University of Arkansas, Fayetteville
ScholarWorks@UARK

Graduate Theses and Dissertations

12-2016

Genetic Determinants of Salmonella and Campylobacter Required for In Vitro Fitness

Rabindra Kumar Mandal
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

 Part of the Bioinformatics Commons, Molecular Genetics Commons, and the Pathogenic Microbiology Commons

Citation

Mandal, R. K. (2016). Genetic Determinants of Salmonella and Campylobacter Required for In Vitro Fitness. *Graduate Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/1785

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

Genetic Determinants of *Salmonella* and *Campylobacter* Required for *In Vitro* Fitness

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Poultry Science

by

Rabindra Kumar Mandal
Institute of Agriculture and Animal Science, Tribhuvan University
Bachelor of Veterinary Science and Animal Husbandry, 2010

December 2016
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Dr. Young Min Kwon
Dissertation Director

_____
Dr. Ravi D. Barabote
Committee Member

_____
Dr. Daniel J. Lessner
Committee Member

_____
Dr. Byung-Whi Kong
Committee Member

ABSTRACT

Non-typhoidal *Salmonella* (NTS) and *Campylobacter* play a major role in foodborne illness caused by the consumption of food contaminated by pathogens worldwide. A comprehensive understanding of the genetic factors that increase the survival fitness of these foodborne pathogens will effectively help us formulate mitigation strategies without affecting the nutrition ecology. The objective of this study was to identify the genetic determinants of *Salmonella* and *Campylobacter* that are required for fitness under various *in vitro* conditions. For the purpose, we used a high throughput Transposon sequencing (Tn-seq) that utilizes next generation sequencing (NGS) to screen hundreds of thousands of mutants simultaneously. In Chapter 1, we reviewed the technical aspects of different Tn-seq methods along with their pros and cons and compressive summary of recently published studies using Tn-seq methods. In Chapter 2, we exposed complex Tn5 library of *Salmonella* Typhimurium 14028S (*S*. Typhimurium) to the mimicked host stressors *in vitro* conditions. Such as low acidic pH (pH 3) found in the stomach, osmotic (3% NaCl) and short chain fatty acid (SCFAs, 100 mM Propionate) found in intestine, and oxidation (1mM H_2O_2) and starvation (12-day survival in PBS) found in macrophage. There was an overlapping set of 339 conditionally essential genes (CEGs) required by *S*. Typhimurium to overcome these host stressors. In Chapter 3, we screened of *S*. Typhimurium Tn5 library for desiccation survival. *Salmonella* spp. is the most notable and frequent cause of contamination in low-water activity foods. We identified 61 genes and 6 intergenic regions required for fitness during desiccation stress. In Chapter 4, the essential genome of *Campylobacter jejuni* (*C. jejuni*) NCTC 11168 and *C. jejuni* 81-176 was investigated using Tn-seq. We identified 166 essential protein-coding genes and 20 essential transfer RNA (tRNA) in *C. jejuni* NCTC 11168 which were intolerant to Tn5 insertions during *in vitro* growth. The reconstructed library *C. jejuni* 81-

176 had 384 protein coding genes with zero Tn5 insertions. The genetic determinants *Salmonella* and *Campylobacter* identified in this study have high potential to be explored as food safety intervention, therapeutic and vaccine target to curb the spread of the foodborne pathogens making world a safer place.

ACKNOWLEDGEMENT

I would like to express my heartiest thanks and appreciation to my major advisor Dr. Young Min Kwon. He not only filled in me with enthusiasm, motivation and desire to accomplish the research projects but also shared his experience and taught me about life in general. He is a good teacher, mentor and advisor.

My sincere gratitude goes to my dissertation committee members Dr. Ravi Damodar Barabote, Dr. Daniel J. Lessner, and Dr. Byung-Whi Kong for their constructive criticism and guidance throughout my research. I would also like to express my thanks to my collaborators Dr. Jiangchao Zhao, Dr. Steven C. Ricke, Dr. Robert F. Wideman, Dr. Douglas D. Rhoads, and Dr. Billy Hargis for their help on different projects.

I would also like to convey my thanks to the staffs of High Performance Computing Center (AHPCC) Dr. David Chaffin, Dr. Jeff Pummill and Dr. Pawel Wolinski for their assistance. I am very thankful to our lab manager Dr. Tieshan Jiang who helped me to get started with the experiments. I am also grateful to my lab mates Dr. Turki Dwoud, Yichao, Deepti, Sardar, Bishnu, Dr. Nicole Calhoun and Adil for making lab an enjoyable place. Thank you Miss Donna for being my best lab neighbor.

Above all, I am indebted to my mom Sumitra Devi Mandal and father Ram Prasad Mandal whose courage helped me to excel in every realm of my life. I am grateful to my sister Lilam and brother Birendra and my wife's family. Finally, I acknowledge my wife Dr. Anita Mandal (Nimmi) for her endurance, patience, understanding and love during the past few years that made this dissertation possible.

DEDICATION

This dissertation is dedicated to my mother Sumitra Devi Mandal whose dream helped me achieve for what I am today.

TABLE OF CONTENTS

LIST OF PUBLICATIONS (Published)

1. Kwon, Young Min, Steven C. Ricke, and **Rabindra K. Mandal**. "Transposon sequencing: methods and expanding applications." ***Applied microbiology and biotechnology*** 100, no. 1 (2016): 31-43. (Chapter 1).

LIST OF PUBLICATIONS (In Preparation)

2. **Mandal, Rabindra K**, Tieshan Jiang and Young Min Kwon. "Genetic mechanisms of *Salmonella* enterica serovar Typhimurium for overcoming host stressors". (Chapter 2).

3. **Mandal, Rabindra K** and Young Min Kwon. "Global screening of *Salmonella* genes for desiccation survival". (Chapter 3).

4. **Mandal, Rabindra K**, Tieshan Jiang and Young Min Kwon. "Essential genome of *Campylobacter jejuni*". (Chapter 4).

INTRODUCTION

Foodborne illness is an illness resulting from the consumption of food contaminated with harmful microbes or chemical substances. Foodborne illness causes more than 250 diseases ranging from diarrhea to cancers where most of illness result from foodborne pathogens like bacteria, viruses, and parasites. Many foodborne pathogens are present in healthy animal usually in intestine that contaminate food either by direct contact or cross-contamination with intestinal contents during any point in the farm-to-table continuum. Raw food of animal origin i.e. raw meat and poultry, raw eggs, unpasteurized milk, raw shellfish, raw fruits, and vegetables are the most likely to be contaminated with foodborne pathogens. Despite the efforts to subdue the foodborne illness, this common and costly but yet preventable disease, continues to pose significant threat to human health and economy globally.

Recently, World Health Organization (WHO) Foodborne Disease Burden Epidemiology Reference Group (FERG) estimated 600 million illnesses and 420,000 deaths due to foodborne disease worldwide in 2010 where diarrheal foodborne illness, caused by bacteria, virus and protozoa were solely responsible more than 50% of deaths (230,000). *Campylobacter* spp. and norovirus were the most prominent cause of diarrheal foodborne illness while non-typhoidal *Salmonella enterica* (NTS) was the major cause of diarrheal death. Globally, 18 million Disability Adjusted Life Years (DALYs), a measure that calculates overall burden of disease expressed as the number of years lost because of ill-health, disability or early death, were attributed to foodborne pathogens especially NTS and enteropathogenic *Escherichia coli* (World Health Organization 2015).

In the United States (US), foodborne pathogens acquired domestically is estimated to cause 9.4 million foodborne illness, 55,961 hospitalizations, and 1,351 deaths. Each year 1 in 6 people gets sick due to the consumption of contaminated food and water. The most common cause of foodborne illness was norovirus (58%) followed by NTS (11%), *Clostridium perfringens* (10%), and *Campylobacter* spp. (15%) whereas NTS was the leading cause of death (Scallan et al. 2011). NTS and *Campylobacter* are among the leading cause of foodborne illnesses resulting in hospitalization and death. Furthermore, foodborne illness acquired in the US places over 15.5 billion economic burden on public annually (Hoffmann 2015). Cost of treatment varies greatly per case ranging from the $202 for *Cyclospora cayetanensis* to $3.3 million for *Vibrio vulnificu*s (Hoffmann, Maculloch, and Batz 2015).

Controlling of these foodborne bacterial pathogens become more challenging through additional aspects of the ecology of the foodborne pathogens, including antibiotic resistance, cross-protection, and stress adaptation. Antibiotic resistance represents a major public health problem and is a threat to management and treatment of bacterial disease of both medical and veterinary importance. Historically, the use of antibiotics in animal production system was blamed for the proliferation of antibiotic resistant organisms. This led more countries to ban antibiotics as growth promoter. However, the end result with regard to emerging foodborne pathogens such as *Campylobacter jejuni*, *Salmonella spp.* and Verotoxigenic *Escherichia coli* (VTEC) do not represent decrease in antibiotic resistance (Ricke et al. 2015, 107; Koluman and Dikici 2013, 57-69). This may be expected because the most microorganism can acquire and express antibiotic resistance. Moreover, the plasmids carrying antibiotic resistance genes are maintained stably because it also carries essential genes such as those responsible for efficient colonization of gut (Pendleton et al. 2015, 215).

Cross protection is a phenomenon where exposure of bacteria to a stress leads to increased fitness in the same or other stresses by altering gene expression (Wesche and Ryser 2013, 422-437). Bacteria encounter various stresses throughout their life cycle particularly at various points from farm to fork like heat, pressure, or osmotic shock, acids, detergents and bacteriocins. In addition, these stresses can alter the virulence properties of pathogens and can contribute to the survival of these pathogens during infection process that can aggravate the host condition (Begley and Hill 2015, 191-210). Consequently, the ability of foodborne pathogens to mount the stress adaptation response add in the morbidity and mortality resulting from the foodborne illness.

Potentially, other phenomenon that can contribute to prevalence of foodborne pathogen illness could be predictive adaptation, which provide microbe the ability to anticipate and pre-emptively respond to the regular environmental fluctuations (temporally distributed stimuli) that confers a considerable fitness advantage for survival of an organism. Predictive adaptation is thought to be ubiquitous in all three domains of life (Bacteria, Archaea, and Eukaryotes). Human body is the perfect ecological niche where a bacterium can find seamless temporal trends in distribution of stresses encountered during the infection process and can increase the severity of foodborne illness (Mitchell et al. 2009, 220-224; Tagkopoulos, Liu, and Tavazoie 2008, 1313-1317).

Thus, identifying the genetic basis of foodborne pathogen in stress conditions encountered during infection cycle will help to elucidate the mechanism of antibiotic resistance, cross protection / stress adaptation and adaptive prediction. Ultimately, this can pinpoint the Achilles heel of foodborne pathogen and can possibly be used for more effective strategy to mitigate the foodborne pathogen from food in a manner that will preserve the integrity of ecological food systems.

In this doctoral dissertation, we have tried to fill the gaps in understanding the genetic determinants of *Salmonella* and *Campylobacter*, the leading bacterial causes of foodborne illness, *in vitro* stress conditions. A high throughput transposon sequencing (Tn-seq) was applied to define the genetic factors of required by *Salmonella* and *Campylobacter* during various condition.

The Chapter 1 is a review of transposon sequencing (Tn-seq) methods. This chapter elaborate the recent development in Tn-seq methods and its growing applications from fitness profiling *in vitro* or *in vivo*. Furthermore, it describes the implementation of Tn-seq methods to novel experimental design for discovery of bacterial factors involved in more specific biological processes.

In Chapter 2, we have mimicked the host stressors encountered by *Salmonella* Typhimurium during host infection *in vitro* like low acidic pH in the stomach, osmotic and short chain fatty acid (SCFAs) in intestine, and oxidation and starvation in macrophage and applied Tn-seq method to unveil the stress resistance genes of bacteria. The research investigates into the genetic mechanisms of *Salmonella* Typhimurium to overcome host stressors.

Furthermore, in Chapter 3, we performed global screening of *Salmonella* genes for desiccation survival. *Salmonella* has the ability to survive desiccation in foods and food processing facilities for years and is the most notable and frequent cause of contamination in low-water activity foods.

Finally, Chapter 4, investigates into the essential genome of *Campylobacter* with a view to possibly identify some novel genetic factors that can be used as the food safety intervention, therapeutic target or vaccine development.

References

Begley, Máire and Colin Hill. 2015. "Stress Adaptation in Foodborne Pathogens." *Annual Review of Food Science and Technology* 6: 191-210.

Hoffmann, Sandra. 2015. "USDA Economic Research Service-Quantifying the Impacts of Foodborne Illnesses." .

Hoffmann, Sandra, Bryan Maculloch, and Michael Batz. 2015. "Economic Burden of Major Foodborne Illnesses Acquired in the United States." *Econ Res Serv, US Dep Agric, Washington, DC), Pp EIB–140*.

