c mice and macrophage cell culture. Since then TraSH and its variants have been developed and tested on different organisms. Another microarray based method was developed by Badarinarayana et al. (2001) for genome-wide analysis of quantitative growth phenotypes using insertional mutagenesis and DNA microarrays. The method was applied to assess the fitness contributions of *Escherichia coli* gene domains under specific growth conditions. Transposon-containing genomic DNA fragments from the selected libraries were compared with the initial unselected transposon insertion library on DNA microarrays to identify insertions that affect fitness. The result from the study validated the approach by identifying previously known conditionally essential genes in *Escherichia coli* that are required for growth in minimal medium.

8.2B. Microarray Tracking of Transposon Mutants (MATT)

A method known as microarray tracking of transposon mutants (MATT) based on similar methodology as TraSH but employing a different experimental protocol to retrieve signals corresponding to numerous insertion sites was used to identify essential genes in *Helicobacter pylori* (Salama *et al.*, 2004). In this method, labeled primers were used to perform semi-random PCR directly on genomic DNA from pools of clones to obtain transposon-flanking sequences for hybridization to microarrays unlike previous method that uses transcription of size-selected restriction digested genomic DNA to which T7 promoter has been ligated. The data obtained from this experiment represented the first comprehensive analysis of the full set of essential genes in *H. pylori* (Salama *et al.*, 2004). This experimental approach has been modified and used to analyze genetic requirements of other bacterial species such as *Francisella tularensis* (Weiss *et al.*, 2007) and *Bacillus anthracis* (Day *et al.*, 2007).

8.2C. Transposon-Mediated Differential Hybridization (TMDH)

A new quantitative method termed transposon-mediated differential hybridization (TMDH) was developed to study approximately 10,000 *S. Typhimurium* random transposon mutants in mice (Chaudhuri *et al.*, 2009). They employed Tn5 and Mu transposon with outward facing T7 and SP6 promoters. The use of high-density tiling microarray allowed sub-genic resolution of individual transposon insertion site by automated algorithm. Also, tiling array allowed probe coverage of entire genome thus removing the reliance on annotated genome features (Chaudhuri *et al.*, 2009). Based on the result of TMDH, 47 genes were targeted for deletion studies out of which three were found to provide protection against intravenous and/or oral challenge (Chaudhuri *et al.*, 2009).

The microarray-based methods are recognized as powerful tools for categorizing gene function on a genomic scale that is applicable to a variety of microorganisms. The well-characterized capability of a microarray in detecting target sequences supports the accuracy of the mapping of transposon insertions. However microarray based methods have their drawbacks such as limited resolution and difficulty in distinguishing positive signal from negative.

8.3 Next-generation sequencing (NGS)-based Approaches

The nucleotide sequencing technologies has facilitated the high-throughput identification of gene (Hall, 2007; MacLean *et al.*, 2009). However, the development of broadly applicable tools for directly testing the role of these genes has not been able to keep up with the recent surge in microbial genome sequencing projects and many fundamental questions connecting genome content to functions still remains unexplored (Goodman *et al.*, 2009). Transposon is a powerful tool that can help us connect gene to its function. The most important property of

transposon is its ability to insert randomly into new DNA sites and this makes it ideal source of portable priming sites to determining nucleotide sequence of unknown regions. Primers designed to bind to these priming sites can be used to generate a set of overlapping sequences and these sequences can be assembled into entire sequence of fragment (Butterfield *et al.*, 2002).

The four NGS based approaches namely INSeq (Goodman *et al.*, 2009), Tn-seq (van Opijnen *et al.*, 2009), HITS (Gawronski *et al.*, 2009) and TraDIS (Langridge *et al.*, 2009) came into existence in the year 2009. A new NGS based approach called “the Tn-seq circle method” was described by Gallagher *et al.* (2011) and was used to examine antibiotic resistant trait in *Pseudomonas aeruginosa*. Similarly Christen *et al.* (2011) used a Tn5 derivative transposon (Tn5Pxyl) in combination with Illumina to identify all essential coding and non-coding chromosomal elements in *Caulobacter crescentus* with high accuracy and resolution. All of the approaches make use of transposon mutagenesis in combination with NGS technology to connect phenotype to the gene. Any digital sequence read of 10bp transposon tag with adjacent genomic sequences is considered with almost certainly as a precise position of a transposon insertion site (Langridge *et al.*, 2009). Therefore, the combination of hyper-saturated transposon mutant pool with high-throughput Illumina sequencing has provided unparalleled degree of resolution to a transposon mutagenesis screening. These transposon and NSG combined approaches are described below and summarized in Table 2.

8.3A Insertion Sequencing (INSeq)

A broadly applicable Mariner transposon was integrated with second-generation sequencing technology and mouse model to study the functional genomics of human gut microbes (Goodman *et al.*, 2009). A negative selection scheme was adopted to identify genes

required for colonization *in vivo*. In this method termed insertion sequencing (INSeq), a single G–T transversion was introduced at a non-conserved position of the inverted repeat (IR) sequences to create a recognition sequence for the MmeI restriction enzyme. MmeI would cleave 16 bp outside of the transposon, capturing a genomic fragment that identifies the insertion site. The short genomic DNA sequences of mutagenized *B. thetaiotamicron* captured by this MmeI digestion were used to uniquely map transposon location on *B. thetaiotamicron* known genome sequence. The resulting transposon mutants were subjected to *in vitro* and *in vivo* selective conditions to highlight mutants that change in relative abundance, and thereby identify genes and pathways critical for fitness under the selective conditions.

8.3B Transposon Sequencing (Tn-seq)

van Opijnen et al. (2009) introduced MmeI site in *magellan6* mini transposon to develop a widely applicable high-throughput tool for gene disruption, called Tn-seq. This method enabled determination of each *Streptococcus pneumoniae* gene's role under a specific condition of interest employing massive parallel sequencing. First, a mutant library was constructed by transposing mini-transposon *magellan6* that contained Mme restriction site within each inverted repeat, and then the bacteria were transformed with the transposed DNA resulting in a pool of bacteria containing transposon insertions. The genomic DNA was isolated from a portion of the pool (t1) and another portion was used to seed a culture for *in vitro* selection. DNA was isolated from the bacteria recovered after selection (t2). The DNA obtained from both time points t1 and t2 were digested with MmeI, then a PCR was performed to amplify 160bp sequence with 20bp of bacterial-specific DNA flanked by Illumina specific sequences to enable sequencing. After sequencing, samples were identified based on barcodes and 20bp reads were mapped to the genome, counted for insertion and fitness for each gene identified were determined. The

quantitative nature of massive parallel sequencing enabled quantitative determination of fitness of a gene knockout as a direct measure of the growth rate of the mutant.

8.3C High-throughput Insertion Tracking by Deep Sequencing (HITS)

HITS is a genome scale negative selection technology that combines high-density transposon mutagenesis and massive parallel sequencing to identify mutants lost from a library after exposure to selective condition(s) (Gawronski *et al.*, 2009). This approach was applied for identification of *Haemophilus influenza* genes responsible for delayed clearance in murine pulmonary model. The first step is the transposon mutagenesis with minitransposon derived from the *HimarI-mariner* transposon followed by shearing of the genomic DNA containing high density transposon insertion. Fragments of DNA were enriched with PCR using transposon-specific and adapter-specific primers and resulting amplicons were purified by affinity capture. Then sequencing was performed *en masse* on the Illumina next-generation sequencing platform. After sequencing reads were mapped to the reference genome to identify the transposon insertion sites. The relative abundance of the mutants within the library before and after the selection was determined by number of insertion sites detected per gene and number of sequencing reads per site. Insertions in genes that are essential for growth or survival in the selective conditions confer attenuated growth or survival during mouse infection.

8.3D Transposon-Directed Insertion-site Sequencing (TraDIS)

TraDIS is another high through-put approach introduced to investigate the essential genes of *S. Typhi* under laboratory and biologically relevant conditions (Langridge *et al.*, 2009). Transposon Tn5 was used to generate *S. Typhi* transposon mutant library containing an estimated 1.1 million individual mutants. The portion of mutants from the input library was passaged

through selective conditions. Genomic DNA was extracted after selection and fragmented to an average size of 300bp. Illumina paired end adapters were ligated to these genomic DNA fragments producing mixture of fragments some of which containing the desired end of transposon and the adjoining bacterial genomic sequence. Transposon insertion sites were amplified with PCR using transposon-specific and Illumina-specific primers. The amplified DNA fragment libraries were sequenced using Illumina and mapped to *S. Typhi* genome. The number and frequency of insertions mapping to each nucleotide in the *S. Typhi* genome were determined for each selective condition and genes were identified. They were able to map 370,000 unique transposon insertion sites to *S. Typhi* genome, an insertion every 13bp on average. With this high density and resolution they were able to simultaneously assay every gene in the genome for essentiality and generated genome-wide list of conditionally essential genes. They determined 356 genes essential for growth under standard laboratory conditions, out of which 256 genes were previously were found to be essential in *E.coli* (Baba *et al.*, 2006). They demonstrated that every gene in the genome could be assayed simultaneously to identify niche-specific essential genes. Eckert *et al.* applied TraDIS retrospectively to assign the genotype and fitness score of enterohemorrhagic *Escherichia coli* O157:H7 (EHEC) mutants previously screened in calves using STM. Using TraDIS they were able to reproduce the fitness defects of the mutants detected by STM with substantial time and cost savings. TraDIS was recognized as a significant advancement towards the principles of reduction, refinement and replacement of animals in research (Eckert *et al.*, 2011).

8.3E Tn-seq Circle

The Tn-seq circle method is a deep-sequencing procedure for tracking large number of *Pseudomonas aeruginosa* transposon mutants (Gallagher *et al.*, 2011). It employs a new Tn-seq

method to generate and amplify single-strand circles carrying transposon junction sequences. Briefly, a transposon mutant library was constructed by Tn5-based transposon (*ISlacZhah-tc*) mutagenesis. The DNA from the mutant pool was sheared, end repaired and all free ends were ligated to one Illumina adaptor. After the ligation products are digested with BamHI for size selection, the resulting single stranded fragments containing transposon ends were circularized by template ligation. This was followed by exonuclease step to eliminate all non-circularized DNA fragments. The transposon-genome junctions were PCR amplified from circularized fragments and second Illumina adaptor was introduced in this step. The PCR products were sequenced on an Illumina flow cell and each sequence read was then mapped to the *P. aeruginosa* genome. Using this method they were able to identify the genes responsible for an intrinsic antibiotic resistance trait in *P. aeruginosa* and validate their Tn-seq method by comparing results with the previous study (Gallagher *et al.*, 2011).

8.3F An additional NSG-based strategy

Christen *et al.* employed Tn5 derived transposon Tn5Pxyl carrying outward facing Pxyl promoter at one end. Depending on the insertion orientation these element is capable of activating or disrupting transcription at any site of integration. Hyper saturated *Tn5Pxyl* insertion library containing an estimated 8×10^5 transposon mutants was constructed. DNA fragments that covered transposon junctions from each mutant pool were simultaneously amplified by a semi-arbitrary 2 step PCR strategy. In a first-round of PCR, a transposon specific primer pointing outwards of the *Tn5Pxyl* element was used in combination with three different semi-arbitrary primers. The PCR products obtained from first PCR were used as templates for second nested PCR step using the Illumina paired-end primer and adapter sequence primer. Second-round PCR products were pooled and DNA fragment of 140 bp-700 bp were size selected and purified by

agarose gel electrophoresis prior to Illumina sequencing. The sequencing reads were mapped to 4Mbp *C. crescentus* genome. Using this strategy they were able to determine *C. crescentus* essential genome at 8bp resolution that included 480 ORF, 402 regulatory sequences, 130 non-coding sequences and 90 intergenic sequences of unknown function. Christen *et al.*, 2011 were able to establish this high resolution strategy for *C. crescentus* and suggest that this strategy is scalable and applicable to high through-put whole genome essential gene studies in broad class of bacterial species.

Thus, all of the methods mentioned above reiterate the fact that the combination of large transposon mutant library and sequencing based strategies identifies precise genomic location, enables digital count of individual insertions and provides an alternative to species-specific DNA microarrays required for hybridization-based mutant profiling. The following chapters of this dissertation focuses on the development of new version of aforementioned Tn-seq method using modified EZ:Tn5 transposon and its application for genome wide identification of conditionally essential genes of *Salmonella*, under various selective condition mimicking the environments the bacteria thrive inside the host (Chapter II). In addition, as *Salmonella* is a major food borne pathogen, the Tn-seq method was also used from food safety prospective to identify conditionally essential genes at food storage temperatures.

REFERENCE

- Andrews-Polymenis, H.L., Santiviago, C.A., and McClelland, M. (2009) Novel genetic tools for studying food-borne *Salmonella*. *Curr Opin Biotechnol* **20(2)**: 149-157.
- Antão, E.M., Ewers, C., Gürlebeck, D., Preisinger, R., Homeier, T., Li, G., and Wieler, L.H. (2009) Signature-tagged mutagenesis in a chicken infection model leads to the identification of a novel avian pathogenic *Escherichia coli* fimbrial adhesin. *PLoS One* **4(11)**:e7796.
- Arrach, N., Porwollik, S., Cheng, P., Cho, A., Long, F., Choi, S.H., and McClelland, M. (2008) *Salmonella* serovar identification using PCR based detection of gene presence and absence. *J Clin Microbiol* **46**: 2581-2589.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura Y., Baba, M., *et al.* (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006-0008.
- Badarinarayana, V., Estep, P.W. 3rd, Shendure, J., Edwards, J., Tavazoie, S., Lam, F., and Church, G.M. (2001) Selection analyses of insertional mutants using subgenomic-resolution arrays. *Nat Biotechnol* **19 (11)**: 1060-1065.
- Bhunia, A.K. (2008) Foodborne microbial pathogens: Mechanisms and pathogenesis. United States of America: Springer Science + Business Media, LLC. Pg. 201-215
- Bispham, J., Tripathi, B.N., Watson, P.R., and Wallis, T.S. (2001) *Salmonella* pathogenicity island 2 influences both systemic salmonellosis and *Salmonella*-induced enteritis in calves. *Infect Immun* **69**: 367-377.
- Bouchrif, B., Paglietti, B., Murgia, M., Piana, A., Cohen, N., Ennaji, M., *et al.* (2009) Prevalence and antibiotic-resistance of *Salmonella* isolated from food in Morocco. *J Infect Dev Ctries* **28(3)**: 35-40.
- Boyd, E.F., Wang, F.S., Whittam, T.S., and Selander, R.K. (1996) Molecular genetic relationship of the *Salmonellae*. *Appl Environ Microbiol* **62**: 804-808
- Boyen F., Haesebrouch, F., Maes, D., Van Immerseel, F., Ducatelle, R.F. and Pasmans, F. (2008) Non typhoidal *Salmonella* infections in pigs: a closer look at epidemiology, pathogenesis and control. *Vet Microbiol* **130**: 1-9.
- Butterfield, Y.S., Marra, M.A., Asano, J.K., Chan, S.Y., Guin, R., Krzywinski, M.I., *et al.* (2002) An efficient strategy for large-scale high-throughput transposon-mediated sequencing of cDNA clones. *Nucleic Acids Res* **30(11)**: 2460-2468.

- Carnell, S.C., Bowen, A.J., Morgan, E., Maskell, D., Wallis, T.S., and Stevens, M.P. (2007) Role in virulence and protective efficacy in pigs of *Salmonella enterica* serovar Typhimurium secreted components identified by signature-tagged mutagenesis. *Microbiology* **153**: 1940-1952.
- Chan, K., Kim, C.C., and Falkow, S. (2005) Microarray-based detection of *Salmonella enterica* serovar Typhimurium transposon mutants that cannot survive in macrophages and mice. *Infect Immun* **73(9)**: 5438-5449.
- Chan, K., Baker, S., Kim, C.C., Detweiler, C.S., Dougan, G., and Falkow, S. (2003) Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. *J Bacteriol* **185**: 553-563.
- Chaudhuri, R.R., Peters, S.E., Pleasance, S.J., Northen, H., Willers, C., Paterson, G.K., *et al.* (2009) Comprehensive identification of *Salmonella enterica* serovar Typhimurium genes required for infection of BALB/c mice. *PLoS Pathog* **5(7)**: e1000529.
- Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Collier, J. A., *et al.* (2011) The essential genome of a bacterium. *Mol Syst Biol* **7**: 528.
- Clubb, R.T., Mizuuchi, M., Huth, J.R., Omichinski, J.G., Savilahti, H., Mizuuchi, K., *et al.* (1996) The wing of the enhancer-binding domain of Mu phage transposase is flexible and is essential for efficient transposition. *Proc Natl Acad Sci USA* **93**:1146–1150
- Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **697(12)**: 6640-6645.
- Day, W.A. Jr., Rasmussen, S.L., Carpenter, B.M., Peterson, S.N., and Friedlander, A.M. (2007) Microarray analysis of transposon insertion mutations in *Bacillus anthracis*: global identification of genes required for sporulation and germination. *J Bacteriol* **189(8)**: 3296-3301.
- Eckert, S.E., Dziva, F., Chaudhuri, R.R., Langridge, G.C., Turner, D.J., Pickard, D.J., *et al.* (2011) Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of *Escherichia coli* O157:H7 screened in cattle. *J Bacteriol* **193**: 1771-1776.
- Economic Research Service. 2011. Foodborne Illness Cost Calculator: Salmonella. United States Department of Agriculture. Available at: <<http://www.ers.usda.gov/Data/FoodborneIllness/>> [Accessed 08 September 2011].
- Frye, J., Karlinsey, J.E., Felise, H.R., Marzolf, B., Dowidar, N., McClelland, M., and Hughes, K.T. (2006) Identification of new flagellar genes of *Salmonella enterica* serovar Typhimurium. *J Bacteriol* **188**: 2233-2243.

- Gallagher, L.A., Shendure, J., and Manoil, C. (2011) Genome-Scale Identification of Resistance Functions in *Pseudomonas aeruginosa* Using Tn-seq. *MBio* **2** (1): e00315–e00310
- Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V., and Akeley, B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci USA* **106**(38): 16422-16427.
- Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., *et al.* (2009) identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**(3): 279-289.
- Goryshin, I.Y., Jendrisak, J., Hoffman, L.M., Meis, R., and Reznikoff, W.S. (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. *Nat Biotechnol* **18**: 97-100
- Hall, N. (2007) Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* **210**: 1518-1525.
- Hamer, L., DeZwaan, T.M., Montenegro-Chamorro, M.V., Frank, S.A., and Hamer, J.E. (2001) Recent advances in large-scale transposon mutagenesis. *Curr Opin Chem Biol* **5**: 67-73.
- Hayes, F. (2003) Transposon based strategies for microbial functional genomics and proteomics. *Annu Rev Genet* **37**: 3-29.
- Hendrixson, D.R., Akerley, B.J., and DiRita, V.J. (2001) Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility. *Mol Microbiol* **40**: 214-224
- Hensel, M., Shea, J.E., Gleeson, C., Jones, M.D., Dalton, E., and Holden, D.W. (1995) Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269**: 400-403.
- Hieter, P., and Boguski, M. (1997) Functional genomics: it's all how you read it. *Science* **278**(5338): 601-602.
- Holt, K.E., Parkhill, J., Mazzoni, C.J., Roumagnac, P., Weill, F.X., Goodhead, I., *et al.* (2008) High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nat Genet* **40**: 987-993.
- Holt, R.A., and Jones, S. J. (2008) The new paradigm of flow cell sequencing. *Genome Res* **18**(6): 839-846.

- Huang, Y., Leming, C.L., Suyemoto, M., and Altier, C. (2007) Genome-wide screen of *Salmonella* genes expressed during infection in pigs using in vivo expression technology. *Appl Environ Microbiol* **73**:7522-7530.
- Hughes, D. (2003) Exploiting genomics, genetics and chemistry to combat antibiotic resistance. *Nat Rev Genet* **4**: 432-441.
- Husseiny, M.I., Hensel, M. (2005) Rapid method for the construction of *Salmonella enterica* serovar Typhimurium vaccine carrier strains. *Infect Immun* **73**(3): 1598–1605.
- Janda, J.M., and Abbott, S.L. (2006) The Enterobacteria, 2nd ed., ASM Press, Washington, DC. Pp. 7-14.
- Janes, B.K., Pomposiello, P.J., Perez-Matos, A., Najarian, D.J., Goss, T.J., Bender, R.A. (2001) Growth inhibition caused by overexpression of the structural gene for glutamate dehydrogenase (gdhA) from *Klebsiella aerogenes*. *J Bacteriol* **183**(8): 2709–2714.
- Kobayashi K, Ehrlich SD, Albertini A, Amati G, Andersen KK, Arnaud M, Asai K, Ashikaga S, Aymerich S, Bessieres P, Boland F, Brignell SC, Bron S, Bunai K et al., 2003. Essential *Bacillus subtilis* genes. *Proc Natl Acad Sci U S A*. **100**(8):4678-83.
- Kwon, Y.M., Bhattacharya, D., and Ricke, S.C. (2009). Genomic approaches to bacterial pathogens using transposon mutagenesis: food safety applications. *In: Perspectives on Food Safety Issues of Animal Derived Foods*, S.C. Ricke and F. T. Jones (eds.), University of Arkansas Press, Fayetteville, Ark., pp. 165-176.
- Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., et al. (2009) Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Res* **19**(12): 2308-2316.
- Ledeboer, N.A., Frye, J.G., McClelland, M., and Jones, B.D. (2006) *Salmonella enterica* serovar Typhimurium requires the Lpf, Pef, and Tafi fimbriae for biofilm formation on HEp-2 tissue culture cells and chicken intestinal epithelium. *Infect Immun* **74**: 3156-3169.
- Lesic, B., Rahme, L.G. (2008) Use of the lambda Red recombinase system to rapidly generate mutants in *Pseudomonas aeruginosa*. *BMC Mol Biol* **9**: 20.
- Lo Fo Wong, D.M.A., Hald, T., van der Wolf, P.J., and Swanenburg, M. (2002) Epidemiology and control measures for *Salmonella* in pigs and pork. *Livest Prod Sci* **76**(3): 215-222.
- MacLean, D., Jones, J.D., and Studholme, D.J. (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**: 287-296.
- Mardis, E.R. (2008a) The impact of next generation sequencing technology on genetics. *Trends Genet* **24**: 133-141

- Mardis, ER. (2008b) Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet.* **9**:387-402.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bembien, L.A., *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437(7057)**: 376-380.
- Maxam, A.M., and Gilbert, W. (1977) A new method for sequencing DNA. *Proc Natl Acad Sci USA* **74(2)**: 560-564.
- McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., *et al.* (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852-856.
- McKusick, V.A. (1997) Genomics: structural and functional studies of genomes. *Genomics* **45(2)**: 244-249.
- Mecenas, J. (2002) Use of signature-tagged mutagenesis in pathogenesis studies. *Curr Opin Microbiol* **5**: 33-37.
- Miesel, L., Greene, J., and Black, T.A. (2003) Genetic strategies for antibacterial drug discovery. *Nat Rev Genet* **4**: 442-456.
- Morgan, E., Campbell, J.D., Rowe, S.C., Bispham, J., Stevens, M.P., Bowen, A.J., *et al.*, (2004) Identification of host-specific colonization factors of *Salmonella enteric* serovar Typhimurium. *Mol Microbiol* **54**: 994-1010.
- Murphy, K.C. (2011) Targeted chromosomal gene knockout using PCR fragments. *Methods Mol Biol* **765**: 27-42.
- Murphy, K.C. (2007) The lambda Gam protein inhibits RecBCD binding to dsDNA ends. *J Mol Biol* **371(1)**: 19-24.
- Octavia, S., and Lan, R. (2007). Single-nucleotide-polymorphism typing and genetic relationships of *Salmonella enterica* serovar Typhi isolates. *J Clin Microbiol* **45**: 3795-3801.
- Oh, J., and Nislow, C. (2011) Signature-tagged mutagenesis to characterize genes through competitive selection of bar-coded genome libraries. *Methods Mol Biol* **765**: 225-252.
- Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain J., *et al.* (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* **413**:848-852.

- Pesole, G., and Saccone, C. (2003) Handbook of comparative genomics: principles and methodology. New York: Wiley-Liss. pp. 133.
- Podolak, E. (2010) Sequencing's new race. *BioTechniques* **48 (2)**: 105-111.
- Porwollik, S., Wong, R.M., and McClelland, M. (2002) Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc Natl Acad Sci USA*. **99**: 8956-8961.
- Pui, C.F., Wong, C., Chai, L.C., Tunung, R., Jeyaletchumi, P., Noor Hidayah, M.S., *et al.* (2011). *Salmonella*: A foodborne pathogen. *Int Food Res J* **18**: 465-473.
- Raskin, D.M., Seshadri, R., Pukatzki, S.U. and Mekalanos, J.J. (2006) Bacterial genomics and pathogen evolution. *Cell* **124**: 703-714.
- Reynolds, M.M., Canals, R., McClelland, M., and Andrews-Polymenis, H.L. (2011) High-throughput screening to determine the genetic requirements for *Salmonella* survival under different growth conditions. *Salmonella: from genome to function*. Caister Academic Press, Norflock, UK, pp. 69-89.
- Rothberg, J.M., and Leamon, J.H. (2008) The development and impact of 454 sequencing *Nature Biotechnol* **26**: 1117-1124.
- Salama, N.R., Shepherd, B., and Falkow, S. (2004) Global transposon mutagenesis and essential gene analysis of *Helicobacter pylori*. *J Bacteriol* **186(23)**: 7926-7935.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467.
- Sanger, F., and Coulson, A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* **94**: 441-448.
- Sassetti, C.M., and E.J. Rubin. (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci USA* **100**: 12989-12994.
- Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2001) Comprehensive identification of conditionally essential genes in Mycobacteria. *Proc Natl Acad Sci USA* **98(22)**: 12712-12717.
- Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy S.L., *et al.* (2011) Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis* **17(1)**: 7-15.
- Shah, D.H., Lee, M.J., Park, J.H., Lee, J.H., Eo, S.K., Kwon, J.T., and Chae, J.S. (2005) Identification of *Salmonella gallinarum* virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology* **151**: 3957-3968.

- Shea, J.E., Santangelo, J.D., and Feldman, R.G. (2000) Signature-tagged mutagenesis in the identification of virulence genes in pathogens. *Curr Opin Microbiol* **3**: 451-458.
- Slauch, J., Taylor, R., and Maloy, S. (1997) Survival in a cruel world: how *Vibrio cholerae* and *Salmonella* respond to an unwilling host. *Genes Dev* **11**: 1761-1774.
- Thune, R.L., Fernandez, D.H., Benoit, J.L., Kelly-Smith, M., Rogge, M.L., Booth, N.J., *et al.* (2007) Signature-tagged mutagenesis of *Edwardsiella ictaluri* identifies virulence-related genes, including a salmonella pathogenicity island 2 class of type III secretion systems. *Appl Environ Microbiol* **73(24)**: 7934-7946.
- van Opijnen, T., Bodi, K.L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6(10)**: 767-772.
- Weiss, D.S., Brotcke, A., Henry, T., Margolis, J.J., Chan, K., and Monack, D.M. (2007) *In vivo* negative selection screen identifies genes required for *Francisella* virulence. *Proc Natl Acad Sci USA* **104(14)**: 6037-6042.
- Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285(5429)**: 901-906.
- www.illumina.com
- Yamamoto, S., Izumiya, H., Morita, M., Arakawa, E., and Watanabe, H. (2009) Application of lambda Red recombination system to *Vibrio cholerae* genetics: simple methods for inactivation and modification of chromosomal genes. *Gene*. **438(1-2)**:57-64.
- Yeung, A., Cameron, D.W., Desjardins, M., and Lee, B.C. (2011) Use of signature-tagged mutagenesis to identify virulence determinants in *Haemophilus ducreyi* responsible for ulcer formation. *J Microbiol Methods* **84(2)**: 290-298.

Tables

Table 1. Summary of functional genomics techniques based on Microarray and Transposon Mutagenesis

Table 2. Summary of various functional genomics techniques based on Transposon mutagenesis and next-generation sequencing (Illumina).

Figure Legends

Figure 1. 454 Workflow: library construction ligates 454-specific adapters to DNA fragments and couples amplification beads with DNA in an emulsion PCR to amplify fragments before sequencing. The beads are loaded into the picotiter plate (PTP). The bottom panel illustrates the pyrosequencing reaction that occurs on nucleotide incorporation to report sequencing by synthesis. (Mardis, E.R., 2008a. *Trends Genet* **24**: 133-141. Figure reproduced with permission. License no.2839470406186)

Figure 2. Illumina workflow. Starting from similar fragmentation and adapter ligation steps, the library is added to a flow cell for bridge amplification (an isothermal process that amplifies each fragment into a cluster). The cluster fragments are denatured, annealed with a sequencing primer and subjected to sequencing by synthesis

using 3' blocked labeled nucleotides. (Mardis, E.R., 2008a, *Trends Genet* **24**: 133-141. Figure reproduced with permission, License no.2839470406186.)

Figure 3. SOLiD work flow: (a) The ligase-mediated sequencing approach of the Applied Biosystems SOLiD sequencer. DNA fragments for SOLiD sequencing are amplified on the surfaces of 1- μ m magnetic beads to provide sufficient signal during the sequencing reactions, and are then deposited onto a flow cell slide. Ligase-mediated sequencing begins by annealing a primer to the shared adapter sequences on each amplified fragment, and then DNA ligase is provided along with specific fluorescent-labeled 8mers, whose 4th and 5th bases are encoded by the attached fluorescent group. Each ligation step is followed by fluorescence detection, after which a regeneration step removes bases from the ligated 8mer (including the fluorescent group) and concomitantly prepares the extended primer for another round of ligation. (b) Principles of two-base encoding. Because each fluorescent group on a ligated 8mer identifies a two-base combination, the resulting sequence reads can be screened for base-calling errors versus true polymorphisms versus single base deletions by aligning the individual reads to a known high-quality reference sequence. (Mardis, E.R., 2008a. *Trends Genet* **24**: 133-141 Figure reproduced with permission License no.2839470406186.) (Mardis, E.R. 2008b, *Annu. Rev. Genomics Hum. Genet.* **9**:387-402.)

Table 1. Summary of functional genomics techniques based on Microarray and Transposon Mutagenesis

Name	Features	Advantage	Disadvantage
<p>TraSH Sassetti <i>et al.</i>, 2001</p>	<p>Himari1 based Mariner transposon Outward facing T7 promoter promotes in vitro transcription, Identification of conditionally essential genes in <i>Mycobacteria</i></p>	<p>General application</p>	<p>Genes lacking TA dinucleotide cannot be Detected</p>
<p>MATT Salama <i>et al.</i>, 2004</p>	<p>Tn7 Transposon, Essential genes analysis of <i>Helicobacter pylori</i> Maps the genomic locus of transposition insertion for pools of clones, amplifies transposon-flanking sequences by semi-random PCR directly on genomic DNA from pools of clones.</p>	<p>General application</p>	<p>False positive</p>
<p>TMDH Chaudhuri <i>et al.</i>, 2009</p>	<p>Tn5 and Mu transposon, outward facing T7 and SP6 promoters for in vitro transcription, Comprehensive identification of <i>S.Typhimurium</i> genes required for infection in mice. identifies attenuated mutants in vitro and in vivo identified targets for vaccine development</p>	<p>General application High-density tiling microarray Allows high resolution, probe coverage of the entire genome Allows determination of genomic position of the transposon insertions by automated algorithm</p>	<p>Cannot identify insertion within restriction Fragment False positive</p>

Table 2. Summary of various functional genomics techniques based on Transposon mutagenesis and next-generation sequencing (Illumina).

Name	Specific features	Advantages	Disadvantages
In-seq (Goodman et al., 2009) Tn-seq (vanOpjnen et al., 2009)	Mariner transposon, MmeI on both Inverted repeats, 16bp junction sequence Tn-seq: <i>Streptococcus pneumoniae</i> In-seq: <i>Bacteriodes thetaiotamicron</i>	Identical lengths of sequence tags, Robust protocol	Limited to Mariner Transposon
HITS (Gawronski et al., 2009)	mini Himari-1, Fragmentation, size fractionation of genomic DNA, PCR with Biotinylated transposon primer, prolonged survival of <i>Haemophilus influenzae</i> in lungs	General application Recover specific fragments by streptavidine-coated magnetic beads	Shearing of gDNA
TraDIS (Langridge et al., 2009)	Tn5, Transposon specific primer for Illumina sequencing, simultaneous assay of every <i>Salmonella</i> Typhi gene	Uses transposon specific primer for Illumina sequencing	Shearing of gDNA Expensive
Tn-seq Circle (Gallagher et al., 2011)	Tn5 based, circularization of DNA Resistance function in <i>Pseudomonas aeruginosa</i> .	General application Robust protocol	Shearing, Biased circularization due to varying DNA fragment size
Christen et al., 2011	Tn5 derivative transposon (Tn5Pxyl), carries an inducible outward pointing Pxyl promoter at one end, pair end adapter Essential genome of <i>Caulobacter crescentus</i>	Robust protocol, general application	Method to be tested in large size bacterial genome
Tn-seq (this study)	Tn5, BsmFI in one of the ME regions, Barcoding strategy for simultaneous sequencing of multiple samples, <i>Salmonella</i> Typhimurium conditionally essential genes	Simple and powerful Uniform amplicon size	Short 12bp junction sequence

Figure 1.

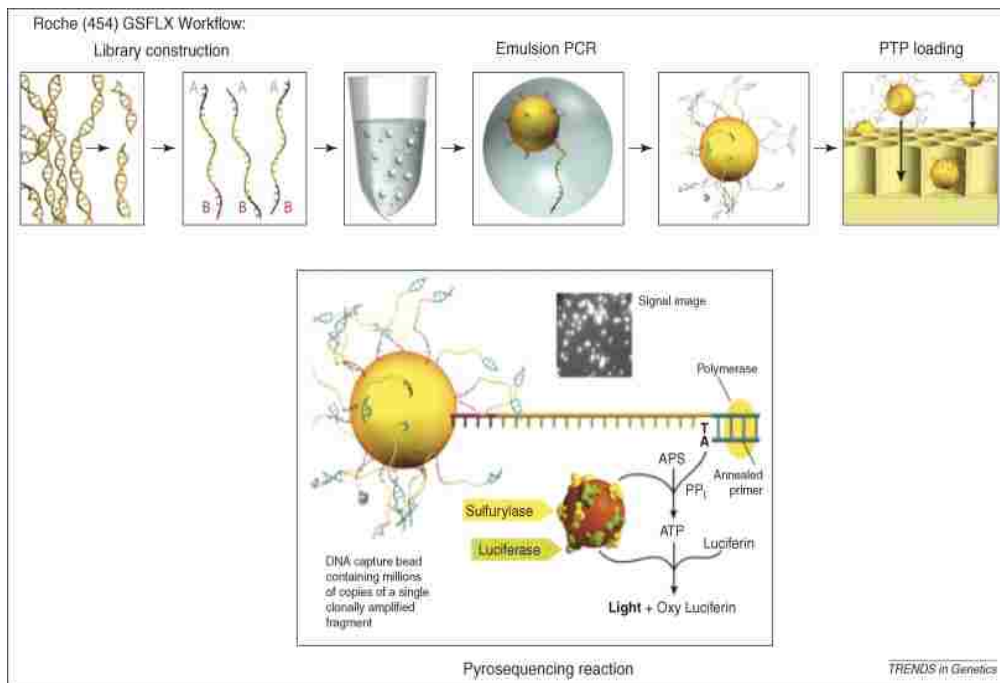


Figure 2.

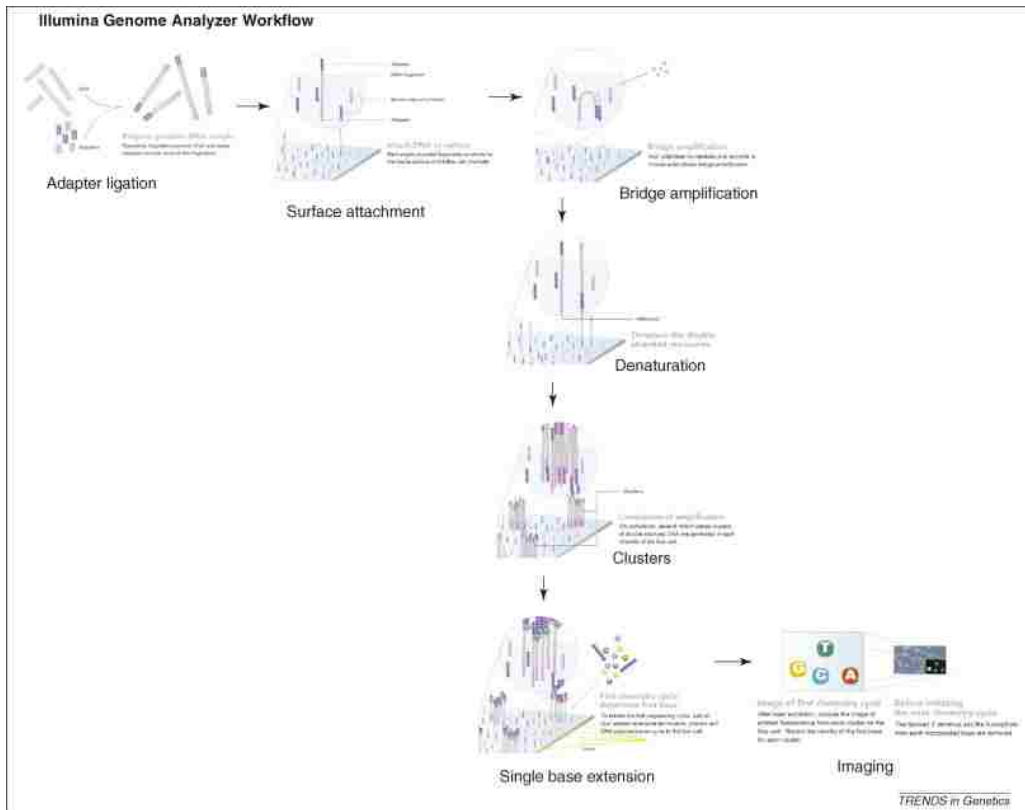
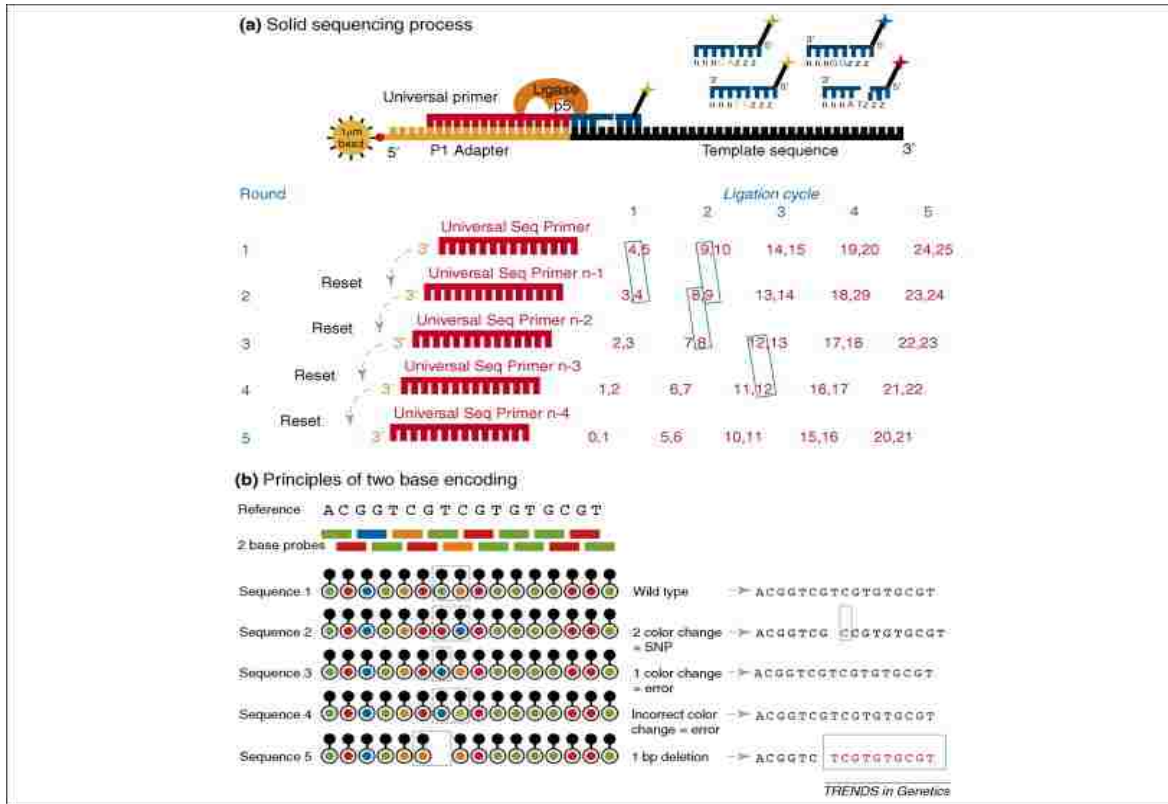


Figure 3.



CHAPTER II

GENOME SCANNING FOR CONDITIONALLY ESSENTIAL GENES IN *SALMONELLA*

Accepted for publication in Applied and Environmental Microbiology Journal

Genome scanning for conditionally essential genes in *Salmonella*

Anita Khatiwara¹, Tieshan Jiang¹, Sam-Sun Sung², Turki Dawoud^{1, 3}, Jeong Nam Kim^{1†}, Dhruva
Bhattacharya^{1‡}, Hee-Bal Kim², Steven C. Ricke^{3, 4, 5}, and Young Min Kwon^{1, 3, 4*}

¹Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR 72701,

²Department of Food and Animal Biotechnology, Seoul National University, Seoul, Korea,

³Center for Food Safety, University of Arkansas, Fayetteville, Arkansas 72704,

⁴Department of Poultry Science, University of Arkansas, Fayetteville, AR72701

⁵Department of Food Science, University of Arkansas, Fayetteville, AR72701

[†]Present address: JHM Health Science Center, Department of Oral Biology, University of
Florida, Gainesville, FL 32610

[‡]Present address: Department of Molecular Genetics and Microbiology, University of New
Mexico Health Sciences Center, Albuquerque, NM 87131

*Corresponding author. Mailing address:

Department of Poultry Science

University of Arkansas

Fayetteville, AR 72701, USA.

Phone : 479-575-4935

Fax : 479-575-8775

ABSTRACT

As more whole genome sequences become available, there is an increasing demand for high-throughput methods that link genes to phenotypes, facilitating discovery of new gene functions. In this study, we describe a new version of Tn-seq method involving a modified EZ:Tn5 transposon for genome-wide and quantitative mapping of all insertions in a complex mutant library utilizing massively parallel Illumina sequencing. This Tn-seq method was applied to a genome-saturating *S. Typhimurium* mutant library recovered from selection under 3 different *in vitro* growth conditions (diluted LB medium, LB medium + bile acid, and LB medium at 42°C), mimicking some aspects of host stressors. We identified an overlapping set of 105 protein-coding genes in *S. Typhimurium* that are conditionally essential in at least one of the above selective conditions. Competition assays using 4 deletion mutants (*pyrD*, *glnL*, *recD* and STM14_5307) confirmed the phenotypes predicted by Tn-seq data, validating the utility of this approach in discovering new gene functions. With continuously increasing sequencing capacity of next generation sequencing technologies, this robust Tn-seq method will aid in revealing unexplored genetic determinants and the underlying mechanisms of various biological processes in *Salmonella* and the other approximately 70 bacterial species for which EZ:Tn5 mutagenesis has been established.

Key words: Tn-seq, functional genomics, conditionally essential genes, *Salmonella*

INTRODUCTION

In recent years an increasing number of complete genome sequences is becoming available for numerous bacterial species, mainly due to rapidly developing DNA sequencing technologies (MacLean *et al.*, 2009). This situation has raised pressing needs for understanding the functional significance of the numerous genes identified or predicted in these complete genomes. Bioinformatics analyses of the complete genomes have brought significant insights into our understanding of the biological implications of the genomic contents (Horner *et al.*, 2010). However, these approaches would provide little help in assigning functions to the genes with no homology to any known genes or predicting the genes responsible for the phenotypes that have never been explored yet. In this view, experimental approaches to study gene functions are essential tools in understanding the genome biology of bacteria. Particularly, high-throughput approaches to study gene functions in a genome-wide scale are especially attractive in this post-genomic era with overwhelming amounts of bacterial genomes to be explored, each containing thousands of genes.

Several high-throughput methods to study gene functions in bacteria have been developed using transposon mutagenesis based on the same framework of negative selection, yet using different strategies to compare mutant pools before and after selection (Bossé *et al.*, 2006). More recently, the extension of these strategies have undergone dramatic improvement in their capacity to determine gene functions in terms of the accuracies in genome mapping of transposon insertion sites and quantitative measurement of each insertion by employing next generation sequencing (Gawronski *et al.*, 2009; Goodman *et al.*, 2009; Langridge *et al.*, 2009; van Opijnen *et al.*, 2009; Christen *et al.*, 2011; Eckert *et al.*, 2011; Gallagher *et al.*, 2011). These methods use different procedures to capture transposon-junction sequences and sequence them

by massively parallel Illumina sequencing to accomplish an in-depth profiling of transposon-junction sequences in a complex mutant library. Although distinct names have been used, all these methods can be appropriately referred to as different variations on Tn-seq method (van Opijnen *et al.*, 2009; Gallagher *et al.*, 2011).

Here, we describe a new version of Tn-seq conveniently tailored to EZ:Tn5 transposon system, which has been broadly tested and used in more than 70 bacterial and archaeal species (<http://www.epibio.com/transcite.asp>). We applied this Tn-seq method to conduct genome-wide identification of *S. enterica* serotype Typhimurium (*S. Typhimurium*) genes that are conditionally essential for growth or survival under the 3 selected *in vitro* growth conditions. These conditions include low-nutrient condition, bile-rich environment, and body temperature of avian species (42°C), mimicking some aspects of the conditions in the host *Salmonella* cells are expected to encounter during infection. As a common human bacterial pathogen, *Salmonella* has to survive when it goes through a variety of stress conditions in the environments and the host during infection (Foster and Spector, 1995; Slauch *et al.*, 1997; Durant *et al.*, 1999; Rychlik and Barrow, 2005). Although a wealth of information on the genetic determinants and the underlying mechanisms of stress resistance have been obtained on *Salmonella*, there is no doubt that many gaps still exist in our knowledge and understanding in this area. This is especially true, in view of the most comprehensive phenomic profiling performed in *Escherichia coli* K12 recently (Nichols *et al.*, 2011).

This study demonstrated the utility and efficiency of this Tn-seq method for the comprehensive identification of conditionally essential genes in *Salmonella*. The genes identified here could be an important resource for better understanding or control of *Salmonella*, including development of novel antimicrobials and vaccines.

MATERIALS AND METHODS

Bacterial strains and culture conditions

S. Typhimurium 14028 wild type strain, isogenic deletion mutants strains, and a spontaneous mutant resistant to Nalidixic Acid (NA) were grown in Luria Bertani (LB) medium or LB agar plates and stored at -80°C in 30% glycerol. The cultures were incubated at 37°C unless described otherwise. Where appropriate, the LB agar plates contained NA (25µg/ml) or kanamycin (Km; 50µg/ml).

Construction of transposon mutant library

QuikChange Site-directed mutagenesis kit (Agilent Technologies La Jolla, CA) was used to change one nucleotide of pMODTM-6 <KAN-2/ MCS> plasmid (Epicentre BioTechnologies, Madison, WI) in one of the mosaic end (ME) sequences using the oligonucleotides in Table S1. This introduced the recognition sequence of TypeIIS enzyme BsmFI in one ME sequence (Figure 1A). The nucleotide change (A/T→G/C) was confirmed by DNA sequencing, and the modified plasmid, pMOD-BsmFI, was used for transposon mutagenesis of *S. Typhimurium* ATCC 14028 wild type strain. Briefly, pMOD-BsmFI was digested with Pvu II enzyme and the 1,116bp fragment of the modified EZ:Tn5TM (EZ:Tn5-BsmFI) was extracted from agarose gel using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) avoiding exposure of the fragment to UV light. The EZ:Tn5-BsmFI was then incubated with EZ-Tn5TM Transposase (Epicentre BioTechnologies) to form a transposon complex according to the instruction manual. Two µl of the complex was subsequently used to transform electrocompetent *S. Typhimurium* 14028 wild type cells by electroporation. Km-resistant transformants were selected on LB plates

supplemented with Km. The resulting mutants were combined to form a complex library of EZ:Tn5-BsmFI mutants containing approximately 1.6×10^4 mutants. The library was stored at -80°C in 30% glycerol.

In vitro selection of transposon mutant library

The mutant library was subjected to selection under 4 different *in vitro* conditions: LB medium, 10X diluted LB medium (dLB), LB medium supplemented with 5% crude ox bile extract (Sigma; LB-bile), and LB medium incubated at 42°C (LB- 42°C). To prepare the inoculum, 1 ml of the library in glycerol stock was diluted by adding 9 ml of LB medium and incubated at 37°C for 1 hr with vigorous shaking (rpm=225). After washing 3 times with Phosphate Buffered Saline (PBS) solution, the library was resuspended in 10 ml LB medium, and diluted with LB medium to reach $\text{OD}_{600} \approx 0.9$. An aliquot (2 ml) of the inoculum was used as the input pool. One ml of the inoculum ($\sim 7.3 \times 10^7$ CFU/ml) of the mutant library was used to inoculate each selection medium (100 ml in 200 ml Erlenmeyer flask) and incubated at the relevant temperature with vigorous shaking (rpm = 225). After 24 hrs, 1 ml of the culture was transferred to the fresh selection medium of the same kind. This selection was repeated 3 times to increase the selection sensitivity of the screening. After 3 consecutive selections, 2 ml of the final culture in each selection was collected to be used as the output pools. The cell pellets from one input and 4 output pools were used to isolate genomic DNA using DNA mini-kit (Qiagen, CA). The quantity and purity of the purified genomic DNA was measured using NanoDrop (Thermo Fisher Scientific, Wilmington, DE).

Sample preparation and Illumina sequencing

Extracted genomic DNA was digested with BsmFI restriction enzyme (New England BioLabs, Ipswich, MA) at 65°C for 3 hrs. After heat inactivation at 80°C for 20 min, the digested DNA was treated with calf intestinal alkaline phosphatase (NEB) at 37°C for 1hr to prevent self-ligation in the following ligation step by incubation for additional 1 hr. DNA was then phenol-chloroform extracted, ethanol precipitated and dissolved in 10 µl H₂O. DNA digests were subsequently ligated to Tn-seq linker, formed by annealing Tn-seq linker 1 and 2 oligonucleotides (Table S1), by overnight incubation at room temperature. The Tn-seq linker-ligated samples were subsequently used as templates in PCR reaction using cloned Pfu DNA polymerase (Agilent Technologies) with one of the 5 barcoded Tn5 primers and Tn-seq linker primer (Table S1). The PCR cycles consisted of initial denaturation at 94°C for 2 min, 5 cycles of 94°C for 30 sec, 55°C for 1 min, and 72°C for 30 sec, 20 cycles of 94°C for 30 sec, 62°C for 1 min, and 72°C for 30 sec followed by final extension at 72°C for 10 min. The amplicon of 129 bp long was PAGE-purified and dissolved in H₂O (Figure S1). The 5 DNA samples tagged with different barcodes were mixed with the same amount based on measurements with a NanoDrop apparatus (Thermo Fisher Scientific). The final mixed sample was analyzed using Affimetrix BioAnalyzer to check the quality and sequenced using Illumina Genome Analyzer II in the Institute for Integrative Genome Biology at the University of California at Riverside.

Analysis of Illumina sequencing data

A computer program was written in Python programming language to perform the following analysis (The Python script is available upon request). First, the sequence reads obtained from Illumina sequencing was sorted for the reads that contain a perfect 19 bp modified ME sequence

(5'-CTGTCCCTTATACACATCT-3'). Next, the filtered sequences were sorted according to the 6 nt barcode sequences demanding a perfect match to one of the 5 barcodes. The transposon-junction sequences were subsequently extracted from the filtered reads and the junction sequences of 11-13 bp long were selected for further analysis. These selected transposon-junction sequences were then mapped to the complete genome of *S. Typhimurium* 14028 (Jarvik *et al.*, 2010) (Accession Number: Chromosome CP001363.1, Plasmid CP001362.1) to select the reads that perfectly map to the genome. Additional filtering was performed to select the reads that map to the genome only in one locus. The final output data obtained by the Python script contained the information on the transposon-junction sequence, origin (chromosome vs. plasmid), genomic coordinate corresponding to EZ:Tn5-BsmFI insertion site, protein-coding gene containing the insertion within the internal 5 to 80% of the coding region, strand (+ vs. – strand), and the number of the reads in each pool for 5 mutant pools (1 input pool and 4 output pools).