Koluman, Ahmet and Abdullah Dikici. 2013. "Antimicrobial Resistance of Emerging Foodborne Pathogens: Status Quo and Global Trends." *Critical Reviews in Microbiology* 39 (1): 57-69.

Mitchell, Amir, Gal H. Romano, Bella Groisman, Avihu Yona, Erez Dekel, Martin Kupiec, Orna Dahan, and Yitzhak Pilpel. 2009. "Adaptive Prediction of Environmental Changes by Microorganisms." *Nature* 460 (7252): 220-224.

Pendleton, S., D. D'Souza, S. Joshi, and I. Hanning. 2015. "Current Perspectives on Campylobacter." *Emerging Issues, Technologies, and Systems*: 215.

Ricke, Steven C., J. Rivera Calo, Pravin Kaldhone, SC Ricke, JR Donaldson, and CA Phillips. 2015. "Salmonella Control in Food Production: Current Issues and Perspectives in the United States." *Food Safety: Emerging Issues, Technologies and Systems*: 107.

Scallan, Elaine, Robert M. Hoekstra, Frederick J. Angulo, Robert V. Tauxe, Marc-Alain Widdowson, Sharon L. Roy, Jeffery L. Jones, and Patricia M. Griffin. 2011. "Foodborne Illness Acquired in the United States—major Pathogens." *Emerg Infect Dis* 17 (1).

Tagkopoulos, I., Y. C. Liu, and S. Tavazoie. 2008. "Predictive Behavior within Microbial Genetic Networks." *Science (New York, N.Y.)* 320 (5881): 1313-1317.

Wesche, Alissa M. and Elliot T. Ryser. 2013. "Stress Adaptation, Survival and Recovery of Foodborne Pathogens." *Guide to Foodborne Pathogens, Second Edition*: 422-437.

World Health Organization. 2015. *WHO Estimates of the Global Burden of Foodborne Diseases: Oodborne Disease Burden Epiemiology Reference Group 2007-2015* World Health Organization.

CHAPTER 1:

Transposon sequencing: methods and expanding applications

Young Min Kwon[1, 2, 3 *], Steven C. Ricke[1, 2, 3, 4] and Rabindra K. Mandal[1]

[1]Department of Poultry Science, University of Arkansas, Fayetteville, AR, 72701

[2]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, 72701

[3]Center of Food Safety, University of Arkansas, Fayetteville, AR, 72704

[4]Department of Food Science, University of Arkansas, Fayetteville, AR, 72704

*Corresponding author:

Department of Poultry Science

College of Agricultural, Food and Life Sciences

University of Arkansas

Fayetteville, AR 72701

Phone : (479) 575-4935

Email ykwon@uark.edu

Applied microbiology and biotechnology

Abstract

A comprehensive understanding of genotype-phenotype links in bacteria is the primary theme of bacterial functional genomics. Transposon-sequencing (Tn-seq) or its equivalent methods that combine random transposon mutagenesis and next generation sequencing (NGS) represents a powerful approach to understand gene functions in bacteria on a genome-wide scale. This approach has been utilized in a variety of bacterial species to provide comprehensive information on gene functions related to various phenotypes or biological processes of significance. With further improvements in the molecular protocol for specific-amplification of transposon-junction sequences and increasing capacity of next generation sequencing technologies, the applications of Tn-seq have been expanding to tackle questions important, yet difficult to address in the past. In this review, we will discuss the technical aspects of different Tn-seq methods along with their pros and cons to provide a helpful guidance for those who want to implement or improve Tn-seq for their own research projects. In addition, we also provide a comprehensive summary of recent published studies based on Tn-seq methods to give an updated perspective on the current and emerging applications of Tn-seq.

Key Words: transposon sequencing (Tn-seq), functional genomics, bacteria, gene functions, next generation sequencing

Introduction

One of the major goals in bacterial genetics is to understand the genetic mechanisms underlying the phenotypes of interest. Among various approaches to reveal the underlying genetic mechanism(s) of a phenotype, the most common initial step is to identify the genetic factors involved in or responsible for expression of the phenotype. Traditionally the gene discovery process has been a rate-limiting step that slows down the entire process of understanding the mechanism. Transposon mutagenesis has been one of the major tools that has contributed significantly to gene discovery in bacteria mainly through loss-of-function screening. However, the necessity to assess the phenotype of each mutant individually required considerable amount of labor and time, thus limiting the total number of mutants that could be screened. As interest in high-throughput applications has increased, the methods that allow comprehensive screening of a large number of mutants have been developed and progressed significantly over the past two decades to accelerate the screening process, including signature-tagged mutagenesis (STM)(Mazurkiewicz et al. 2006), followed by microarray-based footprinting of transposon mutants (Sassetti and Rubin 2002) and more recently transposon sequencing (Tn-seq)(Barquist et al. 2013a; van Opijnen and Camilli 2013). The Tn-seq method is the most recent addition to the transposon tool box, aided mainly by the development of high-throughput NGS technologies. Since the first reports on the development of Tn-seq in 2009 (Gawronski et al. 2009; Goodman et al. 2009; Langridge et al. 2009; van Opijnen et al. 2009), various modifications have been made and applied to facilitate gene discovery in diverse bacterial species. With this more comprehensive approach, high-resolution functional screening of the whole genome can be performed routinely in a small laboratory for various bacterial species, providing remarkably rich information on gene functions for almost every single gene, including both protein-coding gene

and noncoding genes, involved in a wide range of biological processes. In this review, we will discuss the recent development in Tn-seq methods and its expanding applications from a rather straightforward fitness profiling *in vitro* or *in vivo* to implementation of novel experimental designs for discovery of bacterial factors involved in more specific biological processes.

Transposon mutagenesis

Transposons are genetic elements that can move from one genomic location to another. This "mobile nature" of transposons has been harnessed by microbial geneticists for convenient use of transposons as powerful tools for random mutagenesis in bacteria (Hayes 2003). For Tn-seq analysis, Tn5 and *mariner* transposons have been used most frequently among others, due to their simple procedures, broad-host range, and well-characterized near random nature of those particular transposons (Barquist et al. 2013a). There are various ways to deliver the transposon of choice into the cells for transposon mutagenesis (Maloy 2007). These include the methods based on phage delivery systems (Santiago et al. 2015; Sassetti et al. 2001), plasmid delivery systems (de Lorenzo and Timmis 1994; Martínez-García et al. 2011), *in vivo* mutagenesis by electroporation of transposon-transposase complex (Goryshin et al. 2000), and *in vitro* mutagenesis using a purified transposase enzyme followed by natural transformation (Hendrixson et al. 2001; Reid et al. 2008). For more details, readers are encouraged to retrieve the corresponding references.

Development of Tn-seq methods

The basic and critical step common to all Tn-seq methods and its variations is to amplify transposon-junction sequences in an insertion mutant pool specifically but comprehensively without bias as much as possible (Barquist et al. 2013a; van Opijnen and Camilli 2013). Once

transposon-junction sequences are amplified, they are sequenced in depth by NGS to obtain a quantitative profile of all transposon insertions in the library. From the collected DNA sequence data, the DNA sequence of each read is used to precisely locate each transposon insertion in the genome and accordingly, the number of DNA sequence reads originating from the same insertion serve as a measure of relative abundance of the corresponding transposon mutant in the mutant pool. When Tn-seq profiles of a library are quantitatively compared with an appropriate normalization and statistical method between before and after a selection, the genetic factors that are required for optimal growth or survival under the selection process can easily be identified on a genomic scale.

Since the first versions of Tn-seq methods were reported (Gawronski et al. 2009; Goodman et al. 2009; Langridge et al. 2009; van Opijnen et al. 2009), several variations on the method have been described (Christen et al. 2011; Dawoud et al. 2014; Gallagher et al. 2011; Khatiwara et al. 2012; Klein et al. 2012). These variations differ mainly in the manner in which specific amplification of the transposon-junction sequences is accomplished. More specifically, these methods employ different strategies to attach the common primer binding sites to transposon-flanking regions allowing PCR-amplification to occur between the binding sites for the transposon-specific primer and the common primer binding sites on the transposon-flanking regions. The common strategies for amplification of transposon-junction sequences used in different Tn-seq methods termed in various names, including INSeq (Goodman et al. 2009), Tn-seq (van Opijnen et al. 2009), TraDIS (Langridge et al. 2009), Tn-seq circle (Gallagher et al. 2011), and HITS (Gawronski et al. 2009) are summarized in (Febrer et al. 2011)(*See* Figure 2 in Febrer (2011) for comparative graphical illustration of the different strategies).

Approaches based on C-tailing are the recent technical additions to current Tn-seq methods. The C-tailing procedure uses terminal transferase activity to add poly C tails to 3' end to either single stranded or double stranded DNA. When this reaction is performed in the presence of the mixture of dCTP and dideoxy CTP (ddCTP) at a certain ratio, the average lengths of the C-tail can be efficiently controlled (Lazinski and Camilli 2013). This approach was adopted to attach C-tails to the 3' ends of randomly sheared gDNA of a transposon insertion library. The C-tails served subsequently as a binding site for poly G primer to amplify transposon-junction sequences in conjunction with a transposon-specific primer (Klein et al. 2012). Additional research based on the same Tn-seq method further established the robustness of the method (Carter et al. 2014; Kamp et al. 2013; McDonough et al. 2014; Shan et al. 2015; Valentino et al. 2014). Recently, our lab developed a convenient protocol based on single primer extension of transposon-junction sequences using a transposon-specific primer. The single stranded DNA fragments thus synthesized are subsequently C-tailed using a terminal transferase. The resulting C-tailed transposon-junction fragments can thus be easily amplified with transposon-specific primer and poly G primer (Dawoud et al. 2014).

A more recently developed Tn-seq strategy, termed random barcode transposon-site sequencing (RB-TnSeq), is based on incorporating random DNA barcodes into the transposon and utilizing them for fitness profiling in place of transposon-junction sequences (Wetmore et al. 2015). This RB-TnSeq method simplifies the steps to prepare the PCR library because the random DNA barcodes located internally inside the transposon can be easily PCR-amplified with two universal primers flanking the barcode region. Consequently, this simplified PCR step increases the throughput of mutant fitness profiling significantly. However, it requires additional steps of

random barcode tagging of transposon before construction of a mutant library and initially establishing a database for insertion-barcode pairs (Wetmore et al. 2015).

One issue associated with a transposon mutant library generated using a suicide delivery plasmid is that a significant portion of the mutants could be pseudo-transposon mutants that result from integration of the transposon-delivery plasmid into the chromosome. When this type of library is used for Tn-seq analysis, a large number of sequence reads are from transposon-junctions of delivery vector rather than the true transposon insertions in chromosome or plasmid of the host cell, resulting in a waste of valuable sequence reads. Santiago et al. (2015) recently described a simple strategy to address this issue by incorporating two recognition sites for a rare-cutting restriction endonuclease (e.g. *Not*I) on both sides of one inverted repeat (IR) from which the transposon-junction sequences are obtained (Santiago et al. 2015). The genomic DNA from the library are digested with the rare-cutting restriction enzyme, and the resulting small fragments can be efficiently removed by size-fractionation before the next step for preparation of the Tn-seq amplicon library.

In Table 1 different Tn-seq methods are grouped according to the strategies used to accomplish amplification of transposon-junction sequences accompanied by description of their characteristics.

Comparison of Tn-seq methods

All of the Tn-seq methods that have been described until now have been used successfully to identify genes important for the various biological processes of interest, supporting their utility as a functional genomics tool. However, the sensitivity of the gene discovery can be greatly influenced by the comprehensiveness and quantitative accuracy of the resulting Tn-seq profiles.

It is expected that any bias occurring during the preparation of Tn-seq library would negatively influence the accuracy of the resulting Tn-seq profile, leading to false positive or false negative results. There are four theoretical or practical considerations for an ideal Tn-seq method as discussed in detail in the following sections. The potential pros and cons of the currently existing Tn-seq methods based on these criteria are listed in Table 1.