The above output data were processed separately from this step for chromosome and plasmid data using JMP8 software (SAS, Cary, NC). For additional filtering, insertions with read counts in the input pool <10 were eliminated to remove nonspecific background reads. For normalization, a normalization factor was calculated according to the formula $(R_i/R_o)/(S_i/S_o)$, in which the variables represent the total number of sequence reads (R) and insertion sites (S) detected in the input (i) and output (o) pools, for each output pool (Gawronski *et al.*, 2009). The number of sequence reads for each insertion in each output pool was multiplied by the corresponding normalization factor. After normalization, the insertions in a gene with < 3 insertion sites within the internal 5 to 80% of the coding region were removed. The numbers of all normalized sequence reads within each gene were subsequently combined for each pool to obtain the total number of normalized sequence reads originated from each gene. Following this,

the fitness index value was calculated for each gene by dividing the total number of sequence reads in the output pool by those in the input pool.

Construction of deletion mutants

Single deletion mutants of *S. Typhimurium* 14028 carrying a deletion in *pyrD*, *glnL*, *recD*, and STM14_5307 were constructed with a Lambda Red recombination system by the method described by Cox et al. (Cox *et al.*, 2007) using pKD13 (Datsenko and Wanner, 2000) as a template for amplification of Km^R cassette and the oligonucleotides shown in Table S1. After confirmation by DNA sequencing, the deletions were transferred to the fresh wild type background by P22 transduction followed by selection on LB agar plates supplemented with Km. The final Km^R strains were purified on EBU plates to obtain phage-free colonies for further analysis.

Competition assays

Each deletion mutant ($\Delta pyrD$, $\Delta glnL$, $\Delta recD$, and $\Delta STM14_5307$) was competed against NA^R wild type strain in the control (LB, 37°C) or appropriate test conditions (dLB, 37°C for $\Delta pyrD$, $\Delta glnL$; LB-42°C for $\Delta recD$, $\Delta STM14_5307$). Briefly, 100 μ l of the inoculum consisting of equal volumes of the overnight cultures of wild type and a mutant was used to inoculate 100 ml of LB medium (control) or appropriate test medium in 200 ml Erlenmeyer flask. The culture was incubated at the indicated temperature with vigorous shaking (rpm = 225). After 24 hrs, 1 ml of the culture was transferred to the fresh medium and 0.1 ml of the culture was used for dilution and plating on LB agar plates (NA) and LB agar plates (Km) for selective counting of wild type and mutant strain, respectively. The LB agar plates were incubated overnight at 37°C and the

colonies were enumerated. This selection and counting procedure were repeated every 24 hours for up to approximately 4 to 8 days for different mutants.

RESULTS

Overview of the method

EZ:Tn5TM is a Tn5 derivative with modified inverted repeat (IR) sequences of 19 bp (AGATGTGTATAAGAGACAG), which was termed mosaic ends (ME). We observed that this ME region contains DNA sequence (underlined in the above ME sequence) similar to the recognition sequence of Type IIS restriction enzyme BsmFI (5'-GGGAC(N)₁₀↓-3'/3'-CCCTG(N)₁₄↑-5) except one nucleotide (Figure 1A). Since BsmFI cuts the site 14 bp away from the recognition site, this enzyme site can be exploited to extract 12 nt sequences immediately adjacent to Tn5 insertion sites. If these 12 bp transposon-junction sequences could be selectively amplified from a mutant library and sequenced *en mass* in a massively parallel manner, the resulting profile would provide the information on both identity and relative quantity of each insertion in the library. Therefore, we tested if the EZ:Tn5TM could be still functional even after the substitution of one nucleotide (A→G) in one ME region. We found that the efficiency of mutagenesis with the modified EZ:Tn5TM (EZ:Tn5-BsmFI) as measured by the number of resulting Km^R colonies was very similar to that of original EZ:Tn5TM. In repeated transformation experiments, we routinely obtained approximately 1 to 3 x 10⁴ mutants of *S. Typhimurium* 14028 per electroporation.

This would provide an excellent opportunity to use this EZ:Tn5-BsmFI as a powerful tool for deep profiling of complex insertion mutant library via Illumina sequencing as shown in Figure 1B. A genome-saturating library of Tn5 mutants is subjected to a selection condition of interest

and the mutant pools before and after the selection can be compared to identify all insertions exhibiting any significant changes in relative abundance after selection. This analysis will allow identification of bacterial genes important in various biological processes of interest on a genome-wide scale. However, the length of the transposon-junction sequences extracted by this method will be 12 bp, which may not be long enough for unambiguous identification of the genomic locations from which the insertions were originated. To address this issue, we performed a computer simulation using a custom Python script to extract 100,000 short sequences of different length (10-20 bp) from random locations in both strands of the complete genome of *S. Typhimurium* 14028 strain. The short sequences were subsequently mapped back to the genome to determine the portion of the sequences that perfectly match to the genome only in one location (Figure S2). The result shows that approximately 47% of the 12 bp-transposon junction sequences mapped uniquely to the genome, suggesting that on average one half of the transposon-junction sequences experimentally extracted from a mutant library should be discarded because they cannot be mapped unambiguously to one genomic locus. However, we expected that this shortcoming can be overcome by increasing the size of the mutant library and the huge number of sequence reads that can be obtained from Illumina sequencing.

Selection of the mutant library

To test the feasibility and utility of this method, we used this method for genome-wide identification of *S. Typhimurium* genes conditionally essential for fitness under 3 different *in vitro* conditions. Three selective conditions in this study were chosen to mimic low-nutrient condition in infected host tissues (10X diluted LB medium; dLB), bile-rich intestinal environment (LB medium + bile acid; LB-bile), and the elevated body temperature associated

with chicken and other avian species (LB medium at 42°C; LB-42°C). Our EZ:Tn5-BsmFI mutant library consists of approximately 1.6×10^4 different mutants, and the inoculum of 1 ml contained approximately 7.3×10^7 cells, indicating each mutant in the library was represented by approximately 4,600 cells. Mutants were transferred 3 times in the same selective conditions to increase the selection sensitivity. In each transfer during the selection, the inoculum of 1 ml contained approximately 10^8 to 10^9 cells depending on the selective conditions, indicating each mutant is well represented by sufficient number of cells in the inoculum. The final output pool of 2 ml also contained enough cells to represent all surviving mutants at the end of the final selection. We also included LB medium incubated at 37°C as a control condition to identify the genes that are essential for fitness under an optimal growth condition. Since we are interested in the genes uniquely essential under the 3 selective conditions of interest, the genes identified in control condition were removed from those identified in each of 3 selective conditions.

Analysis of Illumina sequencing data

The summary of the Illumina sequencing data and its analysis is shown in Table S2. Among the total of 23,141,540 sequence reads obtained from a single flowcell lane, 76% (17,490,113 reads) contained the complete 19 bp ME sequence. Among these reads, 15,893,768 reads (91%) contained the complete 6 bp barcode sequences perfectly matching one of the 5 barcodes. When these reads were sorted according to the barcode, we obtained a relatively even distribution across different barcodes: 2,374,190 read (ATCACG; Input), 2,859,671 (CGATGT; LB), 2,652,543 (TTAGGC; dLB), 4,382,503 (TGACCA; LB-bile), and 3,624,861 (ACAGTG; LB-42°C). The transposon-junction sequences were subsequently extracted from the reads for each barcode. As expected the majority (>99%) of the sequence reads were approximately 11 to 13 bp

long (Table S2; Figure S3). These 11 to 13 bp sequence reads for each barcode were further filtered for those that map to the genome at only one genomic loci. Finally, we obtained 1,204,021, 1,312,302, 1,367,315, 2,249,816 and 1,554,249 reads for each barcode, which corresponds to approximately 49% of the total number of the transposon-junction sequences of 11 to 13bp for each barcode.

The normalization factors for the output pools were 0.94 (LB), 0.89 (dLB), 0.57 (LB-bile), and 0.80 (LB-42°C) for chromosome and 0.55 (LB), 0.61 (dLB), 0.23 (LB-bile), and 0.49 (LB-42°C) for plasmid. When the reads for all insertions in each gene were combined, we obtained the data set for 3,806 and 90 protein-coding genes for chromosome and plasmid, respectively. To obtain a more robust and reliable result, the genes that contained less than 3 insertions were removed, resulting in a total of 1,879 and 52 coding genes for chromosome and plasmid, respectively. For all of the genes included in the final dataset, each gene contained 8.8 and 9.1 authentic insertion sites on an average on the chromosome and plasmid, respectively. We also determined the reproducibility of Tn-seq profiling using two biological replicates. When the data was processed separately for chromosome and plasmid, we obtained very high levels of reproducibility for both chromosome ($R^2 = 0.99$) and plasmid ($R^2 = 0.99$) (Figure S4). As the level of genome saturation by insertions in this final dataset was not sufficiently high, we did not investigate *in vitro* essential genes (Hutchison *et al.*, 1999; Glass *et al.*, 2006) but focused on conditionally essential genes in this study.

Identification of conditionally essential genes

Among the 1,931 (1,879 + 52) genes, the genes conditionally essential for each selective condition were first selected by a cut-off fitness index ≤ 0.2 , which indicates at least a 5-fold

reduction in relative abundance during selection. The genes selected by this criterion are expected to exhibit strong fitness defect under each condition. When the mutant were selected in LB medium, which was used as a reference condition for optimal growth, a total of 32 genes were identified as required for optimal *in vitro* fitness. These genes should be distinguished from essential genes because the mutants with insertions in the essential genes cannot be recovered by definition of the essential genes. However, the insertion mutants in these 32 genes were recovered and were well represented in the input pool. Therefore, these genes should be considered dispensable for growth or survival, yet contribute to optimal growth at 37°C in LB medium.

The result in Figure 2 gives a clear overview of the genome-wide fitness profile of *Salmonella* mutants. We obtained 39, 61, and 56 genes required for fitness in dLB, LB-bile, and LB-42°C, respectively (Figure 3A). There were many genes essential for fitness in more than one condition and many genes were also shown to be essential for general fitness under optimal growth condition (LB at 37°C; numbers shown in parenthesis in Figure 3A). In total, 105 genes conditionally essential under either one of the selective conditions were identified (Figure 3A; see Table S3 for the list of the genes). All of the 105 genes identified were chromosomal genes and none of them were located on plasmid.

To obtain insights on functional trends associated with each condition, we assigned the identified genes to functional (COG; Clusters of Orthologous Groups) categories (Tatusov *et al.*, 1997) using BLAST on Orthologous groups (BLASTO) algorithm (Zhou and Landwebber, 2007) (Figure 3B). Not surprisingly, the *in vitro* essential genes (LB medium) were most prominently enriched in COG categories K (Transcription), M (Cell wall/ membrane/ envelope biogenesis), and C (Energy production and conversion). The genes essential for fitness in dLB

exhibited a similar trend except that category K was not enriched while category M was further enriched to represent 25% of all genes identified. In LB-bile, the genes required for fitness were significantly biased toward cell wall/ membrane/ envelope biogenesis (category M). This result corroborates well with previous findings that LPS biosynthesis and membrane integrity are critical in bile resistance of *Salmonella* (Prouty *et al.*, 2002; Langridge *et al.*, 2009).

There are 48 (61 minus 13) genes identified as conditionally essential for fitness in the presence of bile salts. Bile resistance has been studied in-depth, and extensive mutant screenings have revealed a comprehensive list of genes required for bile resistance in *S. Typhimurium* (van Velkinburgh and Gunn, 1999; Prouty *et al.*, 2002; López-Garrido *et al.*, 2010), *S. Typhi* (van Velkinburgh and Gunn, 1999; Langridge *et al.*, 2009;) and *Escherichia coli* (Nichols *et al.*, 2011). Notably, 38 out of the 48 genes identified in this study were previously implicated in bile resistance in *Salmonella* species or *E. coli*. This result validates our experimental and bioinformatics approaches to identify conditionally essential genes. The remaining 10 genes newly identified in this study may reflect the differences in the sensitivity of the screening and selection conditions between the experiments.

Phenotypic characterization of deletion mutants

To further verify the result of the Tn-seq screening, we sought to characterize the functions of additional genes that were identified in this study, which have not been previously linked to the phenotypes. We chose the two genes, *pyrD* and *glnL* encoding dihydroorotate dehydrogenase and nitrogen regulation protein, respectively, among the 39 genes identified in dLB medium. The deletion mutants were analyzed for growth patterns in LB medium (control condition) and dLB medium. Unexpectedly, the $\Delta pyrD$ mutant demonstrated a slight growth defect in both LB and

dLB media compared to the wild type (Figure S5). Conversely, $\Delta glnL$ exhibited a slight growth defect only in LB medium (Figure S5). The reason why the growth phenotypes did not accurately reflect those predicted from Tn-seq data could be due to the difference in the assay conditions. Initially the negative selection of mutants was performed through competition within the complex mutant library. Therefore, in order to more closely mimic the conditions in which the mutants were selected, we characterized the 2 mutants using competition assays in which each mutant was competed against the wild type strain. As expected, both $\Delta pyrD$ and $\Delta glnL$ mutants exhibited a severe competitive disadvantage in dLB medium, but not in LB medium (Figure 4A and 4B, respectively).

Additionally, two genes *recD* and STM14_5307 encoding exonuclease V subunit and putative transcriptional regulator, respectively, were chosen from the genes identified in LB-42°C and the deletion mutants were subjected to competition assays. The $\Delta recD$ mutant showed slight competitive disadvantage at 37°C, but more obvious disadvantage during competition at 42°C (Figure 4C). In case of the mutant $\Delta STM14_5307$, the mutant demonstrated a clear competitive disadvantage even at 37°C (Figure 4D), yet a more severe competitive disadvantage was observed at 42°C (Day 7: Figure 4D).

To determine the accuracy of fitness measurement inferred by the Tn-seq profiles, we compared the fitness indices obtained by Tn-seq data and competition assay for each gene. As shown in Figure 5, Tn-seq data were able to predict the fitness of each mutant strain with high levels of accuracy.

Insertions in the conditionally essential genes are not lethal

We analyzed our data in comparison with those from *in vitro* metabolic reconstruction (MRs) modeling (Feist *et al.*, 2009). A metabolic reconstruction breaks down all known metabolic pathways in the cell into their respective reactions and enzymes, and analyzes them within the perspective of the entire system. One such MR recently reported on *Salmonella* Typhimurium LT2 (Thiele *et al.*, 2011) predicted 144 genes lethal for growth or survival in LB media. We found that the 21 genes out of the 105 genes identified in our study were also listed as lethal genes. However, construction of deletion mutants have been reported previously for at least 12 of those 21 genes (Yethon *et al.*, 2000; Gantois *et al.*, 2009; Karasova *et al.*, 2009; Su *et al.*, 2009; Kong *et al.*, 2011) suggesting that the prediction of lethal genes by this MR is not accurate. The non-essential nature of the 105 conditionally essential genes identified in this study is also supported by additional experimental evidence by Langridge *et al.* (2009), in which 356 essential genes were discovered in *S. Typhi* with high confidence using 1.1 million random Tn5 insertions in the genome. Among the 105 genes identified in our study, only two genes (*icdA* and *pssA*) were reported as essential and 29 genes were classified as advantageous for growth in LB medium (Langridge *et al.*, 2009).

Biological roles of the identified conditionally essential genes during host infection

As an important bacterial pathogen with a battery of genetic tools available, *S. Typhimurium* has been used commonly as a model organism to study bacterial pathogenesis. Consequently, a large quantity of ‘omics’ (genomics, transcriptomics, and proteomics) data have been accumulated for *S. Typhimurium* and other related serotypes (Gillespie *et al.*, 2011). Particularly, a large portion of those studies has been conducted in the context of host-pathogen interactions.

In order to gain deeper insights on the biological implications of the genes identified in this study, we analyzed them in light of those data. We found the results obtained from comprehensive functional profiling of mutant libraries (random insertions or targeted deletions) obtained using animal infection models would be particularly useful in shedding insights on the role of the genes and their products during host infection. These large-scale high-throughput screenings of *Salmonella* mutants have provided almost saturating lists of the *Salmonella* genes essential for *in vivo* infection using different infection models of mouse (Chan, *et al.*, 2005; Lawley *et al.*, 2006; Chaudhuri *et al.*, 2009; Santiviago *et al.*, 2009). For a large portion of these *in vivo* essential genes, however, the biochemical bases for attenuation of the mutants are unknown. Our study revealed comprehensive sets of conditionally essential genes with relevance to stress resistance during host infection. Therefore, the comparison of our data with the existing functional profiling data would provide the biological basis for the requirement of at least a subset of the genes essential for *in vivo* infection.

For each of the 105 genes identified as conditionally essential in this study, we determined if they have been previously identified as essential for *in vivo* fitness during infection in animal models (mouse or chicken). One requirement for a gene to have a role for *in vivo* fitness is that the protein encoded by the gene should be expressed *in vivo*. Therefore, we also used the comprehensive *in vivo* proteomics data of *S. Typhimurium* obtained using both systemic and enteric infection models of mice to determine if protein encoded by each of the 105 genes was expressed *in vivo* during infection in mouse (Becker *et al.*, 2006). However, no *in vivo* protein expression data is available for chicken. The summary of the functional information and proteome data is shown in Figure 6 (Table S3-S6 for more details).

The low nutrient condition reflected in dLB medium is encountered by *Salmonella* cells when they are present inside macrophage vacuoles during systemic infection (Foster and Spector, 1995). Out of 24 genes essential for fitness in dLB, 19 genes were shown to be essential for *in vivo* fitness during a systemic infection in the mouse, among which *in vivo* protein expression during systemic infection was detected for 14 genes (Figure 6A; Table S4). Five proteins previously shown to be required for *in vivo* survival during systemic infection in mice (RecG, GlnL, LepA, RfaQ, and ZntA) were not detected *in vivo* probably due to the limited amount and high complexity of the sample (Becker *et al.*, 2006). Two proteins among the 16 proteins detected *in vivo* (Pnp, and OmpA) are not required for fitness during systemic infection. This analysis suggests that the 19 genes were required for *in vivo* fitness during systemic infection in mice due to their requirements for fitness under low-nutrient condition. However, none of the 24 genes were shown to be important during systemic infection in chicken (Figure 6B). This reflects the fact that all large scale mutant screenings have been performed with the mouse systemic infection model and the number of mutants screened in chickens is very limited (Turner *et al.*, 1998; Shah *et al.*, 2005).

In case of LB-bile, *Salmonella* should encounter bile stress in the intestinal tract of the host during enteric infection (Foster and Spector, 1995). In enteric infection model of mouse, 25 genes out of the total 48 genes required for bile resistance *in vitro* were shown to be expressed from the cecal samples of the infected mouse (Figure 6C; Table S5). Among the 25, 5 proteins (RfaL, RfaJ, RfaI, Rfc, and RfbP) were shown to be important for *in vivo* fitness. Interestingly, all these 5 proteins are involved in biosynthesis of lipopolysaccharide core and O-antigen (Kong *et al.*, 2011). The remaining 20 genes have not been linked to enteric infection in mouse, which probably reflects the lack of comprehensive screening conducted with enteric infection model of

mouse due to the technical difficulty associated with bottle neck existing in enteric infection model. However, the fact that these 20 genes are both expressed *in vivo* and required for bile resistance *in vitro* strongly suggests that these genes play roles in efficient colonization in the intestinal tract via conferring resistance to bile. Interestingly, 12 of the 48 genes were shown to be important for cecal colonization during chicken infection (Figure 6D). Nine of the 12 genes (*acrB*, *rfbN*, *rfbD*, *rfbB*, *tolC*, *rfbI*, *rfbK*, *rbsK*, and *rfbP*) were shown to be expressed *in vivo* in enteric infection model of mouse, yet only one of them (*rfbP*) has been functionally linked to enteric infection model of mouse (Table S5).

The elevated temperature of 42°C is the body temperature of avian species including chickens. Therefore, any mutant with fitness defect at 42°C is very likely to be attenuated during infection in a chicken. Among the 40 genes identified as essential for fitness at 42°C in this study, only 2 genes were previously implicated with *in vivo* fitness defect during chicken infection (Figure 6E; Table S6). The result suggests that these 2 genes, *rfaY* and *rfbP*, are required for infection in chicken through their requirements for fitness at 42°C. However, one of the proteins (RfbP) is also required for both systemic and enteric infection in mouse, whose body temperature is 37°C, indicates that there is other mechanism(s) underlying the attenuation of the mutants during host infection (Table S6). For example, the fitness index for the *rfbP* gene was 0.45 and 0.01 for dLB and LB-bile, respectively, indicating that this gene is also required for fitness under low-nutrient conditions (although the fitness index was >0.2) and bile-rich environments.

DISCUSSION

In this study, we described a new version of Tn-seq method and used it with stringent cutoffs to obtain robust fitness profiling of entire genome of *S. Typhimurium* 14028 strain under 3

different *in vitro* stress conditions, including low-nutrient, bile-rich, and high temperature (42°C) environments. The phenotype characterization of the 4 deletion mutants along with previous studies on bile acid resistance of *Salmonella* demonstrated that our Tn-seq method and bioinformatics analysis assessed gene functions accurately.

This method as a variation on existing Tn-seq methods is based on the use of modified EZ:Tn5 transposon that carries recognition site of TypeIIIS enzyme BsmFI on one ME sequence to allow straightforward and robust extraction of Tn5-junction sequences of the identical length of 12nt from a complex mutant library. We used our Tn-seq method in conjunction with barcoding strategy and Illumina sequencing to allow high-resolution functional genome scanning for multiple selective conditions of interest.

There are other existing methods for comprehensive functional screening of a transposon mutant library with the aid of next generation sequencing (Gawronski *et al.*, 2009; Goodman *et al.*, 2009; Langridge *et al.*, 2009; van Opijnen *et al.*, 2009; Christen *et al.*, 2011; Eckert *et al.*, 2011; Gallagher *et al.*, 2011). Our method is distinct from other methods in that we employed EZ:Tn5 mutagenesis system. The proven broad-host range of this mutagenesis system along with a commercially available EZ-Tn5TM Transposase (Epicentre BioTechnologies) will make this method easily accessible and applicable to variety of bacterial species with appropriate modifications on an antibiotics cassette and its promoter. EZ:Tn5 system was also used in the method developed by Langridge *et al.* (2009), and the strategies developed by Gawronski *et al.* (2009), Gallagher *et al.* (2011), Eckert *et al.* (2011), and Christen *et al.* (2011) should be applicable to EZ:Tn5 system with appropriate modifications. However, our Tn-seq method is technically more simple and straightforward as compared to other methods involving mechanical shearing and fractionation of DNA fragments (Gawronski *et al.*, 2009; Langridge *et al.*, 2009;

Eckert *et al.*, 2011; Gallagher *et al.*, 2011). In addition, the uniform length of PCR amplicons obtained by our protocol provides an efficient means to remove fragments resulting from a possible aberrant PCR reaction.

One drawback of our method is the relatively short length of the transposon-junction sequences, which made it necessary to discard approximately one half of the transposon-junction sequences mapped to the genome. However, increasing the number of mutants in the library along with more sequencing reads would eventually overcome this problem. Particularly with rapidly increasing sequencing capacity of next generation sequencing technologies, this will become a negligible issue (Metzker, 2010).

We also demonstrate for the first time that this Tn-seq approach could be used in conjunction with barcodes to analyze multiple samples simultaneously. This approach will allow high-resolution functional screening of a bacterial genome for multiple selective conditions of interest, opening the door for multidimensional comprehensive understanding of bacterial gene functions.

In this study, we modified the DNA sequence of plasmid pMODTM-6 <KAN-2/ MCS> to obtain the EZ:Tn5-BsmFI fragment. Alternatively, EZ:Tn5-BsmFI could be prepared by amplifying the template plasmid pMODTM-6 <KAN-2/ MCS> or any other derivative plasmids using a pair of primers corresponding to the ME sequences where one of the primers contain one nucleotide change to introduce BsmFI site into the ME region (unpublished data). This can simplify the procedure by eliminating the step for site-directed mutagenesis.

We have identified 105 genes conditionally essential for fitness under 3 different conditions reflecting stressors *Salmonella* would encounter during survival in the environments and infected hosts. We examined the biological significance of these genes during host infection by analyzing the data in light of currently available mutant fitness data and proteomics data obtained from

animal infection studies. Through this analysis, we assigned biological bases for *in vivo* requirements of the proteins for all 58 genes among the 105 genes identified in this study. This process resembles the virulence-attenuated pool (VAP) screening using signature-tagged mutagenesis (STM), which is performed with a subset of the original mutants shown to be attenuated *in vivo* to reveal roles that the identified factors play in the infection process (Merrell and Camilli, 2002; Merrell *et al.*, 2002). For VAP screening based on STM, the need to reduce the pool for further screening came from the limited (~96) number of mutants that can be screened simultaneously by STM. However, the global scale and cost-effectiveness of Tn-seq method eliminates the need to prepare a smaller size of VAP pool for secondary screening under various stress conditions. Instead, a complex library of transposon mutants could be screened simultaneously or one at a time under multiple conditions including animal infection and other host-associated stressors. If this approach is used in an animal infection model in conjunction with screenings for a variety of virulence-associated phenotypes representative of the all known host barriers to overcome for successful infection, it is expected to provide a wealth of functional information for most of the *in vivo* essential factors.

The genes identified in this study include many putative or hypothetical genes or the genes with unknown functions. Understanding the functions of these genes and products is expected to reveal unknown mechanisms of *Salmonella* survival and persistence during its life cycle. In addition, these genes have great potentials to be used as good candidate targets for development of vaccines and novel antimicrobials. The mutants with deletions in some of those genes that would still allow *in vivo* survival, yet at appropriately reduced levels could be very effective in eliciting an adaptive immune response in the host, while they are likely be cleared from the host

faster than the wild type strain. It would be interesting to test *S. Typhimurium* mutant with reduced fitness at 42°C as an attenuated live vaccine for poultry.

With rapidly increasing sequencing capacity of the 2nd generation sequencing technologies (Metzker, 2010) and emergence of the 3rd generation sequencing methods with even greater potential (Rothberg *et al.*, 2011) our Tn-seq method will allow exploration into comprehensive understanding of the functional implications of genetic elements (both coding and non-coding genes) at increasingly higher resolution for a variety of biological contexts (Christen *et al.*, 2011).

ACKNOWLEDGEMENTS

This work was supported by NIH R21 AI063137 grant and a USDA Food Safety Consortium grant.

REFERENCES

- Becker, D., Selbach, M., Rollenhagen, C., Ballmaier, M., Meyer, T.F., Mann, M., and Bumann, D. (2006) Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* **440**: 303-307.
- Bossé, J.T., Zhou, L., Kroll, J.S., and Langford, P.R. (2006) High-throughput identification of conditionally essential genes in bacteria: from STM to TSM. *Infect Disord Drug Targets* **6**: 241-262.
- Chan, K., Kim, C.C., and Falkow, S. (2005) Microarray-based detection of Salmonella enterica serovar Typhimurium transposon mutants that cannot survive in macrophages and mice. *Infect Immun* **73**: 5438-5449.
- Chaudhuri, R.R., Peters, S.E., Pleasance, S.J., Northen, H., Willers, C., Paterson, G.K., *et al.* (2009) Comprehensive identification of Salmonella enterica serovar typhimurium genes required for infection of BALB/c mice. *PLoS Pathog* **5**: e1000529.
- Christen, B., Abeliuk, E., Collier, J.M., Kalogeraki, V.S., Passarelli, B., Collier, J.A., *et al.* (2011) The essential genome of a bacterium. *Mol Syst Biol* **7**: 528.
- Cox, M.M., Layton, S.L., Jiang, T., Cole, K., Hargis, B.M., Berghman, L.R., *et al.* (2007) Scarless and site-directed mutagenesis in Salmonella enteritidis chromosome. *BMC Biotechnol* **7**: 59.
- Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640-6645.
- Durant, J.A., Corrier, D.E., Byrd, J.A., Stanker, L.H., and Ricke, S.C. (1999) Feed deprivation affects crop environment and modulates Salmonella enteritidis colonization and invasion of leghorn hens. *Appl Environ Microbiol* **65**: 1919-1923.
- Eckert, S.E., Dziva, F., Chaudhuri, R.R., Langridge, G.C., Turner, D.J., Pickard, D.J., *et al.* (2011) Retrospective application of transposon-directed insertion site sequencing to a library of signature-tagged mini-Tn5Km2 mutants of Escherichia coli O157:H7 screened in cattle. *J Bacteriol* **193**: 1771-1776.
- Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., and Palsson, B. (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* **7**: 129-143.
- Foster, J.W., and Spector, M.P. (1995) How Salmonella survive against the odds. *Annu Rev Microbiol* **49**: 145-174.
- Gallagher, L.A., Shendure, J., and Manoil, C. (2011) Genome-Scale Identification of Resistance Functions in Pseudomonas aeruginosa Using Tn-seq. *MBio* **2**: e00315-e00310.

- Gantois, I., Ducatelle, R., Pasmans, F., Haesebrouck, F., and Van Immerseel, F. (2009) The Salmonella Enteritidis lipopolysaccharide biosynthesis gene rfbH is required for survival in egg albumen. *Zoonoses Public Health* **56**: 145-149.
- Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci USA* **106**: 16422-16427.
- Gillespie, J.J., Wattam, A.R., Cammer, S.A., Gabbard, J.L., Shukla, M.P., Dalay, O., *et al.* (2011) PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* **79**: 4286-4298.
- Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, C., *et al.* (2006) Essential genes of a minimal bacterium. *Proc Natl Acad Sci USA* **103**: 425-430.
- Goodman, A.L., McNulty, N.P., Zhao, Y., Leip, D., Mitra, R.D., Lozupone, C.A., *et al.* (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell Host Microbe* **6**: 279-289.
- Horner, D.S., Pavesi, G., Castrignanò, T., De Meo, P.D., Liuni, S., Sammeth, M., *et al.* (2010) Bioinformatics approaches for genomics and post genomics applications of next-generation sequencing. *Brief Bioinform* **11**: 181-197.
- Hutchison, C.A., Peterson, S.N., Gill, S.R., Cline, R.T., White, O., Fraser, C.M., *et al.* (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. *Science* **286**: 2165-2169.
- Jarvik, T., Smillie, C., Groisman, E.A., and Ochman, H. (2010) Short-term signatures of evolutionary change in the Salmonella enterica serovar typhimurium 14028 genome. *J Bacteriol* **192**: 560-567.
- Karasova, D., Sebkova, A., Vrbas, V., Havlickova, H., Sisak, F., and Rychlik, I. (2009) Comparative analysis of Salmonella enterica serovar Enteritidis mutants with a vaccine potential. *Vaccine* **27**: 5265-5270.
- Kong, Q., Yang, J., Liu, Q., Alamuri, P., Roland, K.L., and Curtiss, R. (2011) Effect of deletion of genes involved in lipopolysaccharide core and O-antigen synthesis on virulence and immunogenicity of Salmonella enterica serovar typhimurium. *Infect Immun* **79**: 4227-4239.
- Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., *et al.* (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* **19**: 2308-2316.

- Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R., and Monack, D.M. (2006) Genome-wide screen for Salmonella genes required for long-term systemic infection of the mouse. *PLoS Pathog* **2**: e11.
- López-Garrido, J., Cheng, N., García-Quintanilla, F., García-del Portillo, F., and Casadesús, J. (2010) Identification of the Salmonella enterica damX gene product, an inner membrane protein involved in bile resistance. *J Bacteriol* **192**: 893-895.
- MacLean, D., Jones, J.D., and Studholme, D.J. (2009) Application of 'next-generation' sequencing technologies to microbial genetics. *Nat Rev Microbiol* **7**: 287-296.
- Merrell, D.S., and Camilli, A. (2002) Information overload: assigning genetic functionality in the age of genomics and large-scale screening. *Trends Microbiol* **10**: 571-574.
- Merrell, D.S., Hava, D.L., and Camilli, A. (2002) Identification of novel factors involved in colonization and acid tolerance of Vibrio cholerae. *Mol Microbiol* **43**: 1471-1491.
- Metzker, M.L. (2010) Sequencing technologies-the next generation. *Nat Rev Genet* **11**: 31-46.
- Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell* **144**: 143-156.
- Prouty, A.M., Van Velkinburgh, J.C., and Gunn, J.S. (2002) Salmonella enterica serovar typhimurium resistance to bile: identification and characterization of the tolQRA cluster. *J Bacteriol* **184**: 1270-1276.
- Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., *et al.* (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**: 348-352.
- Rychlik, I., and Barrow, P.A. (2005) Salmonella stress management and its relevance to behaviour during intestinal colonisation and infection. *FEMS Microbiol Rev* **29**: 1021-1040.
- Santiviago, C.A., Reynolds, M.M., Porwollik, S., Choi, S.H., Long, F., Andrews-Polymenis, H.L., and McClelland, M. (2009) Analysis of pools of targeted Salmonella deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog* **5**: e1000477.
- Shah, D.H., Lee, M.J., Park, J.H., Lee, J.H., Eo, S.K., Kwon, J.T., and Chae, J.S. (2005) Identification of Salmonella gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology* **151**: 3957-3968.
- Slauch, J., Taylor, R. and Maloy, S. (1997) Survival in a cruel world: how Vibrio cholerae and Salmonella respond to an unwilling host. *Genes Dev* **11**: 1761-1774.

- Su, J., Timbely, D., Zhu, M., Hua, X., Liu, B., Pang, Y., *et al.* (2009) RfaB, a galactosyltransferase, contributes to the resistance to detergent and the virulence of *Salmonella enterica* serovar Enteritidis. *Med Microbiol Immunol* **198**: 185-194.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**:631-637.
- Thiele, I., Hyduke, D.R., Steeb, B., Fankam, G., Allen, D.K., Bazzani, S., *et al.* (2011) A community effort towards a knowledge-base and mathematical model of the human pathogen *Salmonella Typhimurium* LT2. *BMC Syst Biol* **5**: 8.
- Turner, A.K., Lovell, M.A., Hulme, S.D., Zhang-Barber, L., and Barrow, P.A. (1998) Identification of *Salmonella typhimurium* genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks. *Infect Immun* **66**: 2099-2106.
- van Opijnen, T., Bodi, K.L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**: 767-772.
- van Velkinburgh, J.C., and Gunn, J.S. (1999) PhoP-PhoQ-regulated loci are required for enhanced bile resistance in *Salmonella* spp. *Infect Immun* **67**: 1614-1622.
- Yethon, J.A., Gunn, J.S., Ernst, R.K., Miller, S.I., Laroche, L., Malo, D., and Whitfield, C. (2000) *Salmonella enterica* serovar typhimurium waaP mutants show increased susceptibility to polymyxin and loss of virulence In vivo. *Infect Immun* **68**: 4485-4491.
- Zhou, Y., and Landweber, L.F. (2007) BLASTO: a tool for searching orthologous groups. *Nucleic Acids Res* **35**: W678-W682.

REFERENCES FOR SUPPLEMENTARY MATERIALS

1. Becker, D., Selbach, M., Rollenhagen, C., Ballmaier, M., Meyer, T.F., Mann, M. and D. Bumann, D. (2006) Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* **440**: 303-307.
2. Buckley, A.M., Webber, M.A., Cooles, S., Randall, L.P., La Ragione, R.M., Woodward, M.J., and Piddock, L.J. (2006) The AcrAB-TolC efflux system of Salmonella enterica serovar Typhimurium plays a role in pathogenesis. *Cell Microbiol* **8**: 847-856.
3. Chan, K., Kim, C.C., and Falkow, S. (2005) Microarray-based detection of Salmonella enterica serovar Typhimurium transposon mutants that cannot survive in macrophages and mice. *Infect Immun* **73**: 5438-5449.
4. Chaudhuri, R.R., Peters, S.E., Pleasance, S.J., Northen H., Willers, C., Paterson, G.K., *et al.* (2009). Comprehensive identification of Salmonella enterica serovar typhimurium genes required for infection of BALB/c mice. *PLoS Pathog* **5**: e1000529.
5. Kong, Q., Yang, J., Liu, Q., Alamuri, P., Roland, K.L., and Curtiss, R. (2011) Effect of deletion of genes involved in lipopolysaccharide core and O-antigen synthesis on virulence and immunogenicity of Salmonella enterica serovar typhimurium. *Infect Immun* **79**: 4227-4239.
6. Langridge, G.C., Phan, M.D., Turner, D.J., Perkins, T.T., Parts, L., Haase, J., *et al.* (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* **19**: 2308-2316.
7. Lawley, T.D., Chan, K., Thompson, L.J., Kim, C.C., Govoni, G.R., and Monack, D.M. (2006) Genome-wide screen for Salmonella genes required for long-term systemic infection of the mouse. *PLoS Pathog* **2**: e11.
8. López-Garrido, J., Cheng, N., García-Quintanilla, F., García-del Portillo, F., and Casadesús, J. (2010) Identification of the Salmonella enterica damX gene product, an inner membrane protein involved in bile resistance. *J Bacteriol* **192**: 893-895.
9. Morgan, E., Campbell, J.D., Rowe, S.C. , Bispham, J., Stevens, M.P., Bowen, A.J., *et al.* (2004) Identification of host-specific colonization factors of Salmonella enterica serovar Typhimurium. *Mol Microbiol* **54**: 994-1010.
10. Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., *et al.* (2011) Phenotypic landscape of a bacterial cell. *Cell* **144**: 143-156.
11. Prouty, A.M., Van Velkinburgh, J.C., and Gunn, J.S. (2002) Salmonella enterica serovar typhimurium resistance to bile: identification and characterization of the tolQRA cluster. *J Bacteriol* **184**: 1270-1276.

12. Santiviago, C.A., Reynolds, M.M., Porwollik, S., Choi, S.H., Long, F., Andrews-Polymenis, H.L., and McClelland, M. (2009) Analysis of pools of targeted Salmonella deletion mutants identifies novel genes affecting fitness during competitive infection in mice. *PLoS Pathog* **5**: e1000477.
13. Shah, D.H., Lee, M.J., Park, J.H., Lee, J.H., Eo, S.K., Kwon, J.T., and Chae, J.S. (2005) Identification of Salmonella gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology* **151**: 3957-3968.
14. Su, J., Timbely, D., Zhu, M., Hua, X., Liu, B., Pang, Y., *et al.* (2009) RfaB, a galactosyltransferase, contributes to the resistance to detergent and the virulence of Salmonella enterica serovar Enteritidis. *Med Microbiol Immunol* **198**: 185-194.
15. Sydenham, M., Douce, G., Bowe, F., Ahmed, S., Chatfield, S., and Dougan, G. (2000) Salmonella enterica serovar typhimurium surA mutants are attenuated and effective live oral vaccines. *Infect Immun* **68**: 1109-1115.
16. Turner, A.K., Lovell, M.A., Hulme, S.D., Zhang-Barber, L., and Barrow, P.A. (1998) Identification of Salmonella typhimurium genes required for colonization of the chicken alimentary tract and for virulence in newly hatched chicks. *Infect Immun* **66**: 2099-2106.
17. van Velkinburgh, J.C., and Gunn, J.S. (1999) PhoP-PhoQ-regulated loci are required for enhanced bile resistance in Salmonella spp. *Infect Immun* **67**: 1614-1622.

Tables

- Table S1.** Oligonucleotides used in this study
- Table S2.** Summary of Illumina sequencing data analysis
- Table S3.** List of all conditionally essential genes identified in this study (fitness index ≤ 0.2 under at least one of the conditions)
- Table S4.** The genes conditionally essential for fitness in dLB medium. The genes also essential for fitness in LB were removed from the list.
- Table S5.** The genes conditionally essential for fitness in LB-bile medium. The genes also essential for fitness in LB were removed from the list.
- Table S6.** The genes conditionally essential for fitness in LB-42°C medium. The genes also essential for fitness in LB were removed from the list.

Figure legends

- Figure 1.** Schematic diagrams for Tn-seq method used in this study. (A) Single nucleotide was changed in one mosaic end (ME) of EZ:Tn5TM to introduce BsmFI recognition site (5'-GGGAC(N)₁₀↓-3'/3'-CCCTG(N)₁₄↑-5). (B) Deep profiling of Tn5-junction sequences by Illumina sequencing.
- Figure 2.** Identification of the genes conditionally essential for fitness under 3 different selective conditions: dLB, LB-bile, and LB-42°C. Genome-wide view of the fitness index⁻¹ (= total read counts in the input pool/ total read counts in the normalized output pool) is shown for each gene identified under the optimal growth condition (LB medium; control) and the 3 different selective conditions. Fitness index⁻¹ was used for Y-axis (instead of fitness index) for better

visualization of mutants with fitness defect as indicated by the high peaks. The cutoff line of 5 ($=0.2^{-1}$) used in this study to determine fitness defect is shown (dashed lines). The genes selected in this study for further mutant characterization are shown by arrows.

Figure 3. Conditionally essential genes. (A) The genes identified as conditionally essential for fitness under each of the three selective conditions. The numbers of the genes that are also essential for optimal fitness in LB medium are shown in parenthesis. There were 8 additional genes essential for fitness in LB medium, but dispensable for fitness under all of 3 selective conditions. (B) Functional classification of the identified genes.

Figure 4. Competition between the wild type and each of $\Delta pyrD$, $\Delta glnL$, $\Delta recD$, and $\Delta STM14_5307$ mutants. For competition assay, the wild type and each deletion mutant was mixed in 1:1 ratio and inoculated into 2 different culture conditions (control vs. relevant test condition indicated). The cultures were diluted into respective selective conditions every 24 hr for up to 4-8 days. The cell numbers were determined for the wild type (NA^R) and mutant cells (Km^R) each day using LB plates supplemented with appropriate antibiotics. (A) $\Delta pyrD$ (LB vs. dLB), (B) $\Delta glnL$ (LB vs. dLB), (C) $\Delta recD$ (LB at 37°C vs. 42°C), and (D) $\Delta STM14_5307$ (LB at 37°C vs. 42°C).

Figure 5. Comparison of the fitness indices obtained by Tn-seq data (in Table S3) and a competition assay (CA) for $\Delta pyrD$, $\Delta glnL$, $\Delta recD$, and $\Delta STM14_5307$ mutants.

The comparison was made for 2 different culture conditions (control vs. relevant test condition indicated) for each mutant. The results of competition assay (n=3) were obtained at day 4 ($\Delta pyrD$, $\Delta glnL$, and $\Delta recD$) or 7 ($\Delta STM14_5307$).

Figure 6. *In vivo* functions of the gene products during animal infection. Among the genes conditionally required for fitness under each selective condition, the numbers of the genes required for *in vivo* fitness during animal infection (mouse vs. chicken) through 2 different infection routes (systemic vs. enteric infection) are indicated. The numbers of the genes that produce proteins *in vivo* are also indicated for mouse infection model using the 2 different routes of infection. The genes also essential for fitness in LB were not included in the numbers of the conditionally essential genes.

Figure S1. Structure of the 129 bp PCR amplicon containing Tn-5 junction sequence (12bp), ME sequence (19bp) and barcode (6bp).

Figure S2. The percentage of transposon-junction sequences of different lengths uniquely mapping to the genome.

Figure S3. Percentage of the sequence reads for Tn5-junction sequences of different lengths.

Figure S4. Reproducibility of biological replicates. Two biological replicate samples were analyzed by Tn-seq for (A) chromosomal ($R^2 = 0.99$), and (B) plasmid genes

($R^2=0.99$). Axes represent the total number of sequence reads mapped to insertion sites in all genes (5-80% of the coding region).

Figure S5. Growth curves of the wild type, $\Delta pyrD$ and $\Delta glnL$ strains in LB or 10X diluted LB medium (dLB). Growth patterns of the wild type and mutant strains ($\Delta pyrD$, and $\Delta glnL$) were determined during growth in LB (control) and dLB medium. Single colonies from freshly streaked LB agar plates were inoculated into LB broth and incubated overnight at 37°C with vigorous shaking (rpm = 225). One μ l of overnight cultures were transferred into 200 μ l of fresh medium per well in 96-well polystyrene plate and incubated at 37°C for 18 hrs without shaking. The absorbance (OD₆₀₀) was measured using Infinite[®] 200 (TECAN Group Ltd., USA).

Table S1. Oligonucleotides used in this study.

[The image shows a very dark and low-contrast scan of a table containing oligonucleotide sequences. The text is illegible due to the quality of the scan. The table likely contains columns for oligonucleotide names, sequences, and other relevant information.]

Table S2. Summary of Illumina sequencing data analysis.

Strain	Seq. #	No. of reads			% of total reads		% of total reads	
		No. of reads	No. of reads	No. of reads	% of total reads		% of total reads	
					100%	100%	100%	100%
100%	100%	100%	100%	100%	100%	100%	100%	
100%	100%	100%	100%	100%	100%	100%	100%	
100%	100%	100%	100%	100%	100%	100%	100%	
100%	100%	100%	100%	100%	100%	100%	100%	

Figure 1A.

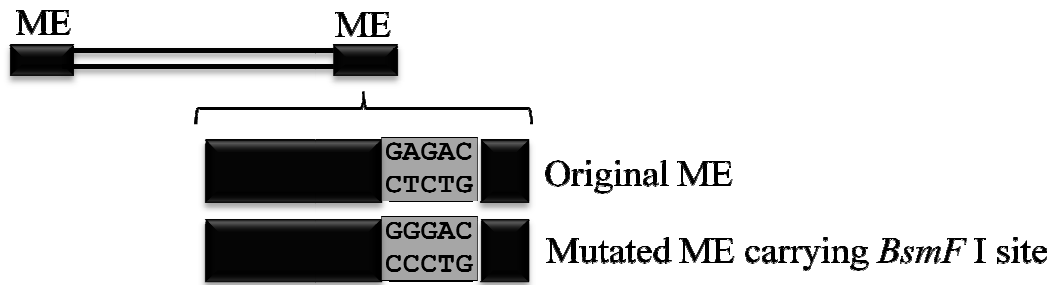


Figure 1B.

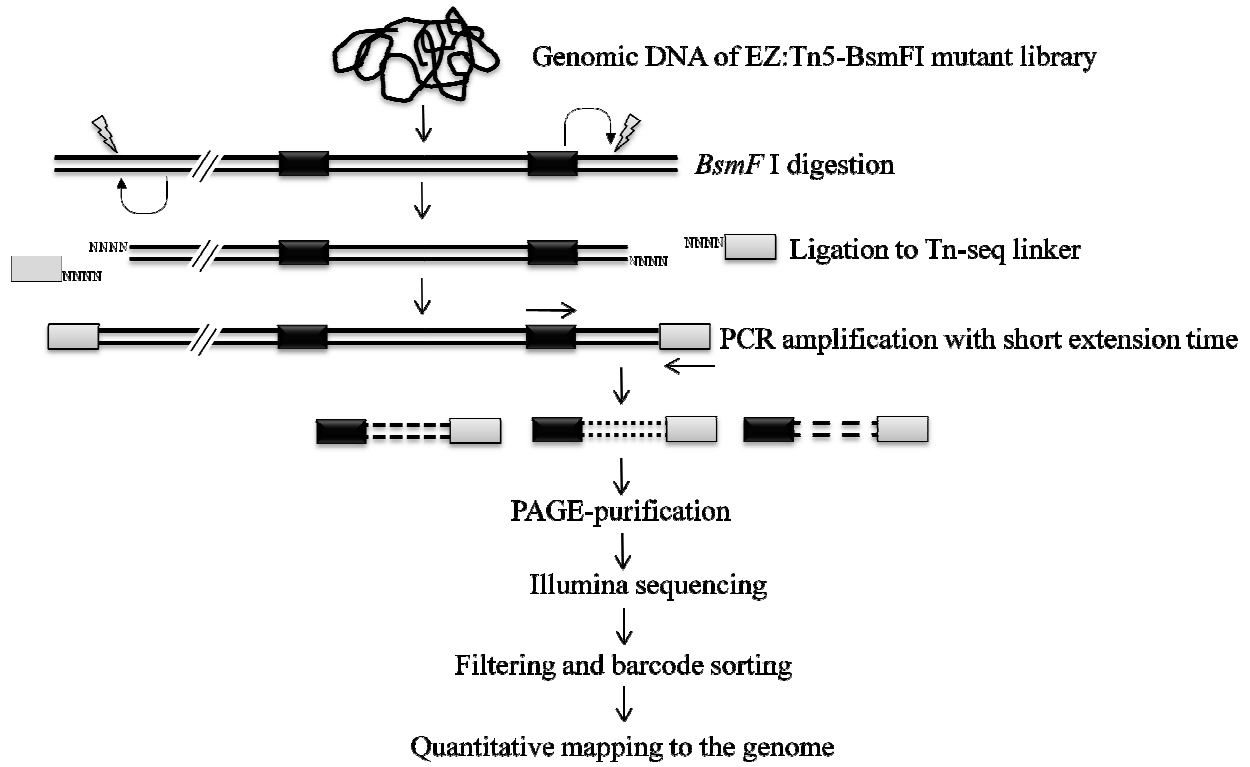


Figure 2.

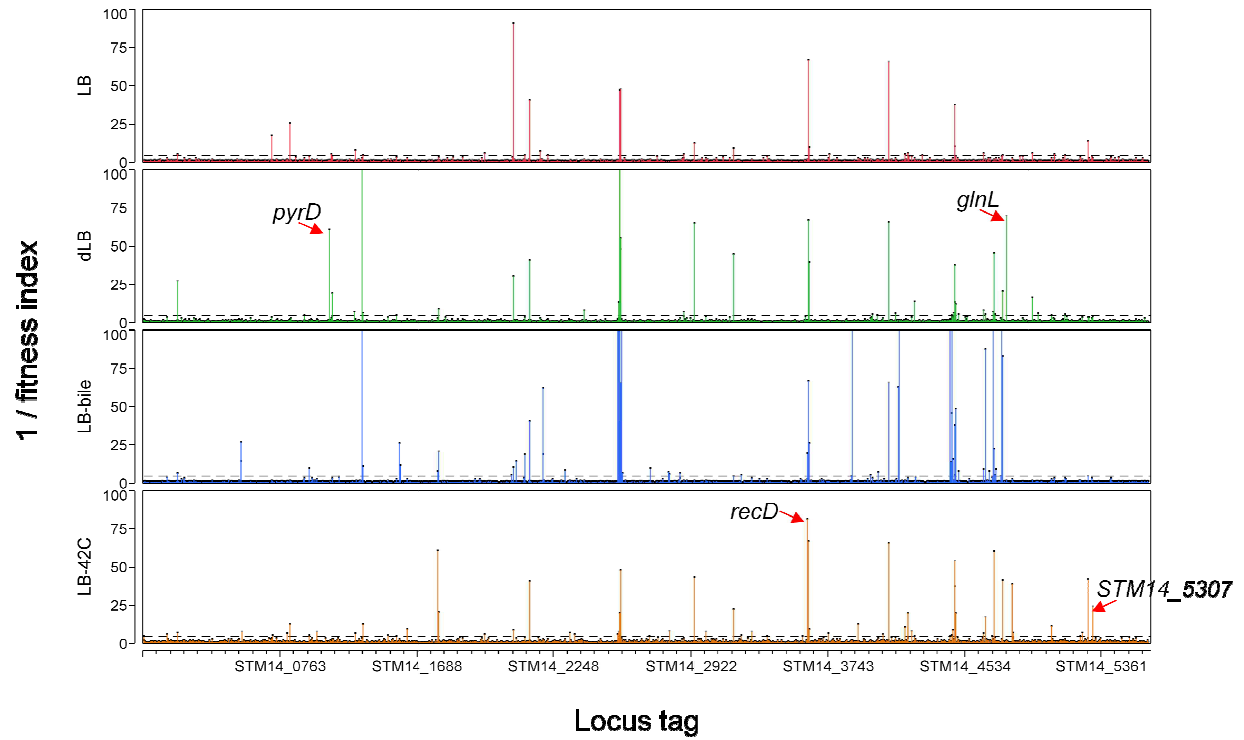


Figure 3A.