Potential bias in Tn-seq library preparation. The most critical requirement for Tn-seq is minimum bias during Tn-seq library preparation. Theoretically, this type of bias can occur during the preparation of the PCR template or PCR amplification. In the "Tn-seq circle" method, the physically sheared genomic DNA fragments are ligated to an adaptor, digested by restriction enzyme, denatured, and circularized through oligonucleotide-mediated ligation (Gallagher et al. 2011). Therefore, the variable lengths of the fragments can cause bias in the ligation reaction and the efficiency of ligation itself would be critical in preparing a template library well representative of the transposon mutant pool. However, Gallagher et al. (2015) recently compared Tn-seq cirle method (Gallagher et al. 2011) with the Tn-seq method based on C-tailing (Klein et al. 2012) by analyzing the same genomic DNA from a complex transposon library of *Acinetobacter baumannii* with the two Tn-seq methods (Gallagher et al. 2015). These two methods provided remarkably similar lists of essential genes, suggesting both methods are robust and the potential bias, if existed, was insignificant. Bias in the final Tn-seq library could also happen during the PCR amplification step due to variable lengths of the PCR products being amplified. In this aspect, only the methods based on the use of the Type IIS restriction enzymes (restriction enzymes that cleave outside of their recognition sequence to one side) and RB-TnSeq (Goodman et al. 2009; Khatiwara et al. 2012; van Opijnen et al. 2009; Wetmore et al. 2015) can avoid this issue, since all other methods produce PCR products of variable lengths. The Tn-seq

method based on nested arbitrary PCR raises concerns for additional bias in PCR amplification due to the nature of primer binding occurring at lower annealing temperatures (Christen et al. 2011). With nested arbitrary PCR, the amplification efficiency could largely be dependent on the nucleotide sequences of the transposon-flanking regions. It is expected that a certain portion of insertions may not allow amplification of transposon-junction sequences at all. However, in this particular research a highly saturating Tn5 library of *Caulobacter crescentus* was used, focusing only on essential gene discovery (Christen et al. 2011). Since gene essentiality can be assessed only with the information on insertion sites without relying on quantitative information on each insertion mutant in the library (Hutchison et al. 1999), potential bias in Tn-seq library preparation may not have been a major obstacle in essential gene discovery (Christen et al. 2011).

Quantities of genomic DNA. Many Tn-seq methods involve physical shearing during the preparing the Tn-seq amplicon library. Physical shearing of genomic or metagenomics DNA is a step commonly used to prepare a DNA fragment library for NGS analysis (Knierim et al. 2011). The random nature of physical shearing makes it an attractive choice because it helps to generate a bias-free fragment library. Although effective, it often requires an optimization step, a relatively large quantity of starting DNA materials, and equipment (e.g. sonicator) to perform this step. For the Tn-seq methods that involve physical shearing, the amount of starting DNA (per sample) ranged from 3 to 6 μg (Gallagher et al. 2011; Langridge et al. 2009; Wong et al. 2011). On the contrary, the methods that begin with the PCR to amplify or extend transposon-junctions directly from the template DNA require much less amount of starting DNA. In the method based on nested arbitrary PCR (Christen et al. 2011), 1 μl of a bacterial culture (OD 0.1) was directly used as a template, and our lab routinely use 50-100 ng of genomic DNA as a

template for the Tn-seq method based on linear PCR followed by C-tailing and PCR (Dawoud et al. 2014; unpublished).

For most Tn-seq applications, collecting a large quantity of bacterial cells representing the entire mutant population is not an issue. However, in certain circumstances where the surviving mutants are recovered from infected host tissues to form a recovered mutant pool, the number of bacterial cell survivors (thus their genomic DNA) can be a limiting factor for performing a physical shearing step, especially when the procedure should be repeated for optimization or due to a mistake. The recovered library can be amplified by bacterial cultivation, but this step may introduce artifacts resulting from differences in mutant *in vitro* growth rates.

Applicability to any transposon elements. Most Tn-seq methods are universally applicable to a mutant library constructed by any type of transposon elements. However, the Tn-seq methods that utilize Type IIS restriction enzymes (*Mme*I or *Bsm*FI) require the presence of the restriction sites at the end(s) of the transposon (Goodman et al. 2009; Khatiwara et al. 2012; van Opijnen et al. 2009). This requirement limits this type of Tn-seq methods only to certain transposons. For example, a *Mme*I site could only be created in the *mariner* transposon that happened to carry sequence in the inverted repeat region that closely matched the *Mme*I site except for one nucleotide (Goodman et al. 2009; van Opijnen et al. 2009). Therefore, these methods cannot be applied to any other transposon elements and thus is not applicable to an existing transposon library that is constructed based on wild type *mariner* transposon or other transposon elements. In the case of RB-TnSeq, the use of barcode regions located within the transposon instead of transposon-junction sequences for quantitative profiling of transposon mutants provides multiple advantages (Wetmore et al. 2015). However, it also requires the use of a modified transposon carrying random barcodes within the transposon for a library construction.

Precise genome mapping. The length of transposon-junction sequence reads should be sufficiently long to allow precise genome mapping of the reads and thus precise determination of the insertion sites. For most Tn-seq methods, the length of transposon-junction reads can be adjusted by choosing and purifying an appropriate range of PCR products. The lengths of the Tn-seq amplicons are uniformly fixed to a relatively short length only for the methods based on the use of Type IIS restriction enzymes (Goodman et al. 2009; Khatiwara et al. 2012; van Opijnen et al. 2009). The question then becomes how long should the transposon-junction sequences be to serve this purpose? This can be estimated. For example, based on computer simulation analysis, minimum lengths of 16 bp would be required for unambiguous genome mapping for 98% of the reads when the genome of *Bacteroides thetaiotaomicron* was used for the test (Goodman et al. 2009). This fact suggests that the majority of the reads from Tn-seq method based on *Mme*I, which produces 16 bp sequence reads, would be sufficient for precise genome mapping (Goodman et al. 2009; van Opijnen et al. 2009). However, Tn-seq method based on the use of *BsmF*I restriction enzyme suffers from short reads of 11 to 12 bp, for which approximately 50% of the reads would have to be discarded due to the inability to achieve unambiguous genome mapping (Khatiwara et al. 2012).

Applications of Tn-seq methods

With continuously increasing read numbers for Illumina sequencing (which currently provides approximately 3.0×10^8 reads per lane on HiSeq2500), Tn-seq analysis provides extraordinary opportunities for gene discovery at an accelerated rate to address various biological questions that were impossible to answer in the past before the development of Tn-seq methods. We have highlighted some of the interesting trends in Tn-seq applications in the following sections.

Essential genes. One of the first applications of global transposon mapping data was to discover essential genes of the bacterium *Mycoplasma genitalium* (Hutchison et al. 1999). An essential gene is defined as the gene that is essential for growth or survival under the optimal growth condition. Therefore an essential gene set would be expected to change depending on how the optimal condition was initially defined. Conventionally, however, essential genes refer to the genes required for growth or survival of a bacterium in the standard rich media commonly used for routine culture of the bacterial species. When global transposon mapping data became available, essential genes could be identified conceptually by the genomic regions that contain no or very few transposons. By the subtractive nature of the approach for essential gene discovery, the accuracy of prediction would be further enhanced by higher level of genome saturation via transposon insertions. With the Tn-seq method, much higher levels of genome saturation can be accomplished, and therefore Tn-seq data obtained from various bacterial species under standard growth media have provided high quality data for essential gene discovery. The complete set of essential genes, termed "essential genome" has been defined by Tn-seq data for numerous bacterial species, including *Burkholderia pseudomallei* (Moule et al. 2014), *Campylobacter jejuni* (Gao et al. 2014), *Pseudomonas aeruginosa* (Lee et al. 2015), and *Streptococcus pyogenes* (Le Breton et al. 2015), and one archaeal species, *Methanococcus maripaludis* (Sarmiento et al. 2013).

In most studies on essential gene discovery using Tn-seq, viable transposon mutants are usually recovered from a single nutrient-rich condition and the resulting Tn-seq data is used to identify essential genes. However, in a more recent study, Lee et al. (2015) studied the essential genes in *Pseudomonas aeruginosa* in six different media, and identified 352 general and 199 condition-specific essential genes. This approach allows discernment of "essential genes" specific to

different growth conditions from truly essential genes, and to define core essential genes that are required for viability under multiple growth conditions.

Conditionally essential genes (in vitro conditions). Conditionally essential genes should be considered as an extension of the essential genes in the sense that conditionally essential genes are required for growth or survival only under the condition of the interest other than the standard media. Of special interest, for example, would be defining a bacterial gene set conditionally required for growth or survival during specific environmental niches related to the life cycle of the species. For a bacterial pathogen, it would be particularly important to understand which genes are essential to overcome the stressors or immune defenses in the host. By comparing the genetic requirements for growth under different *in vitro* conditions with the genes required for *in vivo* growth or survival in the host, the unknown selective pressures that bacterial pathogens encounter in specific host niches can be identified (Khatiwara et al. 2012; Merrell and Camilli 2002; van Opijnen and Camilli 2012). More recently, the potential link between the metabolic capacity of a pathogen and its virulence has been suggested as a critical factor for expression of pathogenic phenotypes (Rohmer et al. 2011). Metabolic genes that enable a pathogen to utilize a nutrient uniquely present in a host niche would play an important role during infection in the host tissues. For example, Griffin *et al*. (2011) used Tn-seq to define a set of genes in *Mycobacterium tuberculosis* that are required for *in vitro* utilization of cholesterol as a sole carbon source. Comparison of the result of this study with previously identified genes in *M. tuberculosis* required for *in vivo* survival during mouse infections (Sassetti and Rubin 2003) demonstrated that 10% of the genes specifically required for bacterial growth *in vivo* are also required for the utilization of cholesterol *in vitro*. Until recently, a large portion of the Tn-seq studies for gene discovery have focused on screening and characterizing conditionally

essential genes under *in vitro* conditions, largely due to the simplicity of the experimental design, the lack of problems associated with bottlenecks that would occur in animal infection studies, and the resulting in-depth insights that can be gained from the comprehensive sets of genes identified. An extensive list of Tn-seq studies on conditionally essential genes is shown in Table 2.

Genes required for in vivo fitness in the host. Genome-wide identification of bacterial virulence genes required for *in vivo* fitness during host infection using Tn-seq is an extremely valuable approach in understanding complex mechanisms of virulence. Such applications of Tn-seq for virulence gene discovery using various pathogen-host infection models has been steadily increasing over the years, leading to identification of numerous previously known as well as unknown virulence factors (Table 2). In some animal infection models, however, this approach involving a complex library is not feasible due to the bottlenecks that cause stochastic removal of bacterial cells during establishment of infection (van Opijnen and Camilli 2013). In such cases, multiple transposon libraries of smaller sizes can be used to identify *in vivo* fitness factors (Chaudhuri et al. 2013). One emerging research area of interest is the comparative analysis of *in vivo* fitness factors in multiple hosts. Chaudhuri et al. (2013) screened the same collection of *Salmonella enterica* Typhimurium Tn5 mutants for mutants with reduced gut colonization in 3 different hosts, chickens, pigs, and calves, and identified a core set of virulence genes as well as host-specific virulence factors. More recently, Weerdenburg et al. (2015) used a similar comparative analysis to identify the factors of a broad-host-range pathogen *Mycobacterium marinum* that are important for survival in phagocytic cells of five different host species. Finally, Tn-seq method has also been applied to understand the genetic mechanisms associated with *Vibrio fischeri* symbiotic colonization of the light organ of squid (Brooks et al. 2014).

Small RNA genes. Initially the focus of Tn-seq application was to identify protein-coding genes important for fitness. Although some sRNA (e.g. GlmZ in *Escherichia coli*) was identified by a phenotypic screening of transposon mutants (Kalamorz et al. 2007), the general applicability of Tn-seq for the comprehensive discovery of sRNA genes remained uncertain. However, several Tn-seq studies have demonstrated that Tn-seq can actually be very effective in identifying conditionally essential sRNA genes (Barquist et al. 2013b; Christen et al. 2011; Mann et al. 2012; Zhang et al. 2012). The challenge in analyzing sRNA genes lies simply in the fact that sRNA genes are much smaller than protein-coding genes, thereby reducing the chance to be hit by a transposon, and sRNA knockout mutants usually do not exhibit strong phenotypes (Sharma and Vogel 2009). The utility of Tn-seq in identification of sRNA genes is very significant for high-throughput analysis of sRNA genes in bacteria. Previously, the main approach for sRNA discovery was through detection of sRNA transcripts either by transcriptome analysis (microarray or RNA-seq) or cloning of reverse transcribed RNA transcripts (RNomics)(Sharma and Vogel 2009). Alternatively, RNA chaperone Hfq protein was used as a bait to capture sRNA transcripts associated Hfq protein (Sharma and Vogel 2009). However, these approaches seldom reveal any information regarding biological functions of the sRNAs. To understand sRNA functions, it requires a time-consuming downstream analysis for individual sRNAs (Sharma and Vogel 2009). However, Tn-seq analysis provides the means to identify sRNAs genes comprehensively and also valuable insights on the sRNA functions. Two studies using Tn-seq have identified sRNA genes required for growth in rich media for *Caulobacter crescentus* (Christen et al. 2011) and *Mycobacterium tuberculosis* (Zhang et al. 2012). More interestingly, Tn-seq analysis has been used in mouse models of colonization to reveal the contribution of the sRNAs in *Streptococcus pneumoniae* to fitness *in vivo* (Mann et al. 2012). Collectively, these

studies demonstrate the general utility of Tn-seq for global discovery of essential as well as conditionally essential sRNA genes.

Additional genetic elements and features: promoters, operons, and domains. For most Tn-seq analysis, the focus of the research is usually on identification of essential genes or conditionally essential genes, whether it is protein-coding genes or noncoding sRNA genes. Usually, no additional attempts have been made to find more information beyond genetic requirements of the genes. However, Christen *et al*. (2011) demonstrated that a small change in the design of the transposon itself can provide additional in-depth information beyond genetic requirements. In their study, a Tn5 transposon containing a strong xylose-inducible promoter facing outward was used to construct a genome-saturation library of *Caulobacter crescentus*. When the mutants were recovered in the presence of the inducer and analyzed for insertions by Tn-seq, the comparative analysis of the insertion groups in two different orientations allowed identification of the promoter regions of essential genes, the operons with essential functions, and domains accountable for essentiality of the corresponding genes (Christen et al. 2011).

Genetic interaction mapping. One powerful approach to tackle functional organization of the genes related to a certain phenotype is to examine genetic interactions among multiple gene products (Dixon et al. 2009). In general, when a double mutant that shows a significant deviation in fitness compared with the expected multiplicative effect of combining two single mutants, it is considered a genetic interaction. Negative genetic interactions refer to a more severe defect in fitness than expected. In extreme cases where the cell is not viable due to mutations in two non-essential genes, it is regarded as a synthetic lethality. Positive interactions refer to double mutants with a less severe fitness defect than expected. Genetic interaction networks can reveal unexpected functional dependencies between genetic loci (i.e. epistasis, wherein the phenotypic

effects of mutation in one gene are modified by one or more other genes). For example, negative genetic interactions often result from loss-of-function mutations in pairs of genes in parallel or compensatory pathways that impinge on a common essential process. Conversely, positive interactions can occur between genes in the same pathway if the loss of one gene alone inactivates the pathway such that loss of a second gene confers no additional defect. Genetic interaction networks can be explored by performing Tn-seq analysis in a wild type strain and its mutant strain counterpart with deletion in a gene of interest (query gene), and comparing the resulting profiles (van Opijnen et al. 2009). This approach was applied to determine genetic interactions in *Streptococcus pneumoniae* (van Opijnen et al. 2009; van Opijnen and Camilli 2012). Similar genetic interaction mapping based on the use of a microarray-based transposon tracking method was employed previously to uncover genetic interactions important for *in vivo* fitness of *Mycobacterium tuberculosis* during infection in mice (Joshi et al. 2006) and for motility of *E. coli* (Girgis et al. 2007) with a focus on a selected set of query genes.

In a more recent study to understand how *trans*-translation by tmRNA encoded by *ssrA* is dispensable in *Caulobacter crescentus*, Feaga et al. (2014) used Tn-seq method to identify gene(s) that are synthetically lethal with *ssrA* gene deletion by performing Tn-seq with *himar1* transposon libraries in wild type and Δ*ssrA* backgrounds, and found that ArfB is a functional homolog of tmRNA that can also release nonstop ribosomes. Genetic interaction mapping can also be performed using an inhibitor that blocks a specific pathway. In a study by Santa Maria et al. (2014) a natural product tunicamycin was used to selectively inhibit TarO, the first enzyme in the wall teichoic acid (WTA) pathway of *Staphylococcus aureus*. They selected a *mariner* transposon library in the presence and absence of tunicamycin, and the resulting Tn-seq profiles were compared to identify genes that affect survival in the presence of tunicamycin, thus

implicating their products in WTA-related activities (Santa Maria et al. 2014). These studies illustrate that Tn-seq has general applicability in mapping genetic interactions for diverse bacterial species (Table 2).

Novel genetic factors involved in specific biological processes. One interesting aspect for Tn-seq application is the development of novel screening strategies that allow genome-wide identification of genetic factors involved in specific biological processes of significance for the bacterial species. These applications require fairly sophisticated experimental designs and optimized experimental conditions to identify the target genes precisely. Some of the examples include identification of genetic factors responsible for (1) Vi capsule expression in *Salmonella enterica* Typhi (Pickard et al. 2013), (2) immunity against killing by Type VI protein secretion system (T6SS) in *Vibrio cholerae* (Dong et al. 2013), and (3) *in vivo*-specific induction of *xds* gene encoding a secreted exonuclease in *Vibrio cholerae* (McDonough et al. 2014). For some bacterial pathogens, the genetic factors required to proceed through specific stages in host-pathogen interactions, such as adhesion (de Vries et al. 2013) or invasion to host cells (Gao et al. 2014) have been identified using Tn-seq. More examples of these types of studies are shown in Table 2.

Discovery of adaptive mutations. Before the development of Tn-seq methods, microarray-based footprinting methods were developed to quantitatively track mutants in a complex transposon library (Sassetti and Rubin 2002). One of the unique applications of the microarray-based method was to identify adaptive mutations that contribute to selectable phenotypic variations (Goodarzi et al. 2009). With current NGS technologies, genome sequencing of the wild type and evolved strains can be easily used to reveal genetic differences (e.g. mutations) between two strains. However, distinguishing adaptive mutations from neutral mutations is a challenging and

labor-intensive process. Goodarzi *et al*. (2009) described a method, termed ADAM for Array-based discovery of adaptive mutations, that employs parallel, genome-wide linkage analysis to simultaneously identify all mutated loci with direct contributions to fitness. Although it has not been realized yet, it is quite conceivable that the Tn-seq method could be used in place of a microarray-based transposon mapping approach to advance strategies currently used in ADAM to identify adaptive mutations at a higher resolution.

Conclusions and Perspectives

With the comprehensiveness and sensitivity of Tn-seq, it has emerged as a method of choice to explore genotype-phenotype relationships of a bacterial genome on a genomic scale. Since initial developments of the method in 2009, several variations on Tn-seq have been described with ever increasing applications in numerous bacterial and archaeal species where an efficient random transposon mutagenesis system can be established. The major driving force behind the development of Tn-seq was NGS technologies, more specifically Illumina sequencing. The ability to sequence hundreds of millions fragments in parallel is the crucial component that provides the comprehensiveness and sensitivity characteristic of Tn-seq methods. With continuous improvements on current NGS platforms (especially increasing read numbers of Illumina sequencing technology), it is expected that the ability to sequence more reads at a reduced cost will occur in the near future (Watson 2014). This will be an advantage in further enhancing the capacity of Tn-seq methods by increasing (1) the number of samples to be analyzed, (2) read depth, or (3) the saturation levels of an insertion library.

Although over 70 research articles based on Tn-seq methods have been published within the past six years, there are still immense chemical or stress conditions encountered by microorganisms

that remain to be explored using straightforward applications of Tn-seq under the corresponding *in vitro* conditions. For example, the 1,144 chemical genomic assays have been performed with the collection of yeast deletion mutants (Hillenmeyer et al. 2008). There has been an increasing number of Tn-seq studies using animal infection models to identify *in vivo* survival genes. Once such a study is done for a given pathogen using a standard animal infection model under "standard" conditions, the next logical step would be to use Tn-seq to understand bacterial genes required for *in vivo* colonization or survival of the pathogen that would be dependent on the altered host conditions. The altered host conditions could be contributed by genetic factors (e.g., different strains of mice, or transgenic animals) or environmental factors (e.g., modified gut microbiota, co-infection, diets, age, stress, gender etc.) as exemplified in several studies shown in Table 2 (Carter et al. 2014; Goodman et al. 2009; Wong et al. 2013; Zhang et al. 2013). Knowing the genetic factors required only under specific host or environmental conditions would be extremely helpful in revealing the mechanisms by which the pathogens cope with the dynamically changing microenvironments in the host. Until now most Tn-seq studies have been conducted to study bacterial species, and only one study was reported in which archaeal species was studied using a Tn-seq method (Sarmiento et al. 2013). Since any haploid microorganisms with an appropriate insertional mutagenesis system can be analyzed by Tn-seq, the method could be applied to the study of more archaeal species and even haploid yeast strains in the future.

Conflict of Interest

The authors declare that they have no conflict of interest.

List of Tables

Table 1. Strategies to amplify transposon-junction sequences in Tn-seq different methods

Amplification strategy	Specific transposon?	Physical shearing	Amplicon Length	Other pros & cons	Reference*
Type IIS restriction enzyme → Adapter → PCR	Yes	No	Uniform → PCR bias ↓	Short reads → ambiguous mapping ↑	(Goodman et al. 2009; Khatiwara et al. 2012; van Opijnen et al. 2009)
Shearing → Adapter → PCR	No	Yes	Variable → PCR bias ↑	Large quantity of DNA	(Gawronski et al. 2009; Langridge et al. 2009)
Nested arbitrary PCR → nested PCR	No	No		Potential target bias	(Christen et al. 2011)
Shearing → Adapter → Restriction enzyme → circularization → PCR	No	Yes		Large quantity of DNA Potential target bias Non-specific background ↓	(Gallagher et al. 2011)
Shearing → C-tailing → PCR	No	Yes		Large quantity of DNA	(Klein et al. 2012)
Single primer extension → C-tailing → PCR	No	No			(Dawoud et al. 2014)
Simple PCR of barcode regions with 2 universal primers	Yes	No	Uniform → PCR bias ↓	Construction of transposon with random barcodes Simple PCR protocol Throughput ↑	(Wetmore et al. 2015)

* Only representative references are shown.

Table 2. Applications of Tn-seq.

Target categories	Microorganisms	Selection conditions	References
Conditionally essential genes under *in vitro* conditions	*Salmonella enterica* Typhi	Bile resistance	(Langridge et al. 2009)
	Mycobacterium tuberculosis	Growth on cholesterol as a carbon source	(Griffin et al. 2011)
	Pseudomonas aeruginosa	Intrinsic resistance to tobramycin	(Gallagher et al. 2011)
	Shewanella oneidensis	Anaerobic minimal broth medium	(Brutinel and Gralnick 2012)
	Salmonella enterica Typhimurium	Growth in 42ºC, bile salts, and limited nutrients	(Khatiwara et al. 2012)
	Streptococcus pneumoniae	17 *in vitro* conditions representing *in vivo* selective pressures	(van Opijnen and Camilli 2012)
	Escherichia coli ST131	Serum resistome (genes required for resistance to human serum)	(Phan et al. 2013)
	Haemophilus influenza	Fitness in ambient air	(Langereis et al. 2013)
	Streptococcus pneumoniae	Growth under CO_2-poor condition	(Burghout et al. 2013)
	Sphingomonas wittichii	Fitness in sand with salicylate as a carbon source	(Roggo et al. 2013)
	Staphylococcus aureus ST398	Whole porcine blood	(Christiansen et al. 2014)
	Escherichia coli	Resistance to ionizing radiation	(Byrne et al. 2014)
	Moraxella catarrhalis	Serum resistome (genes required for resistance to human serum)	(de Vries et al. 2014)
	Haemophillus influenza	Resistance to neutrophils and serum	(Langereis and Weiser 2014)
	Staphylococcus aureus	Growth in blood and ocular fluids	(Valentino et al. 2014)
	Pseudomonas aeruginosa	Cystic fibrosis sputum	(Turner et al. 2015)
	Bacteroides thetaiotaomicron	Fitness in the presence of polymyxin B	(Cullen et al. 2015)
	Staphylococcus auresus	Fitness at different temperatures (16, 23, 30, 37, and 42ºC)	(Santiago et al. 2015)
	Escherichia coli	Fitness in gentamicin during stationary-phase	(Shan et al. 2015)
	Caulobacter crescentus	Genetic factors required for uranium resistance	(Yung et al. 2015)
Genes required for *in vivo* fitness in the host	*Haemophilus influenza*	Murine pulmonary model	(Gawronski et al. 2009)
	Escherichia coli O157:H7	Colonization in the intestinal tract of calves	(Eckert et al. 2011)
	Streptococcus pneumoniae	Nasopharynx colonization, and lung infection models of mice	(van Opijnen and Camilli 2012)
	Salmonella enterica Typhimurium	Gut colonization in chickens, pigs, and calve vs	(Chaudhuri et al. 2013)
	Pseudomonas aeruginosa	Gut colonization, and systemic infection of mice	(Skurnik et al. 2013)
	Vibrio cholerae	Colonization of the infant rabbit intestine	(Kamp et al. 2013)
	Vibrio cholerae	Colonization of the infant rabbit intestine	(Fu et al. 2013; Subashchandrabose et al. 2013)
	Uropathogenic *E. coli* (UPEC)	Murine model of systemic infection	
	Yersinia pestis	Systemic infection in mice	(Palace et al. 2014)
	Acinetobacter baumannii	Persistent infection in the mouse lung	(Wang et al. 2014)
	Vibrio fischeri	Symbiotic colonization in the light organ of squid	(Brooks et al. 2014)
	Campylobacter jejuni	Cecal colonization in chicks	(Johnson et al. 2014)
	Pseudomonas aeruginosa	Murine models of acute vs. chronic wound infection	(Turner et al. 2015)
	Staphylococcus aureus	Murine abscess model	(Valentino et al. 2014)
	Mycobacterium marinum	Intracellular survival in phagocytic cells of different origins (human, mouse, fish, and protozoa)	(Weerdenburg et al. 2015)
	Klebsiella pneumoniae	Mouse model of pneumonia	(Bachman et al. 2015)
	Pseudomonas aeruginosa	Mouse model of lung infection	(Roux et al. 2015)
	Acinetobacter baumannii	Murine infection model	
	Vibrio cholerae	Rabbit infection model	

Table 2. (Cont.) Applications of Tn-seq.