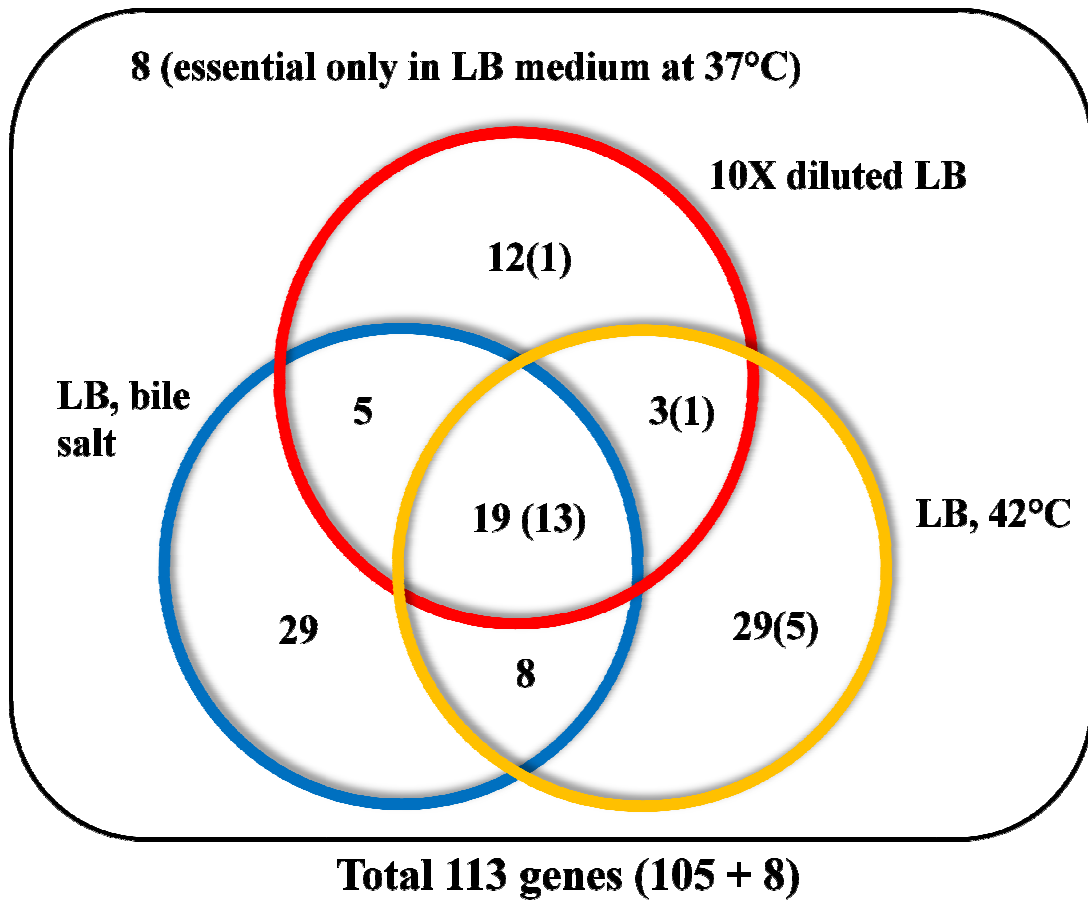


Figure 3B.

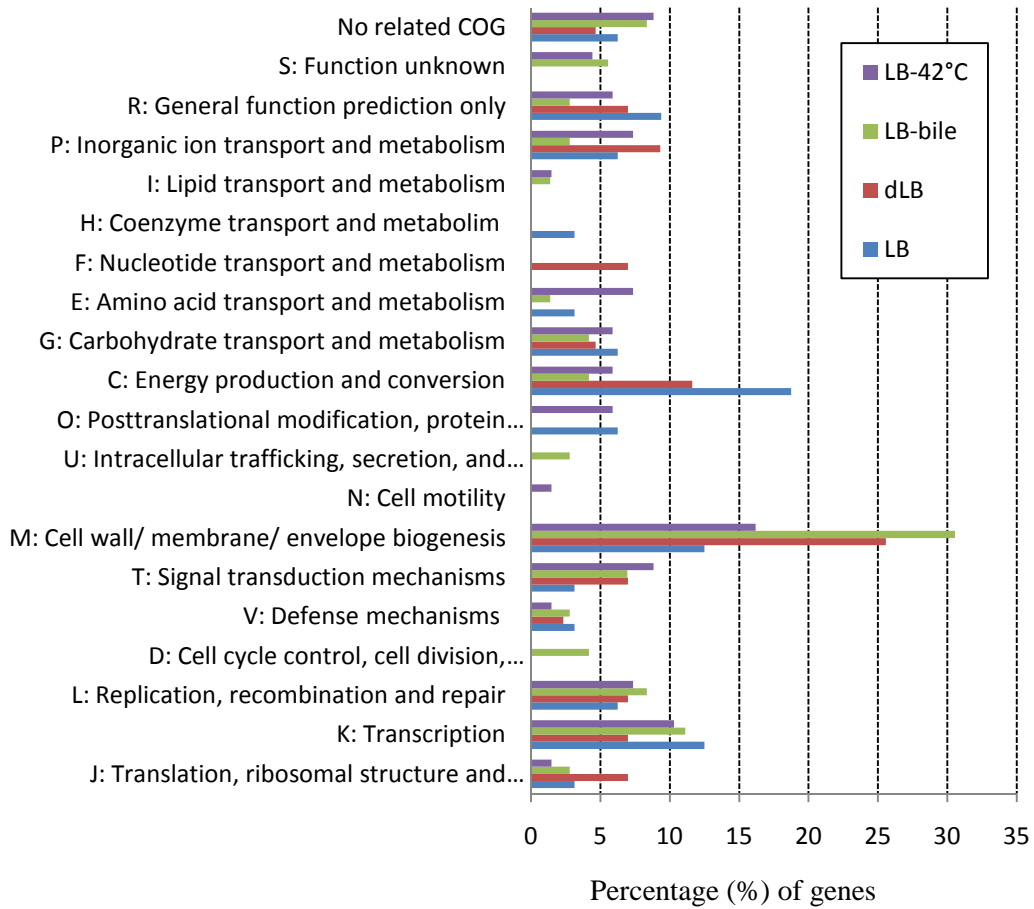


Figure 4A.

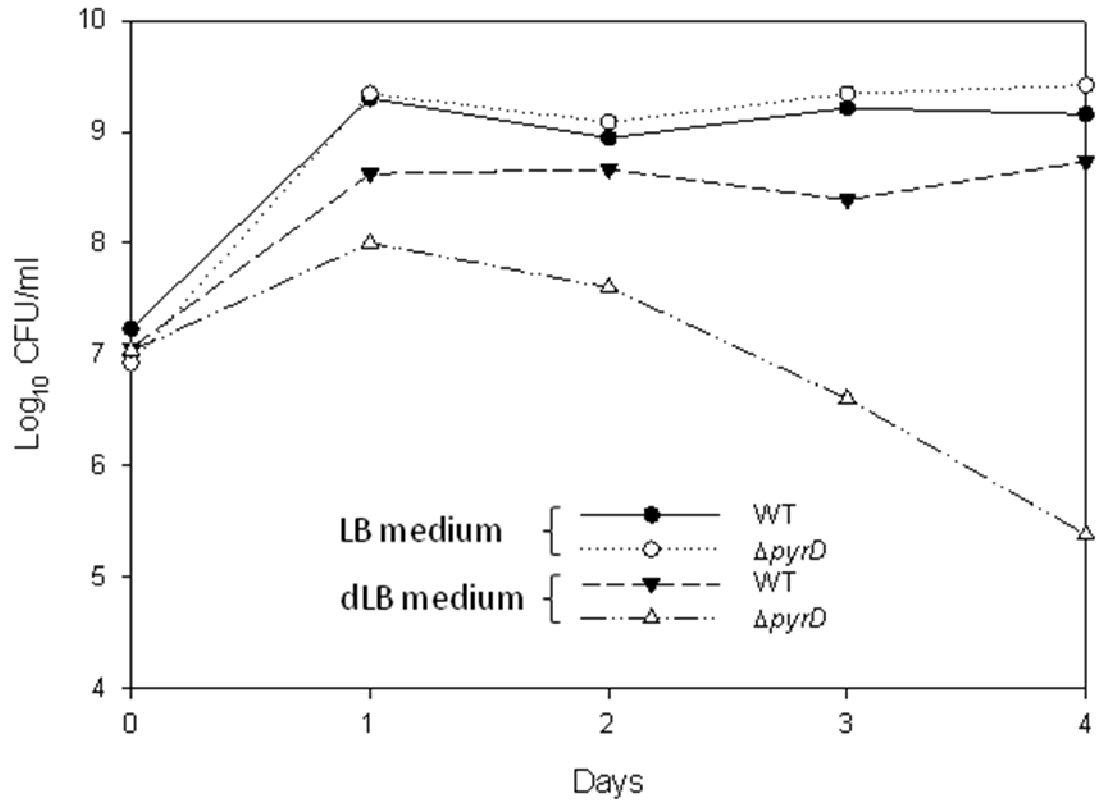


Figure 4B.

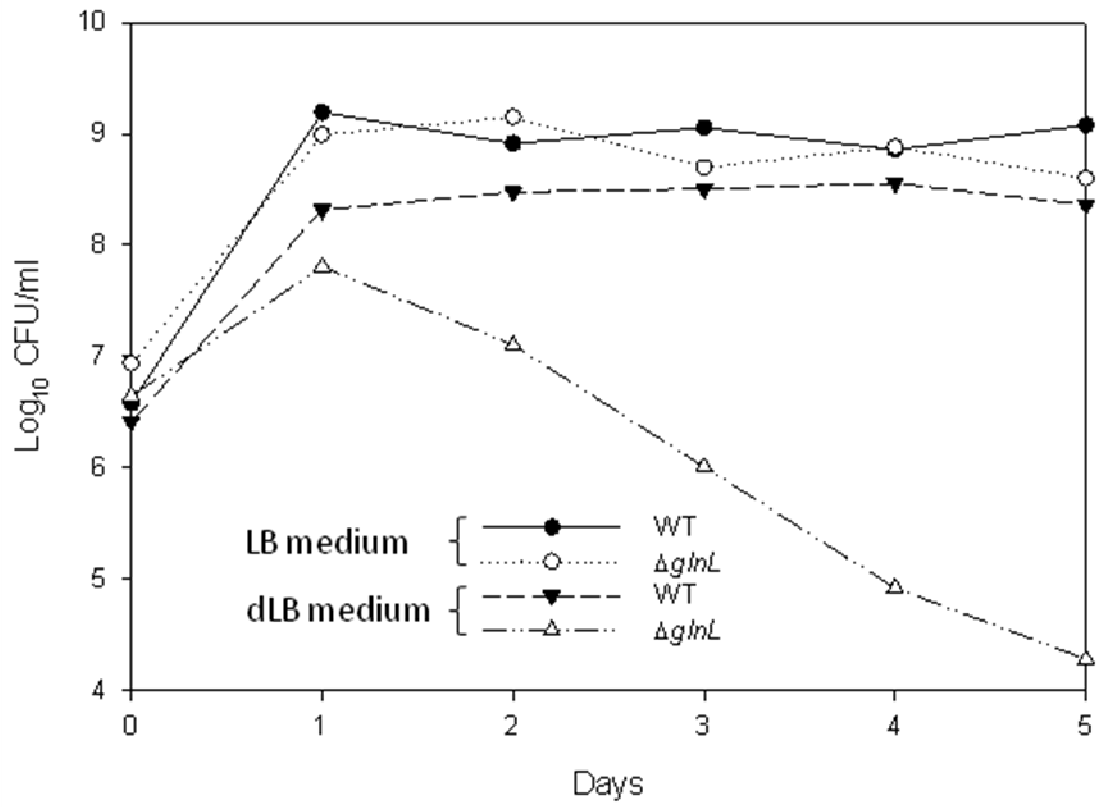


Figure 4C.

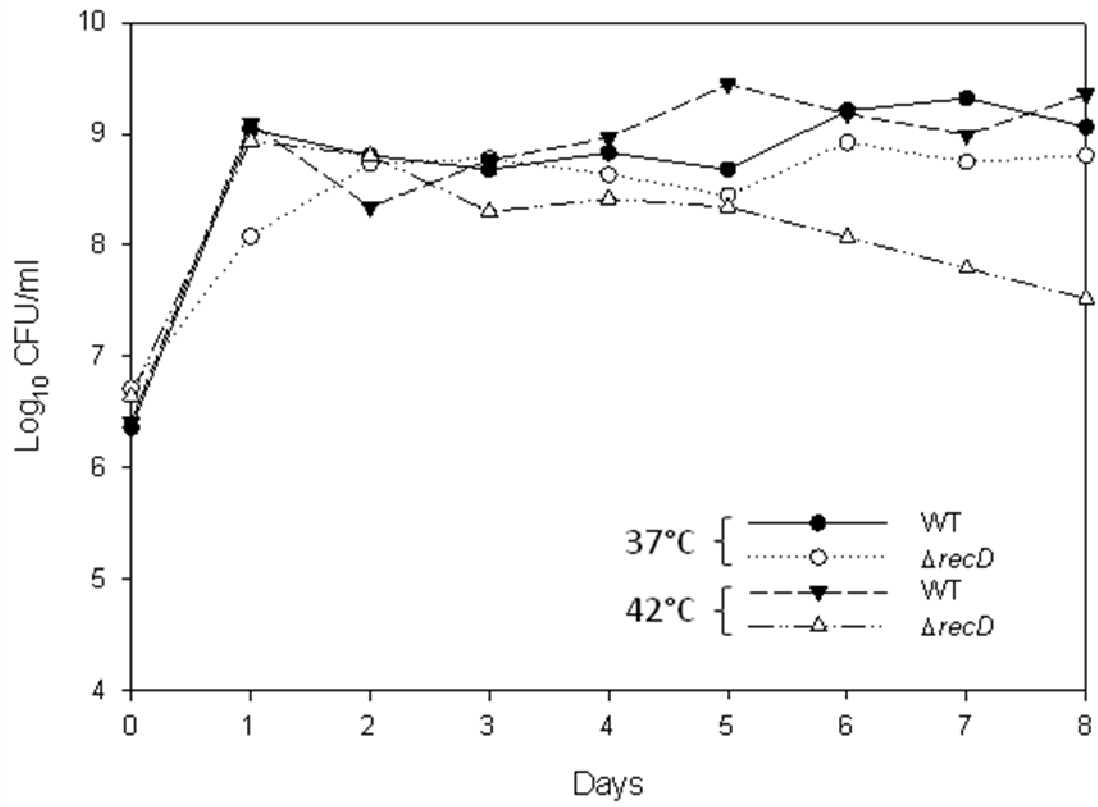


Figure 4D.

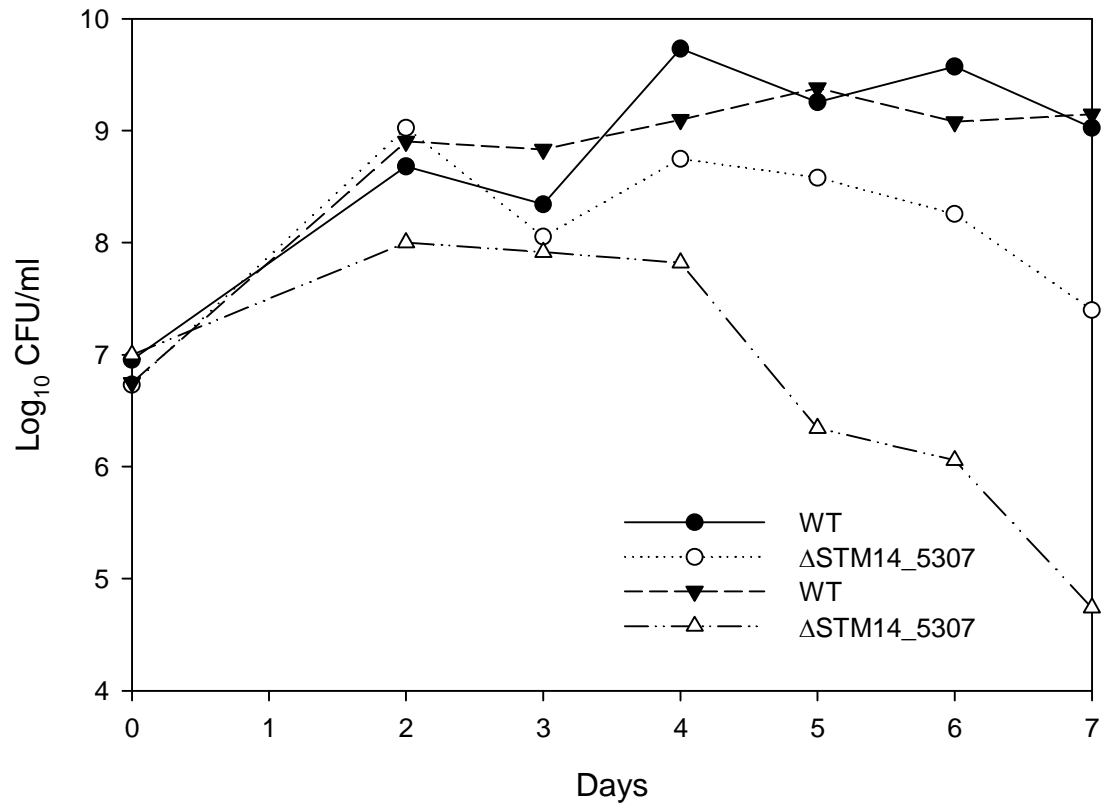


Figure 5.

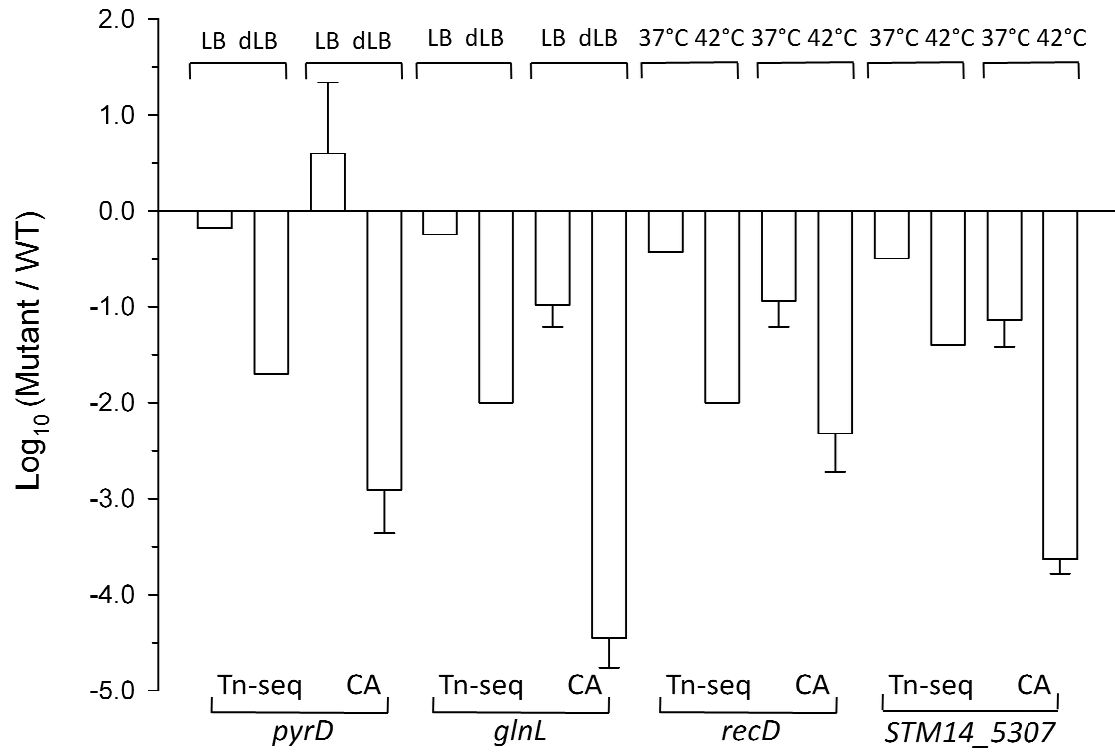


Figure 6.

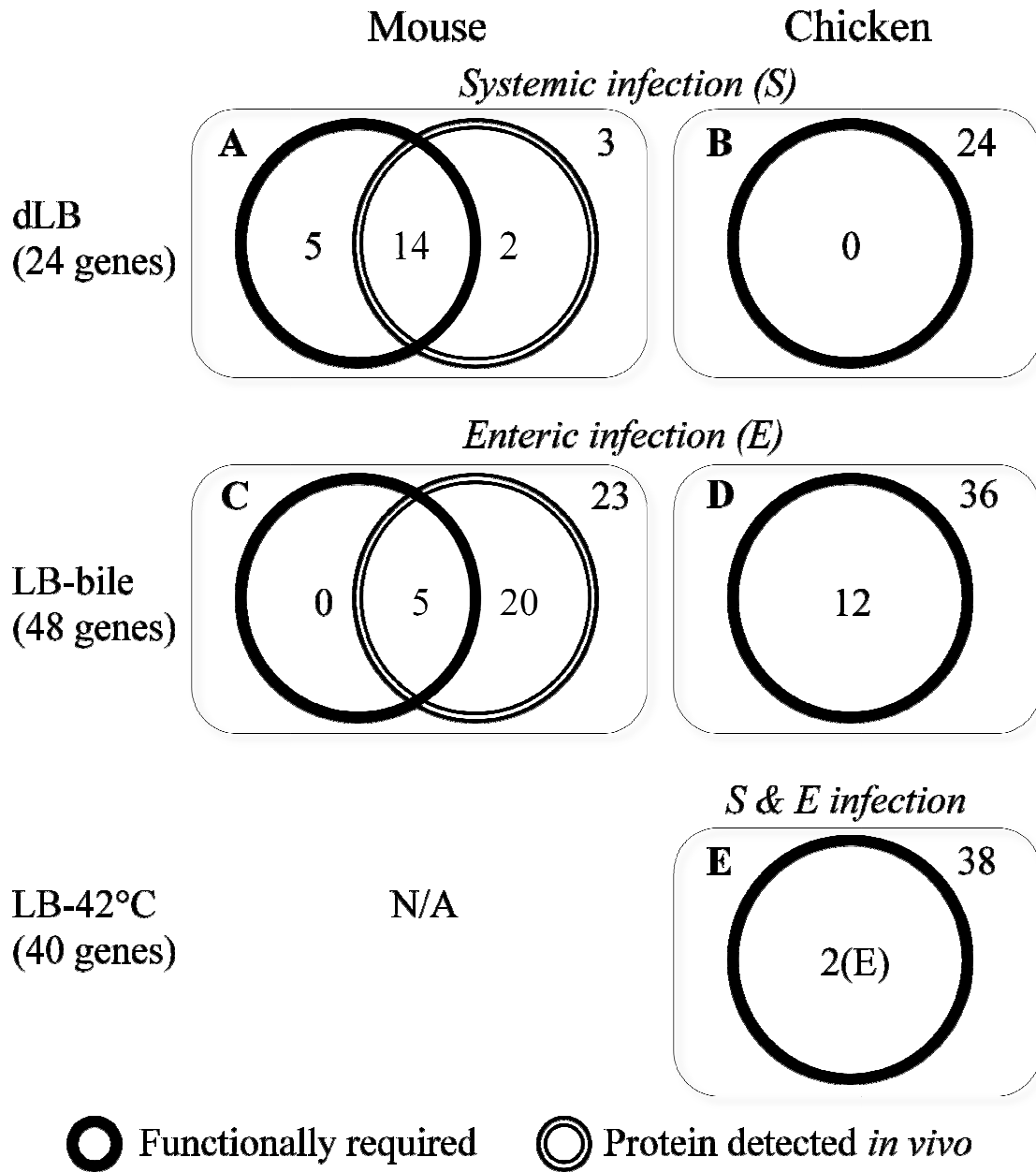


Figure S1.

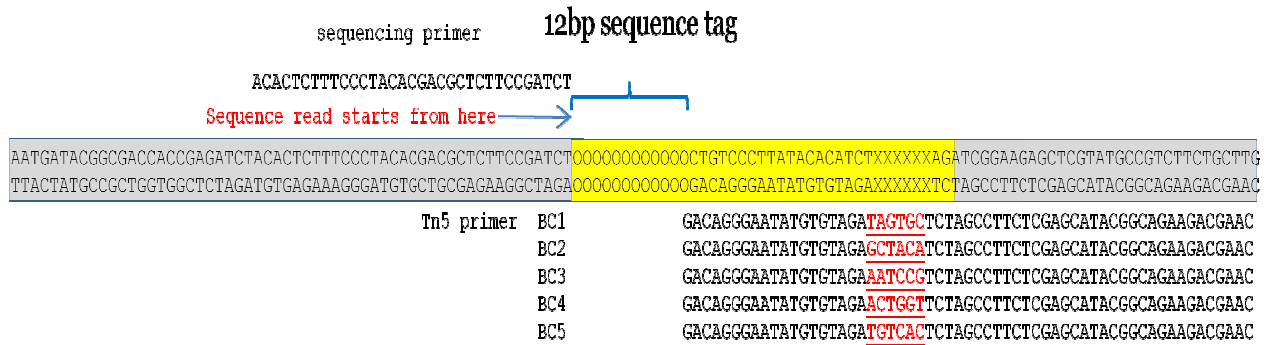


Figure S2.

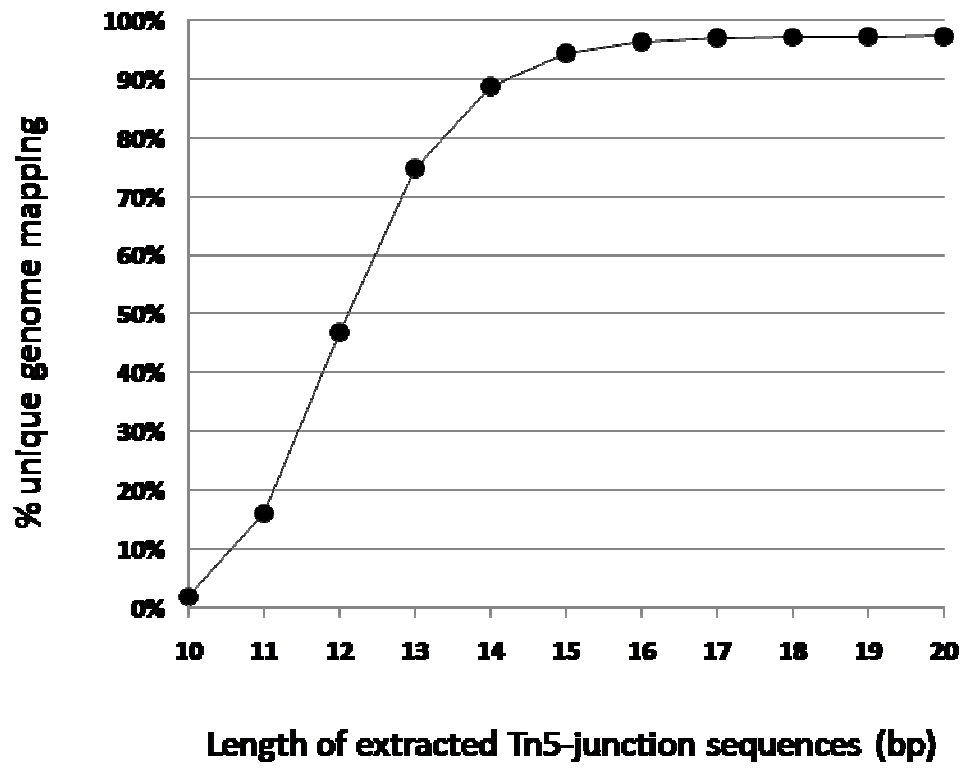


Figure S3.