Target categories	Microorganisms	Selection conditions	References
Genes required for *in vivo* fitness in altered host environment	*Bacteroides thetaiotaomicron*	Germ-free mice with WT, *Rag1*-/- or *Myd88*-/- genetic backgrounds; Germ-free WT mice with 3 different simple bacterial communities	(Goodman et al. 2009)
	Mycobacterium tuberculosis	WT vs. CD4-deficient mice	(Zhang et al. 2013)
	Haemophilus influenza	Murine model of coinfection with influenza A virus	(Wong et al. 2013)
	Streptococcus pneumoniae	WT vs. SCD (sickle cell disease) mice	(Carter et al. 2014)
Genes for required for dissemination from the host	*Vibrio cholerae*	Genes contributing to dissemination from the rabbits to pond water	(Kamp et al. 2013)
Genetic interaction mapping	*Streptococcus pneumoniae*	Genetic interactions connected to 5 query genes	(van Opijnen et al. 2009)
	Streptococcus pneumoniae	Genetic interactions connected to 1 query gene	(van Opijnen and Camilli 2012)
	Staphylococcus aureus	Cellular factors that interact with wall teichoic acids	(Santa Maria et al. 2014)
	Caulobacter crescentus	Genes synthetically lethal with *ssrA gene* (encoding tmRNA)	(Feaga et al. 2014)
Genetic factors involved in specific biological processes	*Salmonella enterica* Typhi	Genetic factors required for efficient Vi capsule expression	(Pickard et al. 2013)
	Moraxella catarrhalis	Adherence to epithelial cells	(de Vries et al. 2013)
	Vibrio cholerae	Type VI protein secretion system (T6SS) immunity proteins	(Dong et al. 2013)
	Vibrio cholerae	Regulators responsible for *in vivo*-specific induction of *xds* gene	(McDonough et al. 2014)
	Bacillus subtilis	Genes contributing to acquisition of a conjugative element	(Johnson and Grossman 2014)
	Campylobacter jejuni	Invasion to cultured mammalian cells	(Gao et al. 2014)
	Escherichia coli O104:H4	Genes required for maintenance and transmission of ESBL plasmid	(Yamaichi et al. 2015)
	Clostridium difficile	Genes required for sporulation	(Dembek et al. 2015)
	Mycobacterium marinum	Essential genes rendered non-essential upon heterologous expression of outer membrane porin	(Ates et al. 2015)
	Escherichia coli ST131	Genes required for the stable maintenance of IncF plasmid	(Phan et al. 2015)
	Vibrio cholerae	Novel facters required for expression of *tcpA* gene.	(Wang et al. 2015)

References

Ates LS, Ummels R, Commandeur S, van der Weerd R, Sparrius M, Weerdenburg E, Alber M, Kalscheuer R, Piersma SR, Abdallah AM, Abd El Ghany M, Abdel-Haleem AM, Pain A, Jiménez CR, Bitter W, Houben EN (2015) Essential Role of the ESX-5 Secretion System in Outer Membrane Permeability of Pathogenic Mycobacteria. PLoS Genet 11(5):e1005190 doi:10.1371/journal.pgen.1005190

Bachman MA, Breen P, Deornellas V, Mu Q, Zhao L, Wu W, Cavalcoli JD, Mobley HL (2015) Genome-Wide Identification of Klebsiella pneumoniae Fitness Genes during Lung Infection. MBio 6(3):e00775 doi:10.1128/mBio.00775-15

Barquist L, Boinett CJ, Cain AK (2013a) Approaches to querying bacterial genomes with transposon-insertion sequencing. RNA Biol 10(7):1161-9 doi:10.4161/rna.24765

Barquist L, Langridge GC, Turner DJ, Phan MD, Turner AK, Bateman A, Parkhill J, Wain J, Gardner PP (2013b) A comparison of dense transposon insertion libraries in the Salmonella serovars Typhi and Typhimurium. Nucleic Acids Res 41(8):4549-64 doi:10.1093/nar/gkt148

Brooks JF, Gyllborg MC, Cronin DC, Quillin SJ, Mallama CA, Foxall R, Whistler C, Goodman AL, Mandel MJ (2014) Global discovery of colonization determinants in the squid symbiont Vibrio fischeri. Proc Natl Acad Sci U S A 111(48):17284-9 doi:10.1073/pnas.1415957111

Brutinel ED, Gralnick JA (2012) Anomalies of the anaerobic tricarboxylic acid cycle in Shewanella oneidensis revealed by Tn-seq. Mol Microbiol 86(2):273-83 doi:10.1111/j.1365-2958.2012.08196.x

Burghout P, Zomer A, van der Gaast-de Jongh CE, Janssen-Megens EM, Françoijs KJ, Stunnenberg HG, Hermans PW (2013) Streptococcus pneumoniae folate biosynthesis responds to environmental CO2 levels. J Bacteriol 195(7):1573-82 doi:10.1128/JB.01942-12

Byrne RT, Chen SH, Wood EA, Cabot EL, Cox MM (2014) Escherichia coli genes and pathways involved in surviving extreme exposure to ionizing radiation. J Bacteriol 196(20):3534-45 doi:10.1128/JB.01589-14

Carter R, Wolf J, van Opijnen T, Muller M, Obert C, Burnham C, Mann B, Li Y, Hayden RT, Pestina T, Persons D, Camilli A, Flynn PM, Tuomanen EI, Rosch JW (2014) Genomic analyses of pneumococci from children with sickle cell disease expose host-specific bacterial adaptations and deficits in current interventions. Cell Host Microbe 15(5):587-99 doi:10.1016/j.chom.2014.04.005

Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang J, van Diemen PM, Buckley AM, Bowen AJ, Pullinger GD, Turner DJ, Langridge GC, Turner AK, Parkhill J, Charles IG, Maskell DJ, Stevens MP (2013) Comprehensive assignment of roles for Salmonella typhimurium genes in intestinal colonization of food-producing animals. PLoS Genet 9(4):e1003456 doi:10.1371/journal.pgen.1003456

Christen B, Abeliuk E, Collier JM, Kalogeraki VS, Passarelli B, Coller JA, Fero MJ, McAdams HH, Shapiro L (2011) The essential genome of a bacterium. Mol Syst Biol 7:528 doi:10.1038/msb.2011.58

Christiansen MT, Kaas RS, Chaudhuri RR, Holmes MA, Hasman H, Aarestrup FM (2014) Genome-wide high-throughput screening to investigate essential genes involved in methicillin-resistant Staphylococcus aureus Sequence Type 398 survival. PLoS One 9(2):e89018 doi:10.1371/journal.pone.0089018

Cullen TW, Schofield WB, Barry NA, Putnam EE, Rundell EA, Trent MS, Degnan PH, Booth CJ, Yu H, Goodman AL (2015) Gut microbiota. Antimicrobial peptide resistance mediates resilience of prominent gut commensals during inflammation. Science 347(6218):170-5 doi:10.1126/science.1260580

Dawoud TM, Jiang T, Mandal RK, Ricke SC, Kwon YM (2014) Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers. Mol Biotechnol 56(11):1004-10 doi:10.1007/s12033-014-9779-4

de Lorenzo V, Timmis KN (1994) Analysis and construction of stable phenotypes in gram-negative bacteria with Tn5- and Tn10-derived minitransposons. Methods Enzymol 235:386-405

de Vries SP, Eleveld MJ, Hermans PW, Bootsma HJ (2013) Characterization of the molecular interplay between Moraxella catarrhalis and human respiratory tract epithelial cells. PLoS One 8(8):e72193 doi:10.1371/journal.pone.0072193

de Vries SP, Rademakers RJ, van der Gaast-de Jongh CE, Eleveld MJ, Hermans PW, Bootsma HJ (2014) Deciphering the genetic basis of Moraxella catarrhalis complement resistance: a critical role for the disulphide bond formation system. Mol Microbiol 91(3):522-37 doi:10.1111/mmi.12475

Dembek M, Barquist L, Boinett CJ, Cain AK, Mayho M, Lawley TD, Fairweather NF, Fagan RP (2015) High-throughput analysis of gene essentiality and sporulation in Clostridium difficile. MBio 6(2):e02383 doi:10.1128/mBio.02383-14

Dixon SJ, Costanzo M, Baryshnikova A, Andrews B, Boone C (2009) Systematic mapping of genetic interaction networks. Annu Rev Genet 43:601-25 doi:10.1146/annurev.genet.39.073003.114751

Dong TG, Ho BT, Yoder-Himes DR, Mekalanos JJ (2013) Identification of T6SS-dependent effector and immunity proteins by Tn-seq in Vibrio cholerae. Proc Natl Acad Sci U S A 110(7):2623-8 doi:10.1073/pnas.1222783110

Eckert SE, Dziva F, Chaudhuri RR, Langridge GC, Turner DJ, Pickard DJ, Maskell DJ, Thomson NR, Stevens MP (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(7):1771-6 doi:10.1128/JB.01292-10

Feaga HA, Viollier PH, Keiler KC (2014) Release of nonstop ribosomes is essential. MBio 5(6):e01916 doi:10.1128/mBio.01916-14

Febrer M, McLay K, Caccamo M, Twomey KB, Ryan RP (2011) Advances in bacterial transcriptome and transposon insertion-site profiling using second-generation sequencing. Trends Biotechnol 29(11):586-94 doi:10.1016/j.tibtech.2011.06.004

Fu Y, Waldor MK, Mekalanos JJ (2013) Tn-Seq analysis of Vibrio cholerae intestinal colonization reveals a role for T6SS-mediated antibacterial activity in the host. Cell Host Microbe 14(6):652-63 doi:10.1016/j.chom.2013.11.001

Gallagher LA, Ramage E, Weiss EJ, Radey M, Hayden HS, Held KG, Huse HK, Zurawski DV, Brittnacher MJ, Manoil C (2015) Resources for genetic and genomic analysis of emerging pathogen Acinetobacter baumannii. J Bacteriol doi:10.1128/JB.00131-15

Gallagher LA, Shendure J, Manoil C (2011) Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq. MBio 2(1):e00315-10 doi:10.1128/mBio.00315-10

Gao B, Lara-Tejero M, Lefebre M, Goodman AL, Galán JE (2014) Novel components of the flagellar system in epsilonproteobacteria. MBio 5(3):e01349-14 doi:10.1128/mBio.01349-14

Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ (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 U S A 106(38):16422-7 doi:10.1073/pnas.0906627106

Girgis HS, Liu Y, Ryu WS, Tavazoie S (2007) A comprehensive genetic characterization of bacterial motility. PLoS Genet 3(9):1644-60 doi:10.1371/journal.pgen.0030154

Goodarzi H, Hottes AK, Tavazoie S (2009) Global discovery of adaptive mutations. Nat Methods 6(8):581-3 doi:10.1038/nmeth.1352

Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI (2009) Identifying genetic determinants needed to establish a human gut symbiont in its habitat. Cell Host Microbe 6(3):279-89 doi:10.1016/j.chom.2009.08.003

Goryshin IY, Jendrisak J, Hoffman LM, Meis R, Reznikoff WS (2000) Insertional transposon mutagenesis by electroporation of released Tn5 transposition complexes. Nat Biotechnol 18(1):97-100 doi:10.1038/72017

Griffin JE, Gawronski JD, Dejesus MA, Ioerger TR, Akerley BJ, Sassetti CM (2011) High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. PLoS Pathog 7(9):e1002251 doi:10.1371/journal.ppat.1002251

Hayes F (2003) Transposon-based strategies for microbial functional genomics and proteomics. Annu Rev Genet 37:3-29 doi:10.1146/annurev.genet.37.110801.142807

Hendrixson DR, Akerley BJ, DiRita VJ (2001) Transposon mutagenesis of Campylobacter jejuni identifies a bipartite energy taxis system required for motility. Mol Microbiol 40(1):214-24

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St Onge RP, Tyers M, Koller D, Altman RB, Davis RW, Nislow C, Giaever G (2008) The chemical

genomic portrait of yeast: uncovering a phenotype for all genes. Science 320(5874):362-5 doi:320/5874/362 [pii]10.1126/science.1150021

Hutchison CA, Peterson SN, Gill SR, Cline RT, White O, Fraser CM, Smith HO, Venter JC (1999) Global transposon mutagenesis and a minimal Mycoplasma genome. Science 286(5447):2165-9 doi:8071 [pii]

Johnson CM, Grossman AD (2014) Identification of host genes that affect acquisition of an integrative and conjugative element in Bacillus subtilis. Mol Microbiol 93(6):1284-301 doi:10.1111/mmi.12736

Johnson JG, Livny J, Dirita VJ (2014) High-throughput sequencing of Campylobacter jejuni insertion mutant libraries reveals mapA as a fitness factor for chicken colonization. J Bacteriol 196(11):1958-67 doi:10.1128/JB.01395-13

Joshi SM, Pandey AK, Capite N, Fortune SM, Rubin EJ, Sassetti CM (2006) Characterization of mycobacterial virulence genes through genetic interaction mapping. Proc Natl Acad Sci U S A 103(31):11760-5 doi:0603179103 [pii]10.1073/pnas.0603179103

Kalamorz F, Reichenbach B, März W, Rak B, Görke B (2007) Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol Microbiol 65(6):1518-33 doi:10.1111/j.1365-2958.2007.05888.x

Kamp HD, Patimalla-Dipali B, Lazinski DW, Wallace-Gadsden F, Camilli A (2013) Gene fitness landscapes of Vibrio cholerae at important stages of its life cycle. PLoS Pathog 9(12):e1003800 doi:10.1371/journal.ppat.1003800

Khatiwara A, Jiang T, Sung SS, Dawoud T, Kim JN, Bhattacharya D, Kim HB, Ricke SC, Kwon YM (2012) Genome scanning for conditionally essential genes in Salmonella enterica Serotype Typhimurium. Appl Environ Microbiol 78(9):3098-107 doi:10.1128/AEM.06865-11

Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LT (2012) Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. BMC Genomics 13:578 doi:10.1186/1471-2164-13-578

Knierim E, Lucke B, Schwarz JM, Schuelke M, Seelow D (2011) Systematic comparison of three methods for fragmentation of long-range PCR products for next generation sequencing. PLoS One 6(11):e28240 doi:10.1371/journal.pone.0028240

Langereis JD, Weiser JN (2014) Shielding of a lipooligosaccharide IgM epitope allows evasion of neutrophil-mediated killing of an invasive strain of nontypeable Haemophilus influenzae. MBio 5(4):e01478-14 doi:10.1128/mBio.01478-14

Langereis JD, Zomer A, Stunnenberg HG, Burghout P, Hermans PW (2013) Nontypeable Haemophilus influenzae carbonic anhydrase is important for environmental and intracellular survival. J Bacteriol 195(12):2737-46 doi:10.1128/JB.01870-12

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK (2009) Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. Genome Res 19(12):2308-16 doi:10.1101/gr.097097.109

Lazinski DW, Camilli A (2013) Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. Biotechniques 54(1):25-34 doi:10.2144/000113981

Le Breton Y, Belew AT, Valdes KM, Islam E, Curry P, Tettelin H, Shirtliff ME, El-Sayed NM, McIver KS (2015) Essential Genes in the Core Genome of the Human Pathogen Streptococcus pyogenes. Sci Rep 5:9838 doi:10.1038/srep09838

Lee SA, Gallagher LA, Thongdee M, Staudinger BJ, Lippman S, Singh PK, Manoil C (2015) General and condition-specific essential functions of Pseudomonas aeruginosa. Proc Natl Acad Sci U S A 112(16):5189-94 doi:10.1073/pnas.1422186112

Maloy SR (2007) Use of antibiotic-resistant transposons for mutagenesis. Methods Enzymol 421:11-7 doi:10.1016/S0076-6879(06)21002-4

Mann B, van Opijnen T, Wang J, Obert C, Wang YD, Carter R, McGoldrick DJ, Ridout G, Camilli A, Tuomanen EI, Rosch JW (2012) Control of virulence by small RNAs in Streptococcus pneumoniae. PLoS Pathog 8(7):e1002788 doi:10.1371/journal.ppat.1002788

Martínez-García E, Calles B, Arévalo-Rodríguez M, de Lorenzo V (2011) pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. BMC Microbiol 11:38 doi:10.1186/1471-2180-11-38

Mazurkiewicz P, Tang CM, Boone C, Holden DW (2006) Signature-tagged mutagenesis: barcoding mutants for genome-wide screens. Nat Rev Genet 7(12):929-39 doi:10.1038/nrg1984

McDonough E, Lazinski DW, Camilli A (2014) Identification of in vivo regulators of the Vibrio cholerae xds gene using a high-throughput genetic selection. Mol Microbiol 92(2):302-15 doi:10.1111/mmi.12557

Merrell DS, Camilli A (2002) Information overload: assigning genetic functionality in the age of genomics and large-scale screening. Trends Microbiol 10(12):571-4 doi:S0966842X02024745 [pii]

Moule MG, Hemsley CM, Seet Q, Guerra-Assunção JA, Lim J, Sarkar-Tyson M, Clark TG, Tan PB, Titball RW, Cuccui J, Wren BW (2014) Genome-wide saturation mutagenesis of Burkholderia pseudomallei K96243 predicts essential genes and novel targets for antimicrobial development. MBio 5(1):e00926-13 doi:10.1128/mBio.00926-13

Palace SG, Proulx MK, Lu S, Baker RE, Goguen JD (2014) Genome-wide mutant fitness profiling identifies nutritional requirements for optimal growth of Yersinia pestis in deep tissue. MBio 5(4) doi:10.1128/mBio.01385-14

Phan MD, Forde BM, Peters KM, Sarkar S, Hancock S, Stanton-Cook M, Ben Zakour NL, Upton M, Beatson SA, Schembri MA (2015) Molecular Characterization of a Multidrug Resistance IncF Plasmid from the Globally Disseminated Escherichia coli ST131 Clone. PLoS One 10(4):e0122369 doi:10.1371/journal.pone.0122369

Phan MD, Peters KM, Sarkar S, Lukowski SW, Allsopp LP, Gomes Moriel D, Achard ME, Totsika M, Marshall VM, Upton M, Beatson SA, Schembri MA (2013) The serum resistome of a globally disseminated multidrug resistant uropathogenic Escherichia coli clone. PLoS Genet 9(10):e1003834 doi:10.1371/journal.pgen.1003834

Pickard D, Kingsley RA, Hale C, Turner K, Sivaraman K, Wetter M, Langridge G, Dougan G (2013) A genomewide mutagenesis screen identifies multiple genes contributing to Vi

capsular expression in Salmonella enterica serovar Typhi. J Bacteriol 195(6):1320-6 doi:10.1128/JB.01632-12

Reid AN, Pandey R, Palyada K, Whitworth L, Doukhanine E, Stintzi A (2008) Identification of Campylobacter jejuni genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis. Appl Environ Microbiol 74(5):1598-612 doi:AEM.01508-07 [pii]10.1128/AEM.01508-07

Roggo C, Coronado E, Moreno-Forero SK, Harshman K, Weber J, van der Meer JR (2013) Genome-wide transposon insertion scanning of environmental survival functions in the polycyclic aromatic hydrocarbon degrading bacterium Sphingomonas wittichii RW1. Environ Microbiol 15(10):2681-95 doi:10.1111/1462-2920.12125

Rohmer L, Hocquet D, Miller SI (2011) Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. Trends Microbiol 19(7):341-8 doi:10.1016/j.tim.2011.04.003

Roux D, Danilchanka O, Guillard T, Cattoir V, Aschard H, Fu Y, Angoulvant F, Messika J, Ricard JD, Mekalanos JJ, Lory S, Pier GB, Skurnik D (2015) Fitness cost of antibiotic susceptibility during bacterial infection. Sci Transl Med 7(297):297ra114 doi:10.1126/scitranslmed.aab1621

Santa Maria JP, Sadaka A, Moussa SH, Brown S, Zhang YJ, Rubin EJ, Gilmore MS, Walker S (2014) Compound-gene interaction mapping reveals distinct roles for Staphylococcus aureus teichoic acids. Proc Natl Acad Sci U S A 111(34):12510-5 doi:10.1073/pnas.1404099111

Santiago M, Matano LM, Moussa SH, Gilmore MS, Walker S, Meredith TC (2015) A new platform for ultra-high density Staphylococcus aureus transposon libraries. BMC Genomics 16:252 doi:10.1186/s12864-015-1361-3

Sarmiento F, Mrázek J, Whitman WB (2013) Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Methanococcus maripaludis. Proc Natl Acad Sci U S A 110(12):4726-31 doi:10.1073/pnas.1220225110

Sassetti C, Rubin EJ (2002) Genomic analyses of microbial virulence. Curr Opin Microbiol 5(1):27-32

Sassetti CM, Boyd DH, Rubin EJ (2001) Comprehensive identification of conditionally essential genes in mycobacteria. Proc Natl Acad Sci U S A 98(22):12712-7 doi:10.1073/pnas.231275498

Sassetti CM, Rubin EJ (2003) Genetic requirements for mycobacterial survival during infection. Proc Natl Acad Sci U S A 100(22):12989-94 doi:10.1073/pnas.2134250100

Shan Y, Lazinski D, Rowe S, Camilli A, Lewis K (2015) Genetic Basis of Persister Tolerance to Aminoglycosides in Escherichia coli. MBio 6(2) doi:10.1128/mBio.00078-15

Sharma CM, Vogel J (2009) Experimental approaches for the discovery and characterization of regulatory small RNA. Curr Opin Microbiol 12(5):536-46 doi:10.1016/j.mib.2009.07.006

Skurnik D, Roux D, Aschard H, Cattoir V, Yoder-Himes D, Lory S, Pier GB (2013) A comprehensive analysis of in vitro and in vivo genetic fitness of Pseudomonas aeruginosa using high-throughput sequencing of transposon libraries. PLoS Pathog 9(9):e1003582 doi:10.1371/journal.ppat.1003582

Subashchandrabose S, Smith SN, Spurbeck RR, Kole MM, Mobley HL (2013) Genome-wide detection of fitness genes in uropathogenic Escherichia coli during systemic infection. PLoS Pathog 9(12):e1003788 doi:10.1371/journal.ppat.1003788

Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M (2015) Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc Natl Acad Sci U S A 112(13):4110-5 doi:10.1073/pnas.1419677112

Valentino MD, Foulston L, Sadaka A, Kos VN, Villet RA, Santa Maria J, Lazinski DW, Camilli A, Walker S, Hooper DC, Gilmore MS (2014) Genes contributing to Staphylococcus aureus fitness in abscess- and infection-related ecologies. MBio 5(5):e01729-14 doi:10.1128/mBio.01729-14

van Opijnen T, Bodi KL, Camilli A (2009) Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. Nat Methods 6(10):767-72 doi:nmeth.1377 [pii]10.1038/nmeth.1377

van Opijnen T, Camilli A (2012) A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res doi:gr.137430.112 [pii]10.1101/gr.137430.112

van Opijnen T, Camilli A (2013) Transposon insertion sequencing: a new tool for systems-level analysis of microorganisms. Nat Rev Microbiol 11(7):435-42 doi:10.1038/nrmicro3033

Wang N, Ozer EA, Mandel MJ, Hauser AR (2014) Genome-wide identification of Acinetobacter baumannii genes necessary for persistence in the lung. MBio 5(3):e01163-14 doi:10.1128/mBio.01163-14

Wang Q, Millet YA, Chao MC, Sasabe J, Davis BM, Waldor MK (2015) A Genome-Wide Screen Reveals that the Vibrio cholerae Phosphoenolpyruvate Phosphotransferase System Modulates Virulence Gene Expression. Infect Immun 83(9):3381-95 doi:10.1128/IAI.00411-15

Watson M (2014) Illuminating the future of DNA sequencing. Genome Biol 15(2):108 doi:10.1186/gb4165

Weerdenburg EM, Abdallah AM, Rangkuti F, Abd El Ghany M, Otto TD, Adroub SA, Molenaar D, Ummels R, Ter Veen K, van Stempvoort G, van der Sar AM, Ali S, Langridge GC, Thomson NR, Pain A, Bitter W (2015) Genome-Wide Transposon Mutagenesis Indicates that Mycobacterium marinum Customizes Its Virulence Mechanisms for Survival and Replication in Different Hosts. Infect Immun 83(5):1778-88 doi:10.1128/IAI.03050-14

Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A (2015) Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio 6(3) doi:10.1128/mBio.00306-15

Wong SM, Bernui M, Shen H, Akerley BJ (2013) Genome-wide fitness profiling reveals adaptations required by Haemophilus in coinfection with influenza A virus in the murine lung. Proc Natl Acad Sci U S A 110(38):15413-8 doi:10.1073/pnas.1311217110

Wong SM, Gawronski JD, Lapointe D, Akerley BJ (2011) High-throughput insertion tracking by deep sequencing for the analysis of bacterial pathogens. Methods Mol Biol 733:209-22 doi:10.1007/978-1-61779-089-8_15