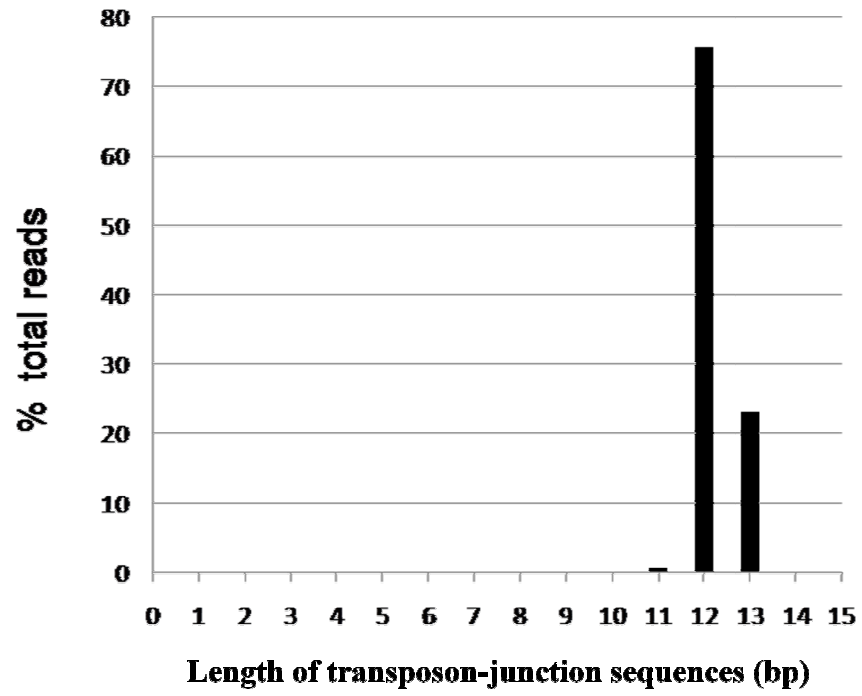
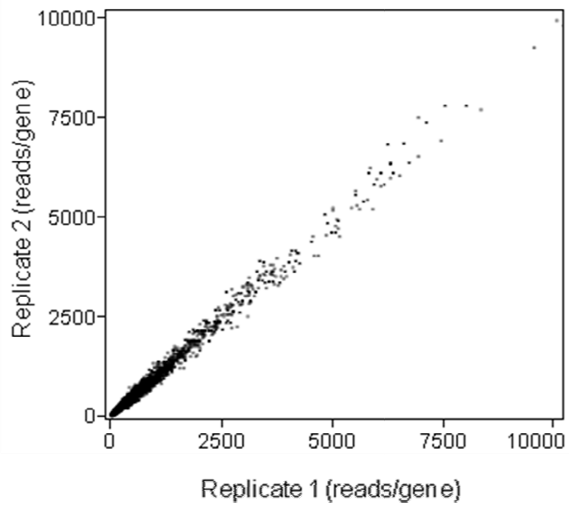


Figure S4.

A



B

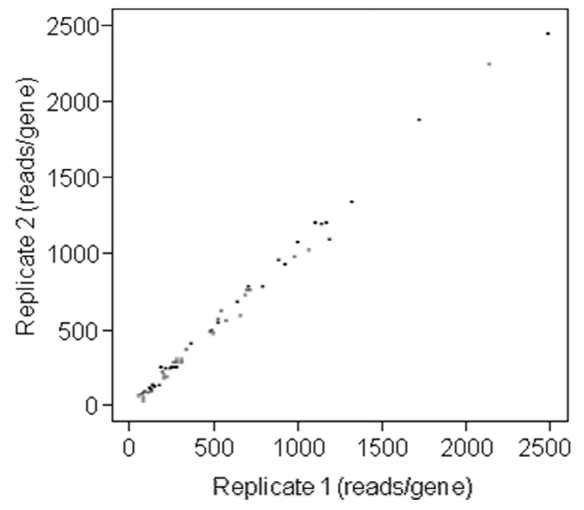
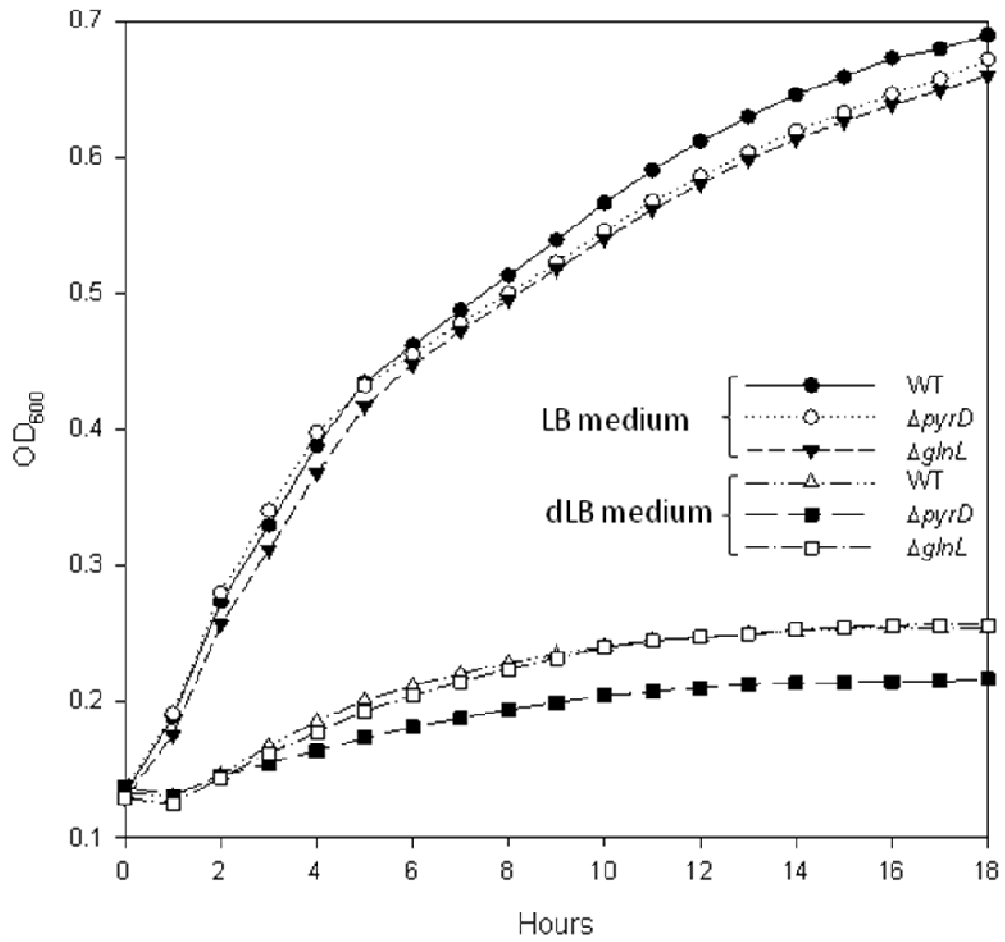


Figure S4.



CHAPTER III

IDENTIFICATION OF *SALMONELLA* TYPHIMURIUM GENES CONDITIONALLY ESSENTIAL FOR SURVIVAL DURING COLD TEMPERATURE STORAGE USING Tn-Seq METHOD

Identification of *Salmonella* Typhimurium Genes Conditionally Essential for Survival During Cold Temperature Storage using Tn-seq Method

ABSTRACT

Non-typhoidal *Salmonellae* are highly prevalent food borne pathogens of worldwide economic and public health importance. Contaminated foods including meat, meat products, eggs, fruits, vegetables are responsible for spread of organisms to humans. The challenges to our ability to reduce incidence of food-borne illness arise from incomplete scientific knowledge of the pathogens. Thus, the knowledge of the bacterial factors necessary for *Salmonella* to persist sub-clinically inside livestock and survive and grow in other reservoirs such as crops and processed foods will allow the development of new strategies to improve food safety. In this study, we applied Tn-seq, a powerful functional genomics tool to dissect the entire *Salmonella* genome, to identify the genes conditionally essential under food storage conditions, namely 4°C refrigeration and -20°C freezing temperatures. We have previously demonstrated the utility and efficiency of this Tn-seq method in comprehensive identification of conditionally essential genes in *Salmonella*. In total Forty two overlapping genes exhibited strong fitness defect under selective conditions and thus were identified as conditionally essential. Thirty seven genes were identified at 4°C, 41 genes were identified at -20°C and 36 of the genes identified were common in both selective conditions. Four genes namely *trkA*, *sapD*, *rfbF* and *rfbI* were picked for further phenotypic characterization. Deletion mutants were subjected to individual survival assays and competition assays. Surprisingly, none of the mutants showed significant difference in survival at cold storage temperature as compared to wild type. This study of *Salmonella* genome to identify the genes conditionally essential under food storage conditions, namely 4°C refrigeration and -20°C freezing temperatures, is first of its kind attempted. The outcome of this study is a

challenge to us to design experiments to study the intricacies of this food borne pathogen further, until we find a reasonable explanation.

INTRODUCTION

Salmonella is a one of the most common food-borne bacterial pathogen affecting about 1.4 million people annually in the U.S, causing 20,000 hospitalizations and 400 deaths resulting in total cost of estimated 3 billion dollars annually in the U.S (Economic Research service 2011; Scallan *et al.*, 2011). Thirty five percent of all food-borne hospitalizations and 28% of food-borne disease deaths are caused by *Salmonella* (Scallan *et al.*, 2011). FoodNet surveillance (1996-1998) compared with 2010 surveillance does not show any change in the incidence of *Salmonella* infection (CDC factsheet, 2010). Among more than 2,500 *Salmonella enterica* serotypes, serotype Typhimurium is the most common one causing human Salmonellosis in the U.S. There are numerous sources of *Salmonella* contamination in human food chain but poultry products still remain the major source of human non-typhoidal salmonellosis, accounting for up to 50% of total incidence in the U.S (CDC, 1996). In recent years the non-typhoidal *Salmonella* infection has been attributed to fruits, vegetables and processed foods such as peanut butter (www.CDC.gov). *Salmonella sp.* can colonize wide range of hosts such as poultry, cattle, pigs, and survive under food processing stresses and eventually lead to production of contaminated meat and other food products. These diverse habitat provide *Salmonella* with an opportunity to adapt and evolve, which is demonstrated by changing trends in salmonellosis (Newell *et al.*, 2010) and exemplified by isolation of *Salmonella* from unconventional food sources like fresh produce (Franz and van Bruggen, 2008). This pathogen continues to remain a great burden for both poultry industry as well as human public safety.

Various physical and chemical methods are used in the food processing industry to reduce or prevent the introduction and survival of microorganisms in food. Physical methods of preservation can include dehydration, freeze-drying, heat treatment, or irradiation. Chemical techniques include use of antimicrobial agents, organic acids, and salt which are added to foods to reduce the levels of contamination. All of these preventive methods work by exerting various forms of stress on the bacterial cell, leading to growth inhibition or death. However, the microorganisms have remarkable adaptive mechanisms that enable them to respond to environmental stresses, such as wide fluctuations in temperature, salt, and sugar concentration, or pH. Therefore, it is not surprising that microorganisms have developed resistance to a number of preservative agents used in the food processing industry (Woteki and Kineman, 2003). In poultry industry, methods that reduce bacterial pathogens in the post-harvest stages of poultry production are based on the general bacteriostatic or bactericidal treatments of poultry products and processing environments. Although effective in reducing bacterial loads in contaminated poultry products, it is obvious that more effective control measures are in great demand to ensure supply of safer poultry products.

To develop more effective strategies, it is important to target specific pathways in *Salmonella* that are essential for persistence in poultry processing, handling and storage environments. The knowledge of genetic factors in *Salmonella* that are essential for counteracting stresses at different stages of poultry production or processing would make it possible to develop control measures targeting those particular pathways. The development and application of novel functional genomics tools have allowed rapid progress in elucidating genetic factors and pathways in various bacterial pathogens that are conditionally essential under various selective conditions. These pathways can be explored as promising targets for development of

novel antimicrobials. In the case of food-borne bacterial pathogens, our understanding of the molecular mechanisms underlying a pathogen's capability to survive and persist in the stressful environments during pre- and post-harvest food processing and distribution would be crucial for development of effective measures to reduce the pathogens in final food products (Kwon *et al.*, 2009). However, the genetic factors of food-borne bacterial pathogens essential for persistence in food environments have been largely unexplored. It is expected that functional genomics approaches that have been extensively used to study bacterial pathogens in the context of host-pathogen interactions would be a powerful tool to increase our understanding in this area of study. Particularly, functional analysis tools that can assess genetic requirements using a complex transposon mutant library and next-generation sequencing (NGS) technologies will be readily available to identify genetic factors of *S. Typhimurium* essential for survival and persistence in association with chicken carcasses during poultry production, processing and distribution (van Opijnen *et al.*, 2009, Khatiwara *et al.*, 2012). The resulting information would be crucial in developing more effective strategies to control Salmonella contaminating poultry products.

Low-temperature storage of poultry and poultry meat products is an important method of controlling microbial growth by slowing the catalytic ability of microbial enzymes resulting in deceleration of metabolic processes. The bacterial cells exposed to low temperature undergo physiological changes such as decrease in membrane fluidity, stabilization of secondary structures of nucleic acids leading to reduced transcription and translation efficiency; inefficient folding of proteins and hampered ribosome function as documented by Phadtare (2004). However, low temperature treatment induces synthesis of number of proteins called cold shock proteins to counteract the harmful effect of temperature change; eventually the bacterial cells

become acclimated to low temperature (Jones *et al.*, 1987). The study and understanding of cold shock response of food-borne pathogens such as *Salmonella* is imperative as refrigeration and freezing are commonly used for food storage.

Recently, a powerful functional genomics tool has been established to dissect the entire *Salmonella* genome for the genes conditionally essential under the selection condition of interest. We have applied this method, Tn-seq, successfully to conduct a genome-wide screening for the genes essential for optimal growth under various stress conditions associated with chicken infection including low nutrient condition, bile salt-rich condition, and at 42°C temperature (Khatiwara *et al.*, 2012). In this study the same Tn-seq method is applied to study conditionally essential *Salmonella* genes as an important step toward comprehensive and deeper understanding of the genetic mechanisms used by *Salmonella* to persist in storage conditions (4°C for refrigeration and -20°C for freezing). Although numerous studies have been conducted to reduce *Salmonella* in poultry products, little is known regarding the genetic mechanisms utilized by *Salmonella* to persist in the environments against various stress conditions and antibacterial treatments. This approach is analogous to the recent strategies to develop novel antibiotics, in which efforts are focused on targeting specific pathways for efficient control while minimizing the probability of development of antibiotic resistance (Clatworthy *et al.*, 2007). A better understanding of the *Salmonella* genes that are required for growth and survival in foods, and surrounding environments will help to focus on future food safety initiatives like development of antimicrobials and to introduce latest science-based approaches to inspection using hazard analysis and critical control points (HACCP) programs.

MATERIALS AND METHODS

Bacterial strains and culture conditions

S. Typhimurium 14028 wild type strain or a spontaneous mutant resistant to Nalidixic Acid (NA) were grown in Luria Bertani (LB) media or LB agar plates and stored at -80°C in 30% glycerol. The cultures were incubated at 37°C unless described otherwise. Where appropriate, the LB agar plates contained NA (25µg/ml) or Kanamycin (50µg/ml).

Transposon mutant library construction

Site-directed mutagenesis was performed to change one nucleotide of pMODTM-6 <KAN-2/ MCS> plasmid (Epicentre, Madison, WI) corresponding to one mosaic end (ME) sequence to introduce the recognition sequence of TypeIIIS enzyme BsmFI using QuikChange Site-directed mutagenesis kit (Agilent Technologies La Jolla, CA) and oligonucleotides in Table 1. The nucleotide change was confirmed by DNA sequencing, and the modified plasmid, pMOD-BsmFI, was used for transposon mutagenesis of *S. Typhimurium* ATCC 14028 wild type strain. Briefly, pMOD-BsmFI was digested with Pvu II enzyme and the 1,221bp fragment of the modified EZ:Tn5TM, EZ:Tn5-BsmFI, was obtained by agarose gel-purification without exposure to UV light. The EZ:Tn5-BsmFI was then incubated with EZ-Tn5TM Transposase (Epicentre BioTechnologies, Madison, WI, USA) to form a transposon complex according to the instruction manual. Two µl of the complex was then used to transform electrocompetent *S. Typhimurium* 14028 wild type cells by electroporation. Km-resistant transformants were selected on LB plates supplemented with Km. The resulting mutants were combined to form a complex library of Tn5 mutants containing approximately 1.6 x 10⁴ mutants. The library was stored at -80°C in 30% glycerol.

In vitro selection of transposon mutant library

The mutant library was subjected to selection under 2 different in vitro conditions; 4°C and -20° C. Approximately 1.6×10^4 Salmonella Typhimurium14028 Tn5 mutants were grown at 37 °C for an hour with vigorous shaking at 225rpm. The culture was then washed three times with 1X phosphate buffered saline (PBS) and OD was adjusted to ≤ 1.0 , corresponding to 7.6×10^8 colony-forming units (CFU) per ml. The culture was dispensed into five tubes with 1ml in each, tube 1 for Input, tube 2&3 for 4° C outputs, and tube 4&5 for -20° C outputs. Genomic DNA was extracted from input sample (tube 1) for Illumina DNA fragment library preparation using QIAamp DNA mini kit (Qiagen, USA). Each pair of output samples were incubated either at 4° C or -20° C for 17 days. After 17days incubation four cultures were recovered on LB agar plates supplemented with kanamycin, CFUs were determined. Genomic DNA was extracted from the harvested output samples.

Sample preparation and Illumina sequencing

Genomic DNA extracted from one input pool and four output pools (Figure.1a) was digested with BsmFI restriction enzyme (New England BioLabs, Ipswich, MA) at 65°C for 3 hrs. After heat inactivation at 80°C for 20 min, the digested DNA was treated with calf intestinal alkaline phosphatase (NEB) at 37°C for 1hr to prevent self-ligation in the following ligation step by incubation for additional 1 hr. DNA was then phenol-chloroform extracted, ethanol precipitated and dissolved in 10 µl H₂O. DNA digests were subsequently ligated to Tn-seq linker, formed by annealing Tn-seq linker 1 and 2 oligonucleotides (Table 1), by overnight incubation at room temperature. The Tn-seq linker-ligated samples were subsequently used as templates in PCR reaction using cloned Pfu DNA polymerase (Agilent Technologies) with one

of the 5 barcoded Tn5 primers and Tn-seq linker primer (Table 1). The PCR cycles consisted of initial denaturation at 94°C for 2 min, 5 cycles of 94°C for 30 sec, 55°C for 1 min, and 72°C for 30 sec, 20 cycles of 94°C for 30 sec, 62°C for 1 min, and 72°C for 30 sec followed by final extension at 72°C for 10 min. PCR amplicons of 129 bp were visualized (Figure. 1b), PAGE-purified and dissolved in H₂O. Five DNA samples tagged with different barcodes were pooled at the same amount based on measurement with a NanoDrop apparatus (Thermo Fisher Scientific). The pooled sample was visualized (Figure 2.), PAGE-purified and finally analyzed using Affimetrix Bio Analyzer to check the quality. Sequencing was conducted at the Institute for Integrative Genome Biology in the University of California at Riverside using Illumina Genome Analyzer II.

Data analysis

Python programming language was used for initial data analysis. The sequence reads obtained from Illumina sequencing was sorted for the reads that contain a perfect 19 bp modified ME sequence (5'-CTGTCCCTTATACACATCT-3'). The resulting filtered sequences were sorted according to 5 bp barcode sequence demanding perfect match to one of the 5 barcodes. Then, transposon-junction sequences were extracted from the filtered reads and the junction sequences of 11-13 bp long were selected for further analysis. These selected transposon-junction sequences were then mapped to the complete genome of *S. Typhimurium* 14028 (Jarvik *et al.*, 2010) (Accession Number: Chromosome CP001363.1, Plasmid CP001362.1) to select the reads that perfectly match to the genome sequence. Additional filtering was performed to select the reads that map to the genome only in one locus. The output data obtained by the Python script contained the information on the transposon-junction sequences, origin (chromosome vs. plasmid), genomic coordinate corresponding to Tn5 insertion site, strand (+ vs. – strand), and the

number of the reads in each pool for 5 mutant pools (1 input pool and 4 output pools). The above data for chromosome and plasmid were processed separately using JMP8 software (SAS, Cary, North Carolina) from this step. For additional filtering, insertions with read counts in the input pool <10 were eliminated to remove background nonspecific reads. For normalization, normalization factor was calculated according to the formula $(R_i/R_o)/(S_i/S_o)$, in which the variables represent the total number of sequence reads (R) and insertion sites (S) detected in the input (i) and output(o) pools, respectively, for each output pool (Gawronski *et al.*, 2009). The number of sequence reads for each insertion in each output pool was multiplied by the corresponding normalization factor. After normalization, the insertions in a gene with <3 insertions within the internal 5 - 80% of the coding region were removed. Then the numbers of all sequence reads within each gene were combined for each pool to obtain total number of normalized sequence reads originated from each gene. Then the fitness value was calculated for each gene by dividing the total number of sequence reads in the input pool by those in the output pool.

Statistical Analysis

The statistical analysis of chromosome and plasmid data was carried out separately using JMP8 software (SAS, Cary, North Carolina). Fitness averages for 4°C and -20° C were calculated and the data were subjected to t-test and P value was calculated. Genes with P value <0.01 and fitness value ≤ 0.33 representing three fold reduction in relative abundance during selection were identified as conditionally essential genes under each selective condition.

Construction of deletion mutants

Oligonucleotides described in (Table1) were designed to produce single deletion mutants of *S. Typhimurium* 14028 harboring a mutation in *trkA*, *sapD*, *rfbF* and *rfbI*. Mutants were constructed using Lamda Red recombination system by the method described by Cox et al. (2007). Briefly, a PCR amplicon containing the pKD13 (Datsenko *et al.*, 2000) kanamycin resistant gene cassette flanked by approximately 200-300 bp of DNA sequences homologous to the up and downstream regions of the target gene was obtained by overlapping extension PCR using Ex Taq TaKaRa polymerase (Takara Bio, Inc). The overlapping extension PCR cycle consisted of initial denaturation at 94°C for 2 min, 29 cycles of 94°C for 30 sec, 58°C for 1 min, and 72°C for 3min followed by final extension at 72°C for 10min. The amplicon was then used to transform *S. Typhimurium* containing plasmid pKD46 to introduce KmR cassette into target gene. The cells were selected on LB plate supplemented with Km. The *trkA* and *sapD* deletions were transferred to the fresh wild type background by P22 transduction and finally selected on LB agar plates supplemented with Km. The *rfbF* and *rfbI* failed repeatedly to produce phage lysate so these mutation could not be transferred to the fresh background. The sequencing result confirmed the successful construction of all four mutants.

Survival assays

Growth patterns of the wild type and mutant strains ($\Delta trkA$, $\Delta sapD$, $\Delta rfbF$ and $\Delta rfbI$) were determined during growth in LB at 4° C and -20°C. Single colonies from freshly streaked LB agar plates were inoculated into three tubes with 2ml LB broth in each tube and incubated 16h at 37°C with vigorous shaking (rpm = 225). Contents from the tubes were combined and washed three times with 1X PBS and finally re-suspended in 6ml LB broth. Then 150uL of the

contents were dispensed into forty micro-centrifuge tubes and stored at 4°C and -20 °C. The content was also used to determine 0 day cell count by serial dilution and plating. Three tubes (replicates) were pulled from 4°C and -20°C on 3, 6, 9, 12, 15 days and serially diluted and plated to determine viable cell count. The data obtained were processed to calculate percentage survivability mutants/ wild type.

Competition Assay

All mutant strains were exposed to 42° C to remove plasmid pKD46 if any. Single colonies from freshly streaked LB agar plates were inoculated into a tubes with 4ml LB broth and incubated 16h at 37°C with vigorous shaking (rpm = 225). After 16h incubation, cultures were washed 3X with 1X PBS and re-suspended in LB broth. Wild type and mutant were mixed at 1:1000 ratios to simulate mutant library selection ratio and 200uL of the mixture was dispensed in micro-centrifuge tubes and stored at 4°C and -20°C. Mixture was also serially diluted and plated to determine 0 day count. After 17day cultures were pulled out of 4°C and -20°C and serially diluted and plated on Kan and Na plates. Next day the colonies were counted and CFUs/ml was determined.

RESULTS

Methodology

We engineered Tn5 derivative transposon (EZ:Tn5TM –BsmFI) that carries modified inverted repeat (IR) sequences that is recognized by restriction enzyme BsmFI. BsmFI cuts 14bp away from the recognition site and this property was used to extract 12bp sequences en mass immediately adjacent to Tn5 insertion site to identify and quantify each insertion in the library. EZ:Tn5-BsmFI is a powerful tool for deep profiling of complex insertion mutant library

via Illumina sequencing as shown by earlier study. (Detailed information is available in the previous chapter)

Selection of the mutant library

Tn-seq method was used for genome-wide identification of *S. Typhimurium* genes conditionally essential for fitness under 2 different in vitro conditions. Two selective conditions in this study were chosen to mimic storage and handling temperatures namely 4°C refrigerating and -20°C freezing environment (LB media), The EZ:Tn5-BsmFI mutant library consists of approximately 1.6×10^4 different mutants, and the inoculum of 1 ml contained approximately 7.3×10^7 cells, indicating each mutant in the library was represented by approximately 4,600 cells. Mutants were incubated at either 4°C or -20°C for 17 days. Following the 17 day incubation, the cultures were recovered on LB agar plates supplemented with 50ug/ml of kanamycin, and the number of CFUs was determined by counting viable cells (colonies) on the selective plates.

Analysis of Illumina sequencing data

The summary of the Illumina sequencing data and its analysis is shown in Table 2. Among the total of 35,688,012 sequence reads obtained from a single flow cell lane, 95% (33,988,628 reads) contained the complete 19 bp ME sequence. Among these reads, 33,179,365 reads (98%) contained the complete 6bp barcode sequences perfectly matching one of the 5 barcodes. When these reads were sorted according to the barcode, we obtained relatively even distribution across different barcodes: 5,442,383 read (ATCACG; Input): 6,154,898 (CGATGT; 4°C-1): 7,397,692 (TTAGGC; 4°C-2): 7,415,434 (TGACCA; -20°C-1): and 6,768,958 (ACAGTG; LB-20°C-2). The transposon-junction sequences were subsequently extracted from the reads for each barcode. As expected the majority (>99%) of the sequence reads were 11~13

bp long. These 11~13 bp sequence reads for each barcode was further filtered for those that map to the genome at only one genomic loci. Finally, we obtained 5,412,187; 6,125,632; 7,362,089; 7,389,502 and 6,733,497 reads for each barcode, which corresponds to ~99% of the total number of the transposon-junction sequences of 11~13bp for each barcode. The normalization factors for the output pools were 0.97 (4°C-1), 0.81 (4°C-2), 0.82(-20°C-1), and 0.90 (-20°C-2) for chromosome and 0.28 (4°C-1), 0.26 (4°C-2), 0.23 (-20°C-1), and 0.26 (-20°C-2) for plasmid. When the reads for all insertions in each gene were combined, we obtained the data set for 4,113 and 91 coding genes for chromosome and plasmid, respectively. To obtain a more robust and reliable result, the genes that contained less than 3 insertions were removed, resulting in a total of 2,350 and 59 coding genes for chromosome and plasmid, respectively. We also determined the reproducibility of Tn-seq profiling using two biological replicates. When the data was processed separately for chromosome and plasmid, we obtained very high levels of reproducibility for both chromosome ($R^2=0.99$) and plasmid ($R^2=0.99$) (Figure 3).

Identification of conditionally essential genes

Statistical analysis was conducted and among the 2,409 (2350 + 59) genes (figure 4), the genes conditionally essential for each selective condition were first selected by a cut-off P value of <0.01 resulting in 200 conditionally essential genes. Then, the genes with fitness index of ≤ 0.33 , which indicates at least a 3-fold reduction in relative abundance during selection, were identified as conditionally essential genes under selective conditions. After this selection, 42 conditionally essential genes were identified (figure 5, table2); the genes selected by this criterion are expected to exhibit strong fitness defect under each condition. We identified 37 and 41 genes required for fitness at 4°C and -20°C, respectively (Table 3). We identified 42 genes in total and 36 of them were common in both conditions. The five genes that were found to be

conditionally essential uniquely at -20°C were *pheT*, *selC*, *tatC*, *fdoG* and *moaC*. The only conditionally essential gene identified uniquely in -20C was *yffG*. All of the 42 genes identified were chromosomal genes and none of them were located on plasmid.

To gain insights on functional trends associated with each selective condition, the identified genes were assigned to functional (COG; Cluster of Orthologous Groups) categories (Tatusov *et al.*, 1997) using BLAST on Orthologous Groups (BLASTO) algorithm (Zhou *et al.*, 2007) shown in Figure 6. Approximately, 15 to 20% of all the genes identified were enriched in COG category M which represents the genes responsible for cell wall, membrane and envelope biogenesis. Almost 10- to- 12 % of the genes identified were enriched in COG category C representing gene playing functional roles in energy production and conversion. Most of the genes identified i.e., 30-to-32% had no related COG category.

Mutant Construction

The basic strategy to replace target gene with an antibiotic gene was accomplished by Red- recombination protocol. Kanamycin resistant gene fragment of 1,320bp was obtained from pKD13 for each gene. Then gene specific upstream fragment were obtained successfully for *trkA* (292bp), *sapD* (329bp), *rfbF* (265bp) and *rfbI* (270bp) genes. Gene specific downstream PCR products were obtained successfully for *trkA* (297bp), *sapD* (311bp), *rfbF* (261bp) and *rfbI* (306bp). Kanamycin resistant gene fragment, upstream and downstream fragments were used as template for overlapping extension PCR to successfully obtain a PCR amplicon containing the pKD13 kanamycin resistant gene cassette flanked by approximately 200-300 bp of DNA sequences homologous to the up and downstream regions of the target gene. The PCR products were of expected size, namely *trkA* (1909bp), *sapD* (1960bp), *rfbF* (1846bp) and *rfbI* (1896bp).

These gene specific PCR products were individually electroporated successfully into *S. Typhimurium* containing plasmid pKD46 to construct deletion mutant. The mutant construction was confirmed by sequencing that showed insertion of Kanamycin gene from pKD13 at precise location and thus confirmed deletion of the target genes.

Phenotypic characterization of deletion mutants

To further verify the result of the Tn-seq screening, we sought to characterize the functions of the identified genes. We chose the four genes, *trkA*, *sapD*, *rfbF* and *rfbI*, among the 42 genes identified based on its fitness score. The *trkA* and *sapD* deletion mutants were analyzed for survivability in LB media at 4° C and -20°C every third day from day 0 to Day 12. The mutants did not show significant difference in survivability compared to wild type (Figure 7 and 8).

Since these genes were initially selected by competition within the mutant library, we characterized the four mutants using competition assays in which each mutant was competed against the wild type strain at 1:1000 ratio. The competition assay results were compared to the fitness indices obtained by Tn-seq data (Table 3) for $\Delta trkA$, $\Delta sapD$, $\Delta rfbF$ and $\Delta rfbI$ mutants. The result is shown in Figure 9. Surprisingly, the fitness measurement inferred by Tn-seq data and competition data for each mutant did not match except for $\Delta sapD$ mutant where Tn-seq index and competition assay show meager but similar trend. The growth assay and competition assay were repeated several times with utmost precision to correct mistakes if any that would explain this unexpected result. Also, the mutants were sequenced for second time and the result did not show any problem. All the repetition led to same outcome.

DISCUSSION

The incidence of food-borne illness caused by *Salmonella* has remained unchanged over the last several years and *Salmonella* is consistently the most common bacterial pathogen in food-borne illness cases. The strategies to reduce *Salmonella* contamination of poultry products are based on variety of antimicrobial techniques, which includes chemical and thermal treatments of poultry products and processing environments. These strategies do not have a defined molecular target for inhibition though they are able to decrease the bacterial load. These strategies are not enough to eliminate bacteria from the poultry products; therefore, there is a need for more effective strategy. Application of currently available antibiotics to poultry processing is not an option because today antibiotic in food is unacceptable. Development of a new class of antibiotic specific to the poultry environment may be an option if we can identify molecular targets uniquely required for growth or survival during poultry processing and storage. In this study, we made an effort to identify genes that are conditionally essential during food storage at 4°C and -20°C using Tn-seq method, which could be molecular targets for future antimicrobial development.

Bacterial cells undergo physiological changes in response to temperature downshift and numbers of cold shock proteins are induced to counteract its harmful effects (Phadtare, 2004). We were also able to detect insertions in some of genes encoding cold shock proteins but none of them were recognized as conditionally essential in our study because these genes did not meet our stringent selection criteria of P value <0.01 and Fitness index ≤ 0.3 (Table 4). The cold stress associated genes identified in our study were *rpoS* encoding RpoS, recognized as alternative sigma factor activated during cold shock in *S. Typhimurium* (Miticka, *et al.*, 2003), *katG* a hydroperoxidase, *prop* a proline/glycine betaine transporter, *otsB* trehalose-6-phosphate

phosphatase, *uspA* encoding universal stress protein A, *csdA* cystein sulfinate desulfinate, *clpB* a protein disaggregation chaperone, and *pnp* a member of mRNA degradosome, all of these genes have been previously identified as genes activated during stresses including cold stress in *E. coli* and other bacteria (Phadtare, 2004).

In this study, we were able to identify 42 conditionally essential genes out of which 36 were common to both conditions. The genes encoding potassium transport systems (*trkA*, *trkH*, *sapD* and *sapB*) were identified in our study as conditionally essential genes, indicating the importance of potassium ions in the survival of *Salmonella* at cold temperatures. Based on these findings, further studies to find natural products that would block the potassium pathway or deplete potassium ions could be an efficient means to eliminate *Salmonella* from food during cold storage. However, the phenotypic study results obtained from the four mutants ($\Delta trkA$, $\Delta sapD$, $\Delta rfbF$, and $\Delta rfbI$) did not correlate to the findings of the Tn-seq method, which is not only difficult to explain but difficult to comprehend in absence of any information available on gene functions for *Salmonella* at 4°C and -20°C selective conditions.

The previous study has demonstrated the utility and efficiency of this Tn-seq method in comprehensive identification of conditionally essential genes in *Salmonella*. The phenotype characterization of the 4 deletion mutants in our previous research (Khatiwara *et al.*, 2012) along with previous studies on bile acid resistance demonstrated that our Tn-seq method and bioinformatics analysis assessed gene functions accurately. The genes identified in replicate samples in this study namely 4°C (Rep 1 and Rep 2) and -20°C (Rep 1 and Rep 2) show very high levels of reproducibility (Figure 3 and Figure 4). The genes identified at 4°C and -20°C are mostly overlapping (Figure 5) indicating that these are not randomly chosen genes but identified by Tn-seq method with high level of efficiency and accuracy. Thus, the results of the phenotypic

assays in this study do not discourage us or make us doubt our methodology and previous results (Khatiwara *et al.*, 2012). Instead, we look forward to design and conduct experiments to understand the intricacies of survival of *Salmonella* in various food matrices and finally to develop methods to eliminate this pathogen from our foods.

REFERENCES

- Centers for Disease Control and Prevention: Fact sheet
<http://www.cdc.gov/foodnet/PDFs/FoodNetMMWRQandA2010.pdf>.
- Clatworthy, A.E., Pierson, E., and Hung, D.T. (2007) Targeting virulence: a new paradigm for antimicrobial therapy. *Nat Chem Biol* **3**: 541-548.
- Cox, M.M., Layton, S.L., Jiang, T., Cole, K., Hargis, B.M., Berghman, L.R., *et al.* (2007) Scarless and site-directed mutagenesis in *Salmonella enteritidis* chromosome. *BMC Biotechnol* **7**: 59.
- Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97**: 6640-6645.
- Economic Research Service. (2011) Foodborne Illness Cost Calculator: Salmonella. United States Department of Agriculture. Available at: <
<http://www.ers.usda.gov/Data/FoodborneIllness/>> [Accessed 08 September 2011].
- Franz, E., and van Bruggen, A.H. (2008) Ecology of *E. coli* O157:H7 and *Salmonella enterica* in the primary vegetable production chain. *Crit Rev Microbiol* **34**: 143-161.
- Gawronski, J.D., Wong, S.M., Giannoukos, G., Ward, D.V., and Akerley, B.J. (2009) Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci USA* **106**: 16422-16427.
- Jarvik, T., Smillie, C., Groisman, E.A., and Ochman, H. (2010) Short-term signatures of evolutionary change in the *Salmonella enterica* serovar typhimurium 14028 genome. *J Bacteriol* **192**: 560-567.
- Jones, P.G., VanBogelen, R.A., and Neidhardt, F.C. (1987) Induction of proteins in response to low temperature in *Escherichia coli*. *J Bacteriol* **169**: 2092-2095.
- Khatiwara, A., Jiang, T., Sung, S.S., Dawoud, T., Kim, J.N., Bhattacharya, D., *et al.* (2012) Genome scanning for conditionally essential genes in *Salmonella*. *Appl Environ Microbiol*. 2012 Feb 24. [Epub ahead of print].
- Kwon, Y.M., Bhattacharya, D., and Ricke, S.C. (2009) Genomic approaches to bacterial pathogens using transposon mutagenesis: food safety applications. In: Perspectives on Food Safety Issues of Animal Derived Foods. Eds. Ricke S. C. and Jones, F.T. University of Arkansas Press.
- Miticka, H., Rowley, G., Rezuchova, B., Homerova, D., Humphreys, S., Farn, J., *et al.* (2003) Transcriptional analysis of the *rpoE* gene encoding extracytoplasmic stress response

- sigma factor sigmaE in *Salmonella enterica* serovar Typhimurium. *FEMS Microbiol Lett* **226(2)**: 307-314.
- Newell, D.G., Koopmans, M., Verhoef, L., Duizer, E., Aidara-Kane, A., Sprong, H., *et al.* (2010) Food-borne diseases - the challenges of 20 years ago still persist while new ones continue to emerge. *Int J Food Microbiol* **139**: 3-15.
- Phadtare, S. (2004) Recent developments in bacterial cold-shock response. *Curr Issues Mol Biol* **6(2)**: 125-136.
- Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R.V., Widdowson, M.A., Roy, S.L., *et al.* (2011) Foodborne illness acquired in the United States--major pathogens. *Emerg Infect Dis* **17(1)**: 7-15.
- Tatusov, R.L., Koonin, E.V., and Lipman, D.J. (1997) A genomic perspective on protein families. *Science* **278**: 631-637.
- van Opijnen, T., Bodi, K.L., and Camilli, A. (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* **6**: 767-772.
- Woteki, C.E., and Kineman, B.D. (2003) Challenges and approaches to reducing foodborne illness. *Annu Rev Nutr* **23**: 315-44.
- Zhou, Y., and Landweber, L.F. (2007) BLASTO: a tool for searching orthologous groups. *Nucleic Acids Res* **35**: W678-W682.

Tables

- Table 1.** Oligonucleotides used in this study
- Table 2.** Summary of Illumina sequencing data analysis
- Table 3.** List of all conditionally essential genes identified in this study (fitness index ≤ 0.5 under at least one of the conditions)
- Table 4.** List of genes with transposon insertions identified in this study, previously recognized as cold stress protein encoding genes (Phadtare, 2004)

Figure Legends

- Figure 1.** Gel picture. (A) Genomic DNA from input pool, 4° C output pool and -20°C output pool in duplicates along with DNA marker in lane1. (B) PAGE gel picture showing PCR amplicon of 129bp size obtained from one input pool and two output pools 4° C and -20°C in duplicate along with DNA marker in lane 1.
- Figure 2.** PAGE gel picture showing pooled sample of 129bp that was purified and sent for Illumina sequencing.
- Figure 3.** Reproducibility of biological replicates. Two biological replicate samples were analyzed by Tn-seq for (A) chromosomal ($R^2= 0.99$), and (B) plasmid genes ($R^2=0.99$). Axes represent the total number of sequence reads mapped to insertion sites in all genes (5-80% of the coding region).

Figure 4. Genome-wide view of the fitness index. Genome-wide view of the fitness index⁻¹ (= total read counts in the input pool/ total read counts in the normalized output pool) for each gene obtained under 2 different selective conditions. Fitness index⁻¹ was used for Y-axis (instead of fitness index) for better visualization of mutants with fitness defect as indicated by the high peaks. The picture shows that the replicates were similar indicating highly reproducibility and the genes identified were overlapping in both selective conditions.

Figure 5. Identification of the genes conditionally essential for fitness under 2 different selective conditions: 4°C and - 20°C. (A) Genome-wide view of the fitness index⁻¹ (= total read counts in the input pool/ total read counts in the normalized output pool) for each gene obtained under 2 different selective conditions. Fitness index⁻¹ was used for Y-axis (instead of fitness index) for better visualization of mutants with fitness defect as indicated by the high peaks. The cutoff line of ≤ 0.3 ($=3^{-1}$) used in this study to determine fitness defect . The underlined italicized genes were selected in this study for further mutant characterization. The gene shown in bold italics are the unique genes identified in each condition. (B) The number of the genes identified as conditionally essential for fitness in two selective conditions. Forty-two genes were identified out of which thirty six were identified in both selective conditions.

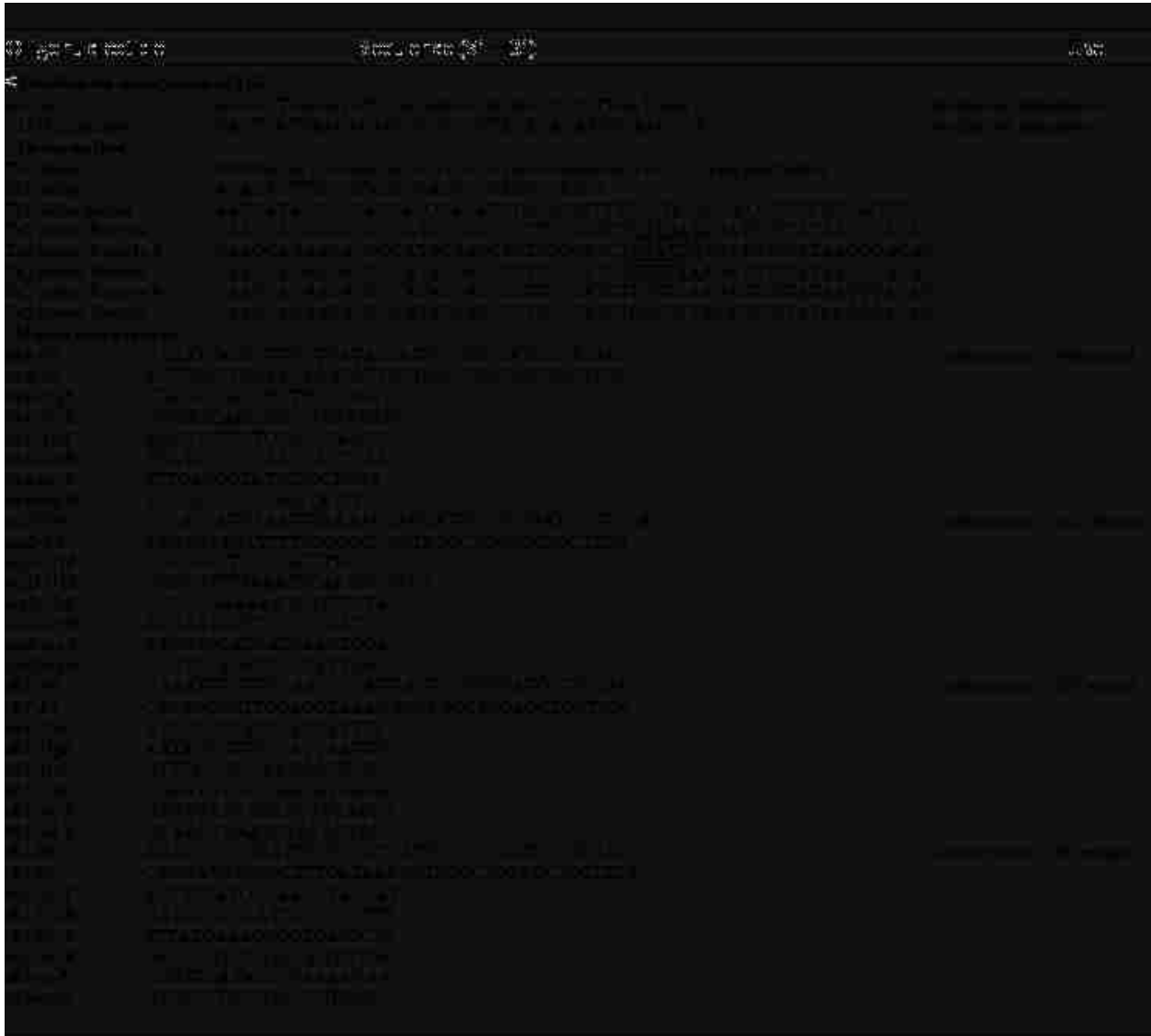
Figure 6. Functional trends associated with two selective conditions. The genes identified in the study were assigned to functional (COG; Clusters of Orthologous Groups) categories using BLAST on Orthologous groups (BLASTO) algorithm.

Figure 7. Individual Survival Assay for $\Delta trkA$ and wild type. Mutants and wild type were incubated separately at 4°C and -20°C. Percentage survivability was calculated at day 0, 3, 6, 9 and 12.

Figure 8. Individual Survival Assay for $\Delta sapD$ and wild type. Mutants and wild type were incubated separately at 4°C and -20°C. Percentage survivability was calculated at day 0, 3, 6, 9 and 12.

Figure 9. Comparison of the fitness indices obtained by Tn-seq data (in Table 3) and competition assay (CA) for $\Delta trkA$, $\Delta sapD$, $\Delta rfbF$, and $\Delta rfbI$ mutants. The comparison was made for 2 different selective conditions (4°C and -20°C) for each mutant. The results of competition assay (n=3) were obtained at day 17 for wild type and mutants ($\Delta trkA$, $\Delta sapD$, $\Delta rfbF$, and $\Delta rfbI$).

Table 1. Oligonucleotides used in this study



The image shows a screenshot of a software application window with a dark background and light text. The window title is "Oligonucleotide Sequences". The interface displays a list of oligonucleotides, each with a unique identifier (e.g., O1, O2, O3, O4, O5, O6, O7, O8, O9, O10, O11, O12, O13, O14, O15, O16, O17, O18, O19, O20, O21, O22, O23, O24, O25, O26, O27, O28, O29, O30, O31, O32, O33, O34, O35, O36, O37, O38, O39, O40, O41, O42, O43, O44, O45, O46, O47, O48, O49, O50, O51, O52, O53, O54, O55, O56, O57, O58, O59, O60, O61, O62, O63, O64, O65, O66, O67, O68, O69, O70, O71, O72, O73, O74, O75, O76, O77, O78, O79, O80, O81, O82, O83, O84, O85, O86, O87, O88, O89, O90, O91, O92, O93, O94, O95, O96, O97, O98, O99, O100). Each entry includes a sequence of nucleotides (A, T, C, G) and some entries have additional information like "Length" and "GC Content". The sequences are listed in a columnar format, with some entries having a "Sequence" column and others having a "Length" column. The text is small and difficult to read, but the overall structure is a table-like list.

Table 2. Summary of Illumina sequencing data analysis

Sample	Total No. of reads	No. of reads with 19bp modified ME sequence	No. of reads with 6bp barcode	No. of with 11-13bp Transposon junction sequence	Transposon junction sequences mapped to the genome	After removal of insertions with < 10 reads	
						No. of reads	No. of insertion sites
Total	35,688,012	33,988,628	33,179,365	33,022,907	33,022,907		
Input				5,412,187	5,412,187	2,816,472	32,544
4°C (1)				6,125,632	6,125,632	2,875,846	32,218
4°C (2)				7,362,089	7,362,089	3,405,066	31,937
-20°C(1)				7,389,502	7,389,502	3,390,063	31,967
-20°C(2)				6,733,497	6,733,497	3,062,067	31,763

Table 3. List of all genes with fitness index ≤ 0.3 under any one condition (Portion of table shown, detail table in CD)

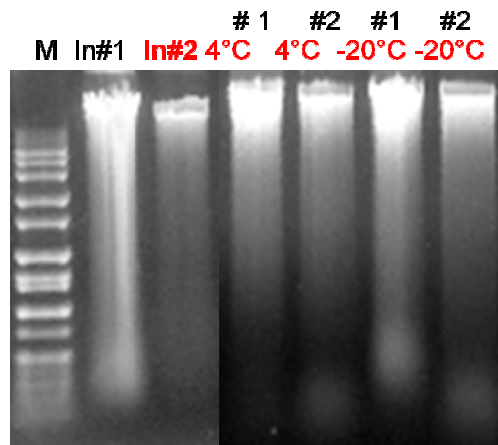
Gene ID	Gene Name	Accession	Description	Fitness Index				
				Condition 1	Condition 2	Condition 3	Condition 4	Condition 5
V1076_0001	0.05	0.10	0.15	0.20	0.25
V1076_0002	0.08	0.12	0.18	0.22	0.28
V1076_0003	0.10	0.15	0.20	0.25	0.30
V1076_0004	0.12	0.18	0.22	0.28	0.32
V1076_0005	0.15	0.20	0.25	0.30	0.35
V1076_0006	0.18	0.22	0.28	0.32	0.38
V1076_0007	0.20	0.25	0.30	0.35	0.40
V1076_0008	0.22	0.28	0.32	0.38	0.42
V1076_0009	0.25	0.30	0.35	0.40	0.45
V1076_0010	0.28	0.32	0.38	0.42	0.48
V1076_0011	0.30	0.35	0.40	0.45	0.50
V1076_0012	0.32	0.38	0.42	0.48	0.52
V1076_0013	0.35	0.40	0.45	0.50	0.55
V1076_0014	0.38	0.42	0.48	0.52	0.58
V1076_0015	0.40	0.45	0.50	0.55	0.60
V1076_0016	0.42	0.48	0.52	0.58	0.62
V1076_0017	0.45	0.50	0.55	0.60	0.65
V1076_0018	0.48	0.52	0.58	0.62	0.68
V1076_0019	0.50	0.55	0.60	0.65	0.70
V1076_0020	0.52	0.58	0.62	0.68	0.72
V1076_0021	0.55	0.60	0.65	0.70	0.75
V1076_0022	0.58	0.62	0.68	0.72	0.78
V1076_0023	0.60	0.65	0.70	0.75	0.80
V1076_0024	0.62	0.68	0.72	0.78	0.82
V1076_0025	0.65	0.70	0.75	0.80	0.85
V1076_0026	0.68	0.72	0.78	0.82	0.88
V1076_0027	0.70	0.75	0.80	0.85	0.90
V1076_0028	0.72	0.78	0.82	0.88	0.92
V1076_0029	0.75	0.80	0.85	0.90	0.95
V1076_0030	0.78	0.82	0.88	0.92	0.98
V1076_0031	0.80	0.85	0.90	0.95	1.00

Table 4.

Gene	Average Fitness Index	
	(4° C)	(-20° C)
<i>csdA</i>	1.05	1.08
<i>clpB</i>	0.88	0.91
<i>pnp</i>	1.02	1.45
<i>rpoS</i>	0.21	1.42
<i>katG</i>	0.63	0.64
<i>prop</i>	0.89	0.69
<i>otsB</i>	0.81	0.60
<i>uspA</i>	1.30	1.01

Figure 1

(A)



(B)

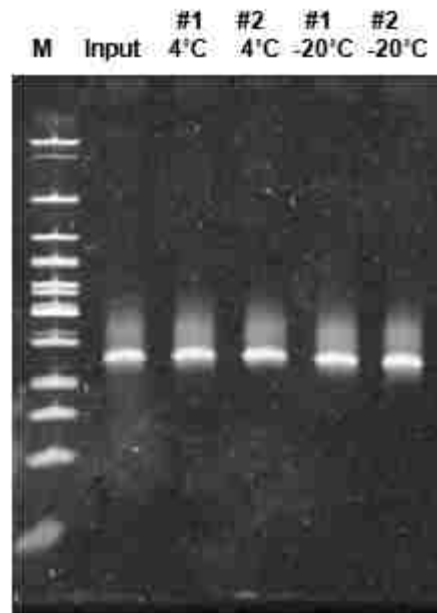


Figure 2.