Yamaichi Y, Chao MC, Sasabe J, Clark L, Davis BM, Yamamoto N, Mori H, Kurokawa K, Waldor MK (2015) High-resolution genetic analysis of the requirements for horizontal transmission of the ESBL plasmid from Escherichia coli O104:H4. Nucleic Acids Res 43(1):348-60 doi:10.1093/nar/gku1262

Yung MC, Park DM, Overton KW, Blow MJ, Hoover CA, Smit J, Murray SR, Ricci DP, Christen B, Bowman GR, Jiao Y (2015) Transposon Mutagenesis Paired with Deep Sequencing of Caulobacter crescentus under Uranium Stress Reveals Genes Essential for Detoxification and Stress Tolerance. J Bacteriol 197(19):3160-72 doi:10.1128/JB.00382-15

Zhang YJ, Ioerger TR, Huttenhower C, Long JE, Sassetti CM, Sacchettini JC, Rubin EJ (2012) Global assessment of genomic regions required for growth in Mycobacterium tuberculosis. PLoS Pathog 8(9):e1002946 doi:10.1371/journal.ppat.1002946

Zhang YJ, Reddy MC, Ioerger TR, Rothchild AC, Dartois V, Schuster BM, Trauner A, Wallis D, Galaviz S, Huttenhower C, Sacchettini JC, Behar SM, Rubin EJ (2013) Tryptophan biosynthesis protects mycobacteria from CD4 T-cell-mediated killing. Cell 155(6):1296-308 doi:10.1016/j.cell.2013.10.045

CHAPTER 2:

Genetic mechanisms of *Salmonella enterica* Typhimurium for overcoming host stressors

Rabindra K. Mandal[1], Tieshan Jiang[1] and Young Min Kwon[1, 2*]

[1]Department of Poultry Science, University of Arkansas, Fayetteville, AR, 72701

[2]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, 72701

*Corresponding author:

Department of Poultry Science

College of Agricultural, Food and Life Sciences

University of Arkansas

Fayetteville, AR 72701

Phone: (479) 575-4935

Email ykwon@uark.edu

Abstract

Salmonella enterica serovar Typhimurium (*S.* Typhimurium), a non-typhoidal *Salmonella* (NTS), result in a range of diseases, including self-limiting gastroenteritis, bacteremia, enteric fever, and focal infections representing a major disease burden worldwide. There is still a significant portion of *Salmonella* genes whose biochemical basis to overcome host innate defense mechanisms, consequently causing disease in host, largely remains unknown. Here, we have applied a high-throughput transposon sequencing (Tn-seq) method to unveil the genetic factors required for the growth or survival of *S.* Typhimurium under various host stressors simulated *in vitro*. A highly saturating Tn5 library of *S*. Typhimurium 14028s (≥ 186, 000 unique insertions) was subjected to selection during growth in the presence of short chain fatty acid (100 mM Propionate), osmotic stress (3% NaCl) or oxidative stress (1mM H_2O_2) or survival in extreme acidic pH (30 min in pH 3) or starvation (12 days in PBS). We have identified an overlapping set of 339 conditionally essential genes (CEGs) required by *S.* Typhimurium to overcome these host insults. Interestingly, entire eight genes encoding F_0F_1-ATP synthase subunit proteins were required for fitness in all the five stresses. Intriguingly, *Salmonella* pathogenicity island (SPI) genes like SPI-1, SPI-2, SPI-3, SPI-5, SPI-6 and SPI-11 are not only required during host infection, but also for fitness under *in vitro* conditions. Additionally, by comparative analysis of the genes identified in this study and the genes previously shown to be required for *in vivo fitness*, we identified novel genes (*marBCT*, *envF*, *barA*, *hscA*, *rfaQ*, *rfbI* and putative proteins STM14_1138, STM14_3334, STM14_485, and STM_5184) that has compelling potential to be exploited as vaccine development and/or drug target to curb the *Salmonella* infection.

Key Words: *Salmonella*, host stress, Tn-seq, conditionally essential genes

INTRODUCTION

Non-typhoidal *Salmonella* (NTS), a gram-negative bacterial pathogen, causes 93 million enteric infections, 155,000 diarrheal deaths, and 3.4 million blood stream infection worldwide annually (Ao et al., 2015; Majowicz et al., 2010). Gram-negative bacterial pathogens, including NTS, are developing resistance against antimicrobial agents including the last resort antibiotics at a startling rate, creating a global crisis in human health. Scientists fear the impending global epidemic of untreatable infections and return to a pre-antibiotic era where a common infection and minor injury can be lethal (Liu et al., 2015; McKenna, 2013; Spencer, 2015; World Health Organization (WHO)). Thus, there is an urgent need to identify genetic factors of pathogenic microorganisms that can serve as targets to develop novel strategies to combat infectious diseases (Medini et al., 2008; van Opijnen and Camilli, 2012). Nonetheless, the insufficiency of the genome-wide data that provide links between genotype and the infection-related phenotypes of bacteria is the major roadblock to discover suitable targets for development of the effective strategies to control infection.

Salmonella enterica serotype Typhimurium (*S.* Typhimurium) is one of the leading cause of NTS (Carden et al., 2015; Crim et al., 2015). Despite *Salmonella* infection has an enormous global burden on disease worldwide and availability of complete genome sequence of *S.* Typhimurium LT2 nearly one and half decade (2002) ago, the biochemical basis of *S.* Typhimurium genes required for *in vivo* survival is still unknown for a large portion of the the genes (Feasey et al., 2012; McClelland et al., 2001). Researchers have tried to delve into the pathogenesis of *S.* Typhimurium using different variations of high throughput screening of transposon mutants, with a limited number of mutants based on a negative selection (Kwon et al., 2016). Chan *et al*. (2005) had detected 157 and 264 genes required by *S.* Typhimurium strain SL1344 for acute

infection in mice (A-Mice) and survival inside macrophage (MΦ), respectively using a microarray-based tracking method (Chan et al., 2005). Lawley *et al*. (2006) used the same method to identify 118 genes of *S.* Typhimurium strain SL1344 required for long-term persistent infection in mice (P-Mice) collected from spleen after 28 day post infection (Lawley et al., 2006). Additionally, Chaudhuri et al. (2013) have comprehensively assigned a core set of 611 genes of *S.* Typhimurium strain ST4/74 required for effective colonization in the calf, pig, and chicken (Chaudhuri et al., 2013). Recently, Silva-Valenzuela et al. identified 224 mutants of *S.* Typhimurium 14028S that were negatively selected using two pools of single gene deletion mutants from spleen and liver at 2 days post infection in mice (Sp-Liv) (Silva-Valenzuela et al., 2015). Previously, our laboratory conducted Tn-seq screening to identify an overlapping set of 105 coding genes of *S*. Typhimurium 14028S required for *in vitro* growth in diluted Luria-Bertani (LB) medium, LB medium plus bile acid and LB medium at 42°C (Khatiwara et al., 2012). However, there is still a gap in the above approach to correlate *in vivo* and *in vitro* survival or growth genes required by *S.* Typhimurium that will help delve into biochemical basis of pathogenesis and potentially pave a roadmap towards the efficient development of novel vaccines, antibiotics, and control strategies.

In this study, we conducted Tn-seq analysis of *S.* Typhimurium 14028s under the five *in vitro* conditions mimicking host stressors during enteric and systemic infection. We have applied a highly efficient method for Tn-seq library preparation that requires only small amount of DNA without the need for enzymatic digestion or physical shearing of genomic DNA (Dawoud et al., 2014). To cause enteric infection *S.* Typhimurium has to overcome gastrointestinal host insult like low acidic pH in the stomach, osmotic and short chain fatty acid (SCFAs) in intestine (Ha et al., 1998; Nava et al., 2005; Sleator and Hill, 2002; Smith, 2003). Eventually, for systemic

infection, *S.* Typhimurium has to vanquish macrophage stress like oxidation, starvation as well as hyperosmotic condition (Lee et al., 2014; Rosenberger and Finlay, 2003; van der Heijden et al., 2015). We hypothesized that the comparative analysis of the comprehensive sets of the *in vivo* (required for host enteric and systemic infection) and *in vitro* fitness genes (for stress resistance, this study and previous) will allow better understanding of the biochemical basis of the genetic requirements of *S.* Typhimurium for host infection and provide enhanced resolution to link genotype to phenotype. Thus, we performed a comparative study between the *in vivo* and *in vitro* fitness genes from previous studies and this study, respectively.

RESULTS AND DISCUSSION

Overall evaluation of resulting Tn-seq profiles

We have constructed a highly saturated transposon mutant library of *S*. Typhimurium 14028S with more than 350,000 transposon mutants created via transformation of EZ-Tn5 transposome complex to electrocompetent cells. The complex Tn5 library, input pool 1(IP1) was then subjected to negative selection under the *in vitro* stress conditions encountered during enteric and systemic infection cycle as described in Materials and Methods. Input pool 2 (IP2) was the technical replicate of IP1 to evaluate the reproducibility of our Tn-seq method (Figure 1). DNA library of the input and output pool for Illumina sequencing were prepared (Figures S1A and S1B). This efficient Tn-seq protocol was developed in our laboratory that offers distinctive advantages over other Tn-seq library preparation methods, including a low amount (~100 ng) of DNA required, and no need for physical shearing or restriction digestion (Dawoud et al., 2014; Kwon et al., 2016).

Illumina sequencing using HiSeq 3000 produced 163,943,475 reads from a single flow cell lane. The raw reads were demultiplexed allowing a perfect match for barcode with exception of few mismatches within Tn5 mosaic end with a custom Perl script as shown in Table S5. H_2O_2 (19,250,956) had the highest number of reads followed by IP1 (10,842,764), Starvation (9,518,226), IP2 (6,345,173), LB (5,004,934), pH3 (3,841,401), PA (2,113,033) and NaCl (1,970,072) (Figure 2A). Although, we pooled equal quantity of PCR products (~10 ng), there was a considerable variation in the number of the reads across the samples, suggesting the need for better strategy for sample pooling.

After demultiplexing, Illumina reads were trimmed of barcode and transposon sequences. The Tn5-junction sequences of 20bp were extracted and mapped to the complete genome of *S.* Typhimurium 14028S (NC_016856.1) using Bowtie. The overall alignment rate throughout the Tn5 libraries were 85.19% (SE ± 1.79). Additionally, we looked for the unique insertion sites in the genome in each library. IP1 had had highest unique insertions (186,621) followed by LB (157,915), H_2O_2 (149,752), IP2 (149,740), PA (127,722), NaCl (125,918), Starvation (118,607) and pH3 (92,008). Similarly, H_2O_2 had highest average read per unique insertion site in the genome (96.007 ± 1.11) with 40 median reads whereas NaCl had the lowest (13.53 ± 0.99) with 5 median reads (Figure 2A).

Pre-aligned reads of the Tn5 library in default SAM mapping file format were fed to 'Analysis of high-Resolution Transposon-Insertion Sequences Technique' (ARTIST) pipeline (Pritchard et al., 2014). Tn5 insertions were mapped into 100 bp genome-wide windows. We observed a higher Spearman correlation of IP1 with IP2 and LB (0.98, $p < 0.0001$). However, there was lower Spearman correlation of IP1 with NaCl (0.97, $p < 0.0001$), PA (0.96, $p < 0.0001$), and H_2O_2 (0.93, $p < 0.0001$). We observed lowest correlation of IP1 with pH3 and starvation (0.84

and 0.91 respectively, $p < 0.0001$) (Figure 2B). This relationship corroborates well with the Tn5 library selection strategies employed, with that based on growth fitness showing higher correlation and that based on survival a lower correlation.

Besides, we looked for the occurrence of any hot spots of Tn5 insertion in the sample libraries. We found an even distribution of Tn5 insertion reads across the libraries throughout the genome. Some of the genomic coordinate lacking insertions have white stripes that are clearly visible (Figure S2) across all the samples that represent essential loci in the *S.* Typhimurium genome.

Identification of Conditionally essential genes (CEGs)

In this study, we used two strategies to identify conditionally essential genes (CEGs) of *S.* Typhimurium to overcome host stressors. The first strategy was a negative selection of complex Tn5 mutant libraries based on growth fitness for mild stressors (3% NaCl, 100mM propionate, 1mM H_2O_2) and second was based on survival (12 day starvation and pH3) of Tn5 mutant libraries for harsher stressors as shown in Figure 1.

ARTIST pipeline can identify if genes are entirely essential or domain essential in a given condition. For simplicity, we assigned genes either entirely essential or domain essential into one category, conditionally essential genes (CEGs). Noteworthy, only a few of the genes were identified as domain essential and the majority of them were entirely essential. We deliberately compared output complex Tn5 libraries PA, NaCl, and H_2O_2 with both IP1 and LB. Interestingly, most of the CEGs were overlapped with this two comparisons. For these conditions PA, NaCl, and H_2O_2, we considered the union set of identified CEGs via comparison of output library with IP1 and LB as CEGs for each condition. However, the output libraries for pH3, and starvation

were compared only with IP1 because the selection of complex Tn5 library was based on survival strategy in which the mutant cells did not multiply in liquid media.