Figure 3

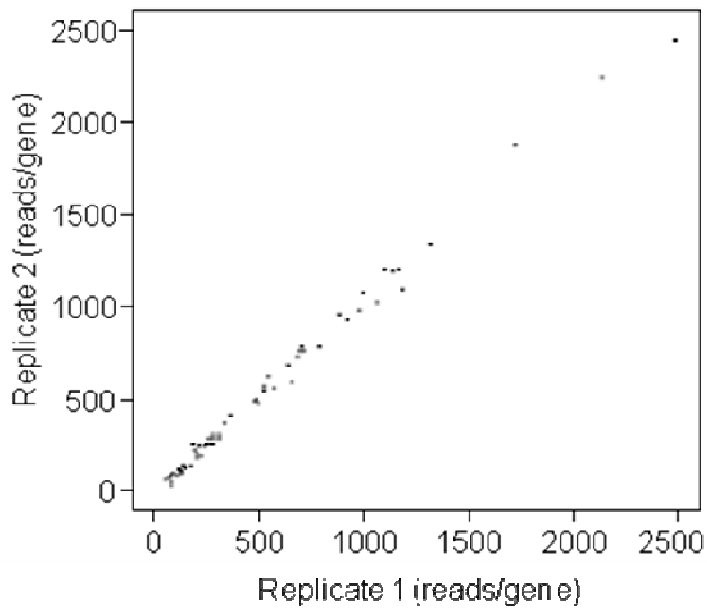
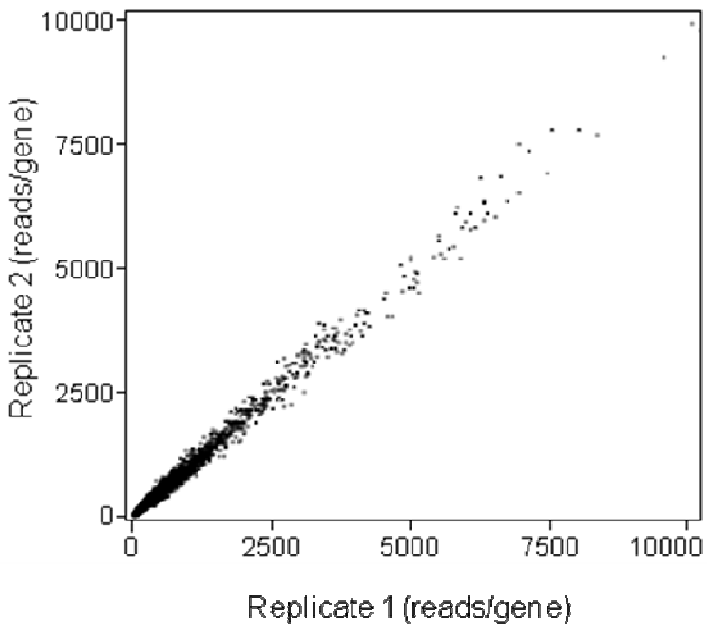


Figure 4

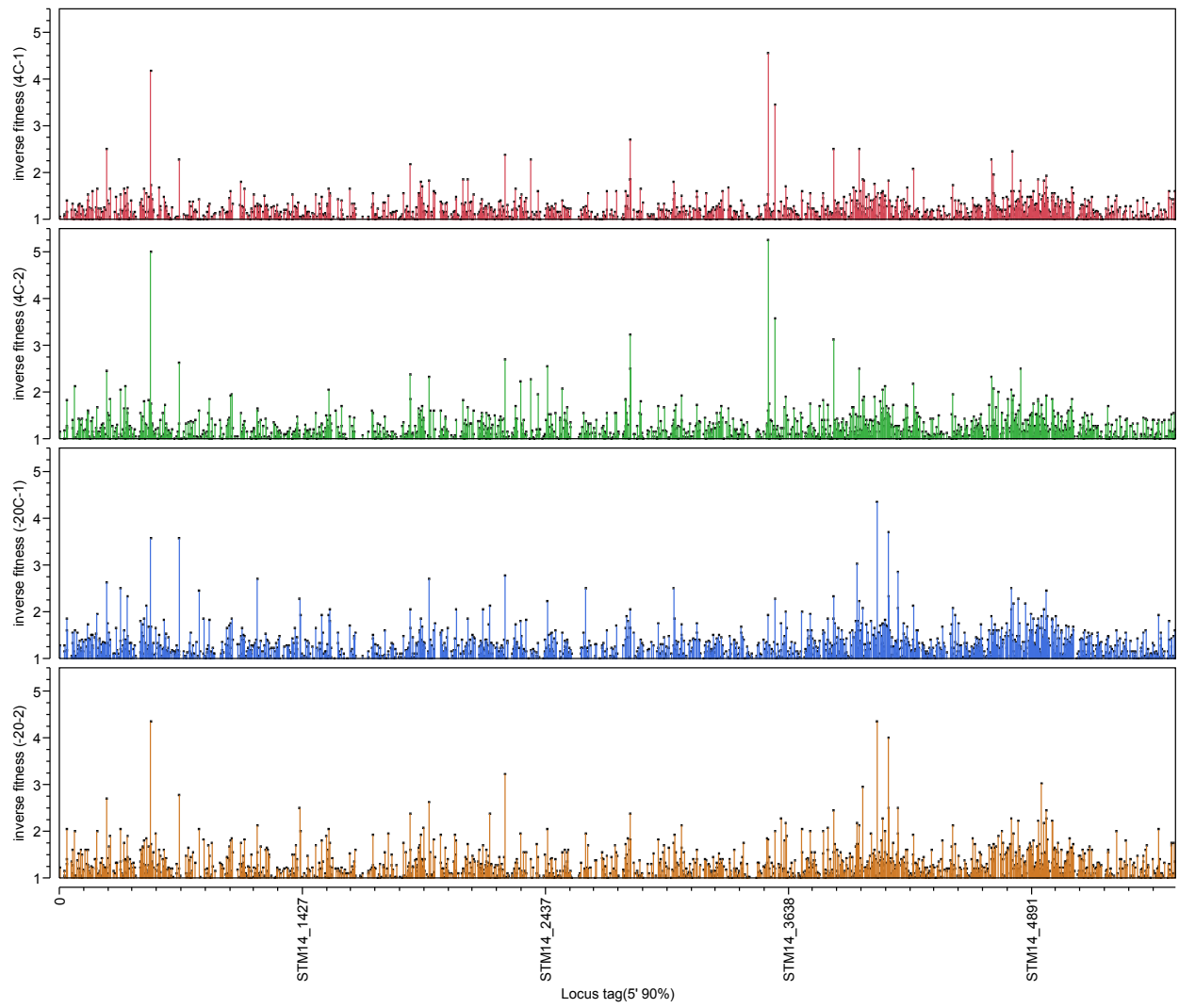
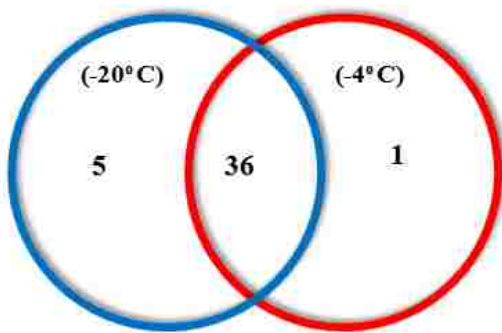
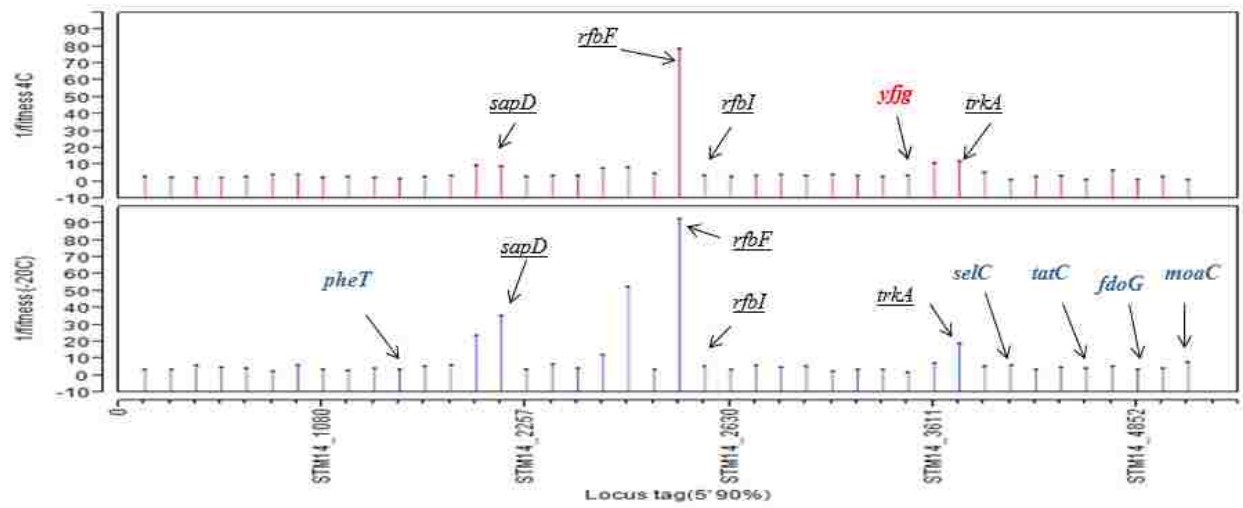


Figure 5.



4°C: *yjg*
 -20°C: *pheT*, *selC*, *tatC*, *fdoG*, *moaC*

Figure 6.

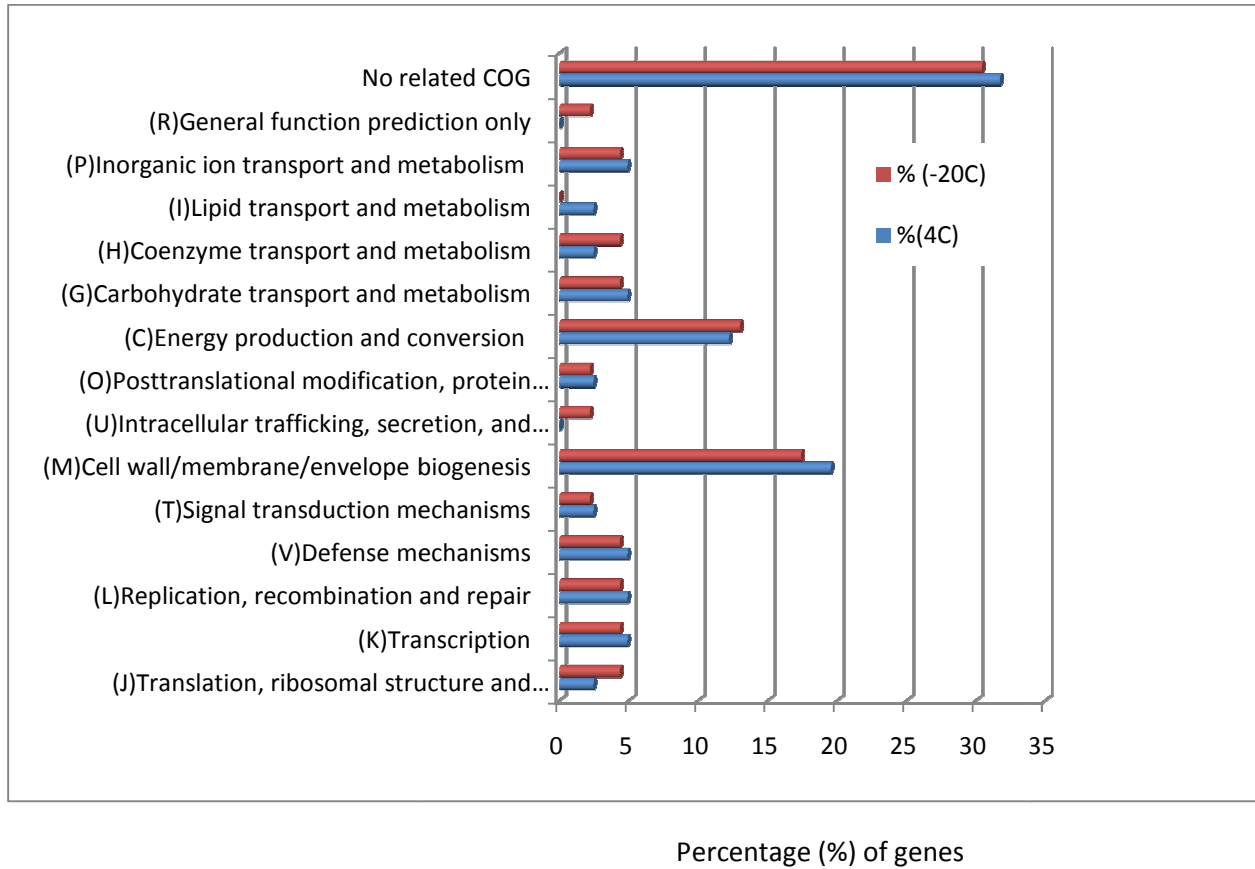


Figure 7.

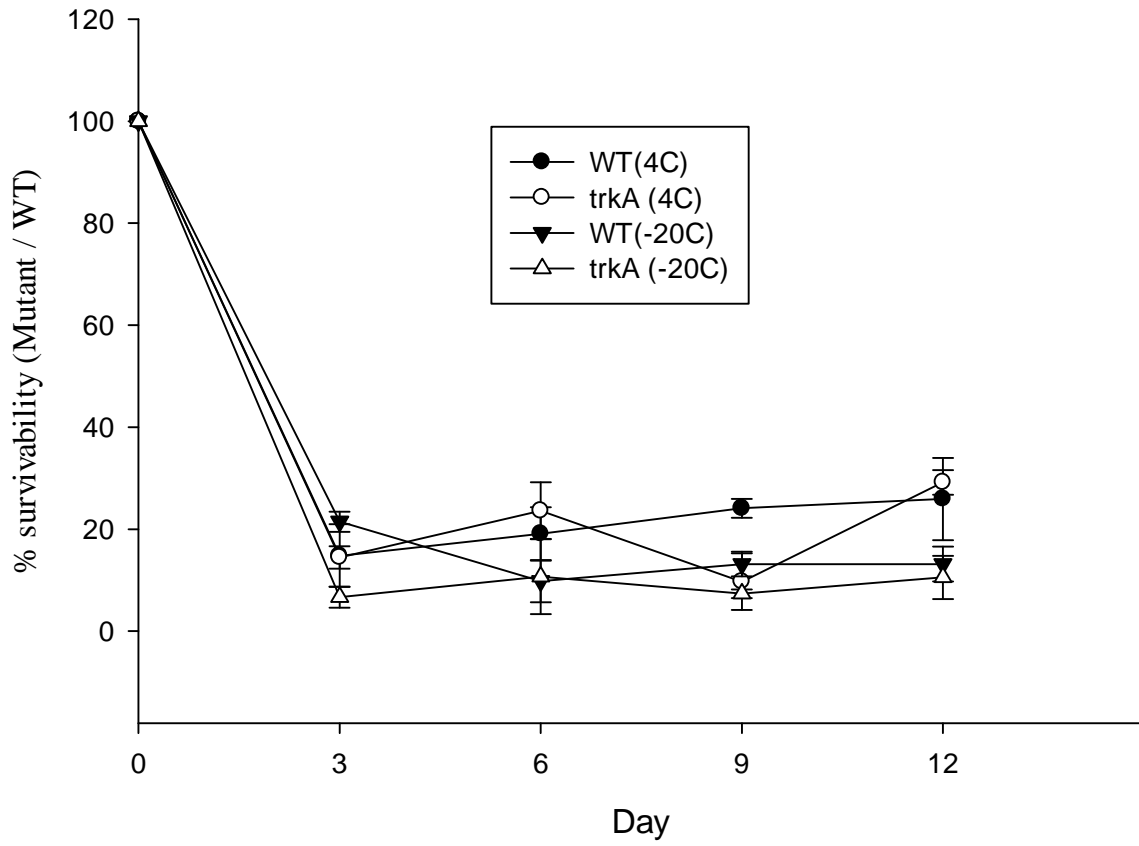


Figure 8.

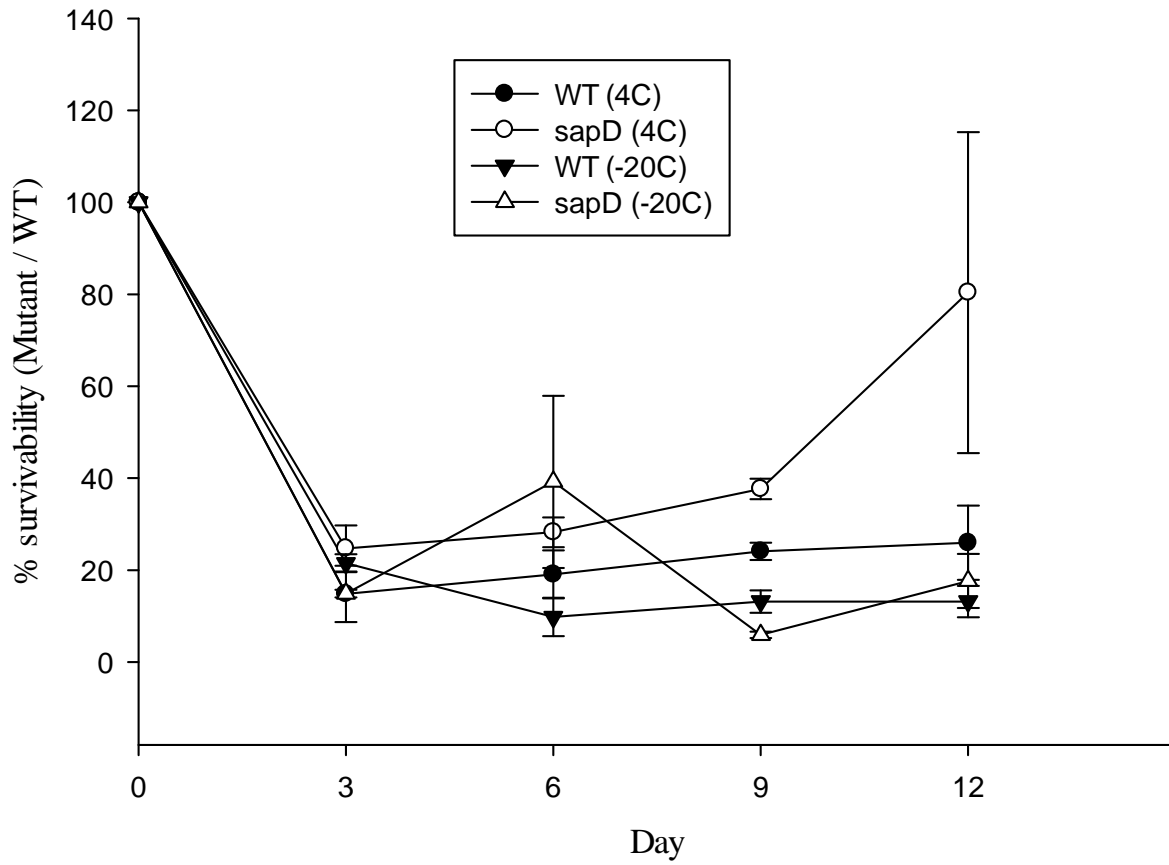
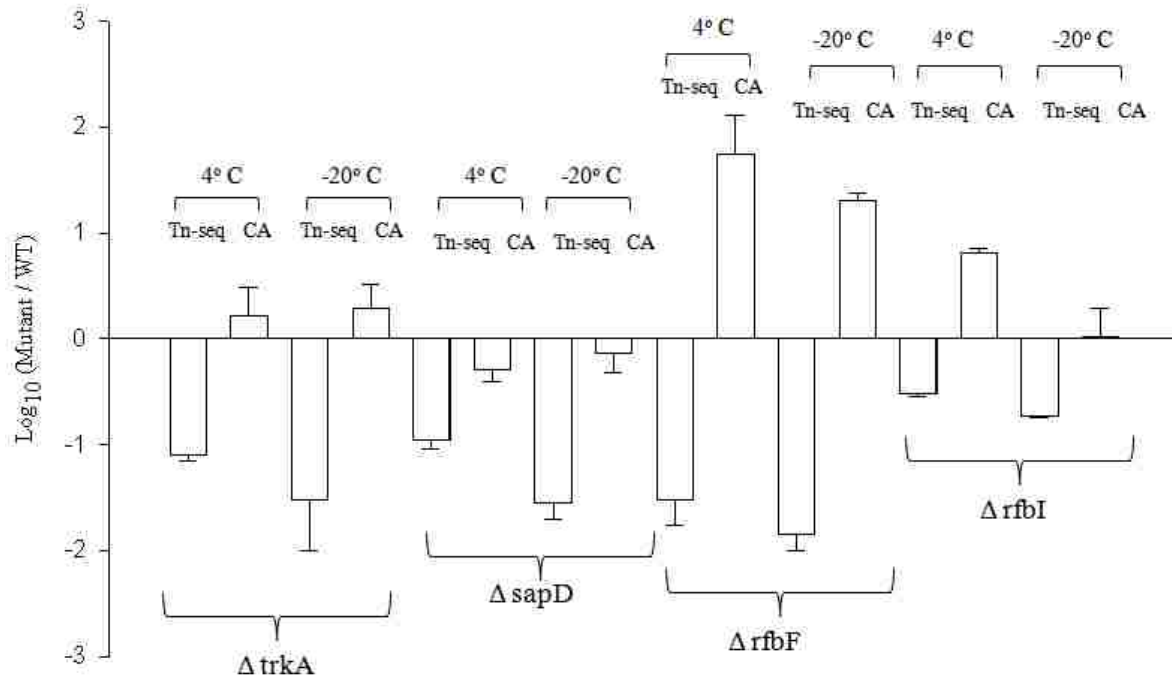


Figure 9.



CHAPTER IV
CONCLUSION

IV. CONCLUSION

Recent rapid development of next-generation sequencing (NGS) has led to generation of complete nucleotide sequences of a large number of genomes. This has facilitated and challenged molecular biology concurrently to develop and use high-throughput techniques to determine the function of the genes identified. Many transposon-based whole-genome analytical techniques have been developed to identify gene functions in bacteria in a high-throughput manner. Most of these approaches are “negative selection” strategies in which large pool of diverse mutants are analyzed to identify mutations that reduce fitness under a particular environment. The latest trend has been the use of whole-genome transposon mutagenesis in combination with massive parallel sequencing to efficiently analyze bacterial genes required for growth and survival at particular selective condition. In this dissertation (Chapter II), we described a new version of Tn-seq method also based on transposon mutagenesis in combination with Illumina sequencing, and used it with stringent cutoffs to obtain robust fitness profiling of entire genome of *S. Typhimurium* 14028 strain under 3 different *in vitro* stress conditions namely low-nutrient, bile-rich, and high temperature (42°C) environments. The phenotypic characterization of the deletion mutants along with previous studies on bile acid resistance of *Salmonella* demonstrated that our Tn-seq method and bioinformatics analysis assessed gene functions accurately. We used our Tn-seq method in conjunction with barcoding strategy and Illumina sequencing to allow high-resolution functional genome scanning for multiple selection conditions of interest. Our method is distinct from other methods in that we employed EZ:Tn5 mutagenesis system. The proven broad-host range of this mutagenesis system along with a commercially available EZ-Tn5TM Transposase (Epicentre BioTechnologies) will make this method easily accessible and applicable

to multitude of bacterial species with appropriate modifications on an antibiotic cassette and its promoter.

Salmonella is an important human foodborne pathogen. The incidence of food borne illness caused by *Salmonella* has remained unchanged over the last several years and *Salmonella* is consistently the most common bacterial pathogen in foodborne illness cases. One common strategy being employed to reduce *Salmonella* contamination in post-harvest poultry production is based on the use of a variety of antimicrobial chemicals that do not have a defined molecular target for inhibition of growth or survival of *Salmonella*. These strategies are not effective enough to eliminate bacteria from the poultry products; therefore, there is a need for more effective strategy. Development of a new class of antimicrobials specific to the poultry environment may be an option if we could identify molecular targets uniquely required for growth or survival during poultry processing and storage. Therefore, we made an effort, first of its kind in chapter III, to identify genes that are conditionally essential during food storage at 4°C and -20°C using Tn-seq method, which could be molecular targets for future antimicrobial development. We were able to identify 42 conditionally essential genes out of which 36 were common to both selective conditions. The genes encoding potassium transport systems (*trkA*, *trkH*, *sapD* and *sapB*) were identified in our study as conditionally essential genes, indicating the importance of potassium ions in the survival of *Salmonella* at cold temperatures. Based on these findings, further studies to find natural products that would block the potassium pathway or deplete potassium ions could be an efficient means to eliminate *Salmonella* from food during cold storage. However, the phenotypic study results obtained from the four mutants did not correlate to the findings of the Tn-seq method. Future study is required to explain the discrepancy between the Tn-seq and the mutant phenotype data before the genes identified in this

study could be explored as potential targets for control of *Salmonella* in poultry processing environments. The concept described in this study could be extended to other stressors unique in poultry processing environments to accomplish pathway-specific and more effective control of *Salmonella* in poultry products.

The genes identified in these studies include many putative or hypothetical genes or the genes with unknown functions. Understanding the functions of these genes and products would reveal various unknown mechanisms of *Salmonella* survival and persistence during its life cycle inside a host or in the surrounding environment. In addition, these genes have great potentials to be used as good candidate targets for development of vaccines and novel antimicrobials. In conclusion, our research work has provided a solid platform and powerful genomics tool for the researchers to design and conduct experiments to understand the intricacies of growth and survival of *Salmonella* in its host, poultry products as well as other food matrices and finally to develop methods to effectively eliminate this pathogen from our foods.

APPENDICES

Permission 1. (Annual review of genomics and human genetics)

Confirmation Number: 10943502

Order Date: 02/26/2012

Customer Information

Customer: Anita Khatiwara

Account Number: 3000493461

Order Details Annual review of genomics and human genetics

Order detail ID:61906399

ISSN:1527-8204

Publication year:2008

Publication Type: Journal

Publisher: ANNUAL REVIEWS

Rights holder: ANNUAL REVIEWS, INC.

Author/Editor: ER Mardis

Permission Status:  **Granted**

Permission type: Republish or display content

Type of use: Republish in a dissertation

Requested use: Dissertation

Republishing organization: University of Arkansas

Organization status: Non-profit 501(c)(3)

Republication date:05/31/2012

Circulation/ Distribution:10


Type of content: Figure/ diagram/ table

Description of requested content: Next generation DNA sequencing methods

Page range(s):390,392,394,395

Translating to: No Translation

Requested content's publication date: 06/24/2008

Payment Method: CC ending in 

Rightsholder terms apply (see terms and conditions)

[Copy](#)

Order Total: \$ 28.50

Permission 2. (Trends in Genetics)

Supplier	Elsevier Limited; The Boulevard, Langford Lane Kidlington,Oxford,OX5 1GB,UK
Registered Company Number	1982084
Customer name	Anita Khatiwara
Customer address	1260 W Maple street Fayetteville, AR 72701
License number	2839470406186
License date	Jan 31, 2012
Licensed content publisher	Elsevier
Licensed content publication	Trends in Genetics
Licensed content title	The impact of next-generation sequencing technology on genetics
Licensed content author	Elaine R. Mardis
Licensed content date	March 2008
Licensed content volume number	24
Licensed content issue number	3
Number of pages	9
Start Page	133
End Page	141
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	3
Format	both print and electronic
Title of your thesis/dissertation	Genome scanning of conditionally essential genes in Salmonella
Expected completion date	May 2012
Estimated size (number of pages)	300
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.0 USD / 0.0 GBP
Total	0.00 USD