We identified an overlapping set of 339 CEGs that are required for fitness of *S.* Typhimurium in at least one of the five conditions. Starvation had the highest CEGs (241), followed by pH3 (103), NaCl (60), H_2O_2 (40) and PA (19) as shown in Table S1. This might likely reflect that starvation is the most severe stressor involving more genetic pathways for survival, while PA is a mild stressor for the fitness of *S.* Typhimurium. More than a half of CEGs were on the lagging strand (56.63%), which is somewhat contrary to responsive genes of *Escherichia coli* and *Streptococcus pneumoniae* (Nichols et al., 2011; van Opijnen and Camilli, 2012). We assigned a functional role to 96 CEGs that were putative proteins and 21 CEGs belonging to hypothetical proteins. The other common stress tolerant proteins in at least 2 of the *in vitro* stressors were ATP synthase, a transcriptional regulator, 3-dehydoroquinate synthase, site-specific tyrosine recombinase *xerC*, flavin mononucleotide phosphatase, ribulose-phosphate 3-epimerase, and DNA-dependent helicase II among others (Table S1).

Intriguingly, we found the *Salmonella* pathogenicity island (SPI) genes were required for *in vitro* stress tolerance. SPI-1 and SPI-2 genes encoding Type III secretion system, SPI-3, SPI-5, SPI-6, and SPI-11 were required for fitness in pH3, NaCl, and Starvation stress. However, no SPI genes were identified for fitness in PA and H_2O_2. SPI-5 and SPI-11 genes were only conditionally essential in pH3, SPI-3 in NaCl and SPI-6 in starvation. Additionally, SPI-1 and SPI-2 were important for fitness in both NaCl and starvation stress (Table S1; Read coverage shown in Figure S3).

For broader insight into stress endurance pathways and functional role, we assigned each CEGs to the cluster of orthologous groups (COG) using eggNOG database (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (Jensen et al., 2008). The CEGs having top hit for the COG in the *S*. Typhimurium str. LT2 were kept and CEGs with no orthologous group were allotted to group XX (**Figure 3B**; Table S1). In overall 21.83% of CEGs belonged to category "function unknown" followed by "intracellular trafficking, secretion, and vesicular transport " (10.91%), "energy production and conversion" (9.44%), and "no orthologs found" (8.26%) among others. With a greater portion of CEGs (30.6%) falling into either "function unknown or "no orthologs found" shows that our data set is rich in novel genotype-phenotype relationships.

Additionally, we were interested to see if the CEGs identified in our study fell into the essential genomes of *S*. Typhimurium. Essential genomes of *S.* Typhimurium strain SL3261 (selected on LB agar) (Barquist et al., 2013) and *S.* Typhimurium strain LT2 (selected on rich medium) (Knuth et al., 2004; Zhang et al., 2004) were compared with the CEGs of *S*. Typhimurium 14028S. Genes in different strain background were looked for the orthologous gene in same strain background (*S*. Typhimurium 14028S). Interestingly, 10 and 15 CEGs in this study were shared with the essential genes of *S*. Typhimurium strain SL3261 and LT2, respectively (Table S2, Figure S4). This indicates that these genes are dispensable in *S.* Typhimurium 14028S strain background, making the essential genome smaller in this strain as compared to other two strains.

Genetic and biochemical basis of CEGs in *S*. Typhimurium

We delve into the genetic and biochemical mechanisms related to the CEGs identified in our study. For convenience, we split the section into specific CEGs, required for fitness in only one

stressor, and common CEGs, shared in at least two stressors out of five host stressors mimicked *in vitro*.

Specific CEGs required for propionate (100 mM PA) stress resistance

CEGs specific for fitness of *S.* Typhimurium in propionate were *yiiD* and *sdhAD*. YiiD is a putative acetyltransferase protein (Read coverage shown in Figure 4C). Acetylation, a post-translation modification of protein was previously shown to enable prokaryotes to increase stress resistance (Ma and Wood, 2011). Additionally, succinate dehydrogenase flavoprotein (*sdhA*) and cytochrome b566 (*sdhD*) subunit proteins were up-regulated by intestinal SCFA in *S.* Typhimurium (Lawhon, 2002). Chowdhury and Shimizu (2008) reported that *sdhA* in the tricarboxylic acid cycle (TCA) were highly induced during temperature upshift in *E. coli* (Hasan and Shimizu, 2008).

Specific CEGs required for osmotic (3% NaCl) stress resistance

Twenty-six resistance genes of *S.* Typhimurium were required for fitness in osmotic stress (3% NaCl) alone. Protein-protein network analysis using STRING database (http://string-db.org) against *S. enterica* LT2 showed three distinct clustering of genes, SPI-3 (*mgtBC*, *misL*, *cigR*, *slsA*, *fidL* and *marT*), two-component system (*dcuBRS*) and sodium ion transport (*yihPO*) along with other nodes (http://bit.ly/2bCKGVG). SPI-3 genes are important for intracellular replication inside phagosome where *Salmonella* experience hyperosmotic stress (Schmidt and Hensel, 2004). The virulence protein *mgtC* and *mgtB*, Mg^{2+} transporter were expressed five-fold when *Salmonella* Typhimurium was exposed to 0.3 M NaCl (Lee and Groisman, 2012). MisL, an autotransporter protein is an intestinal colonization factor (activated by *marT*, a transcriptional regulator) that binds to extracellular matrix fibronectin in an animal host and is also involved in

adhesion to plant tissue (Dorsey et al., 2005; Kroupitski et al., 2013). Deletion of *cigR* in *S.* Pullorum resulted in a significantly decreased biofilm formation and increased virulence (Yin et al., 2016). However, Figueira et al. showed Δ*cigR* strain of *S*. Typhimurium had attenuated replication in mouse bone marrow-derived macrophage (Figueira et al., 2013).

yihPO genes are essential for capsule assembly that is required by *Salmonella* for environmental stress persistence such as desiccation (Gibson et al., 2006). The absence of *ompL* (ortholog of *yshA*) leads to solvent hypersensitivity as it helps in the stabilization of cell wall integrity protecting from solvent penetrance as a physical barrier (Murinova and Dercova, 2014). In *E. coli*, the genes under the control of *dcuS-dcuR*, a two-component system, were not affected upon a hyperosmotic shock (Weber and Jung, 2002). However, *dcuBRS* were conditionally essential in *S*. Typhimurium for fitness during osmotic stress. Putative cytoplasmic protein (STM14_4542, STM14_4828, and STM14_5175), putative inner membrane protein (STM14_4824 and STM14_5184) and putative hydrolase (STM14_4823) were also required for osmotic stress tolerance.

Specific CEGs required for oxidative (1 mM H_2O_2) stress resistance

We identified 16 specific resistance genes required for fitness of *S*. Typhimurium in the presence of 1 mM H_2O_2 and the functional protein association network analysis among the genes was constructed using STRING against *S. enterica* LT2 (http://bit.ly/2bsVKXF). Major resistance genes were those involved in two-component system (*glnD*, *rpoN*, *arcA* (STM4598), and *arcB* (STM3328)), DNA recombination (*recJ*, and *xerD*), and metal ion transport (*corA*, and *trkA*).

Hydrogen peroxide kills *E. coli* cells with two distinct modes, mode-1 killing occurs at a lower concentration of H_2O_2 due to DNA damage and mode-2 killing occurs at a higher concentration

of H_2O_2 due to damage of other structures like proteins and lipids(Imlay and Linn, 1986). Nucleic acid metabolic process genes involved in oxidative stress resistance were *recJ*, *xerD*, *sun,* and *rpoN*. *RecJ* protein, a single-stranded DNA (ssDNA)-specific 5'-3' exonuclease/deoxribophophodiesterase, plays a role in homologous recombination, mismatch repair, and base excision repair (Wakamatsu et al., 2011). In *E. coli*, *xerD* knockout mutants are hypersensitive to tightly bound DNA-protein complexes (TBCs) that block replication forks *in vivo* (Henderson and Kreuzer, 2015). *RpoN*, the alternative sigma factor 54 (σ^{54}), an important regulator of stress resistance and virulence genes in many bacterial species (Riordan et al., 2010). σ^{54} is involved in carbon/nitrogen limitation, nucleic acid damage, cell envelope, and nitric oxide stress (Hartman et al., 2016). However, Hwang *et al.* 2011 found that *rpoN* mutant in *Campylobacter jejuni* was more resistant to 1mM H_2O_2 (Hwang et al., 2011).

Besides, cellular component genes crucial for fitness in H_2O_2 stress were *dsbC*, *glmS*, *trkA*, *corA* including *sun* and *xerD*. *DsbC*, a protein essential for disulfide bond isomerization in the periplasm, has a new role in *E. coli* in protection against oxidative stress (Denoncin et al., 2014). In *E. coli GlmS* plays an important role in cell wall synthesis thus providing protection against cell envelope stress response (Zhou et al., 2009). HscB, a chaperone-encoding gene is upregulated after exposure to oxidative stress in *Burkholderia pseudomallei* (Jitprasutwit et al., 2014). YbgF, an outer membrane vesicle protein, increases the survival of bacteria during exposure to stress or from toxic unfolded proteins by releasing the unwanted periplasmic component (Gogol et al., 2011).

Specific CEG_S required for higher acidic (pH 3) stress resistance

We found 49 specific stress resistance genes required for survival of *S*. Typhimurium in extreme acidic condition (pH 3). Formate dehydrogenase (*fdoHI*, and *fdhDE*) curli proteins (*csgBDEFG*), virulence and envelope proteins (SPI-2: *orf245*, *orf408*, *ssaB*; SPI-5: *pipBC*, *sopB*, and SPI-11: *envEF*, *pagCD*, *msgA*, STM14_1486 where *ssaB*, *pipB,* and *sopB* are effector proteins), and biopolymer transport protein (*exbD* and *exbB*) were clustered in functional protein association network analysis using STRING (http://bit.ly/2bCLVnL).

Formate dehydrogenase catalyzes the oxidation of formate (HCOO-) to CO_2 and H^+. The released electrons from this reaction are used by two cytoplasmic protons to form dihydrogen thus consuming net protons, consequently, counteracting acidification (Leonhartsberger et al., 2002). Curli are major complex extra-cellular proteinaceous matrix produced by *Enterobacteriaceae* that helps pathogenic bacteria like *Salmonella* in adhesion to surfaces, cell aggregation, and biofilm formation (Barnhart and Chapman, 2006). Acidic pH strongly enhances biofilm formation in *Streptococcus agalactiae* (D'Urzo et al., 2014). We hypothesize that curli fibers might potentially protect bacteria from severe acid stress through the physical barrier and likely by the generation of alkaline compounds as in oral biofilms (Cotter and Hill, 2003). PhoP regulates SPI-11 genes such as *envEF*, *pagCD*, and *msgA* where later three are required by *Salmonella* to survive low pH within macrophage (Gunn et al., 1995; Lee et al., 2013). In *Helicobacter pylori,* only the organism to colonize in the acidic human stomach, *ExbB*/*ExbD*/*TonB* complex is required for acid survival and periplasmic buffering (Marcus et al., 2013). Additionally, survival of *ΔexbD* was diminished compared to wild type at pH 3 in *E. coli* (Ahmer et al., 1995). The *metC* gene encoding a key enzyme in methionine biosynthesis,

required for the generation of homocysteine, pyruvate, and ammonia, play a crucial role in bacterial acid stress responses (Reid et al., 2008).

Specific CEGs required for starvation stress resistance

Out of 261 *Salmonella* fitness genes essential for starvation stress, 160 genes were explicitly important for resistance against starvation stress among the five infection-relevant conditions in this study (http://mcaf.ee/k0uhrm). Major enriched gene pathways were oxidative phosphorylation, pathogenesis, two-component system, and lipopolysaccharide biosynthetic process among others.

NADH dehydrogenase, the first component of the respiratory chain, subunit proteins (*nuoCEFGHLMN*) were required for fitness of *Salmonella* during long-term carbon starvation. *Salmonella* defective in NADH dehydrogenase enzyme exhibits defective energy-dependent proteolysis during carbon starvation (Archer et al., 1993). Proteolysis of unbound or unemployed proteins helps bacteria to access nutrients as an important survival strategy during carbon starvation (Michalik et al., 2009). SPI-1 (*hilACD*, *iagB*, *invH*, *orgAC*, *prgHIJK*, STM14_3500, and STM14_3501) and SPI-2 (*ssaMNOPQRSTV*, *sscB*, and *sseDEF*) encoding type III secretion system (T3SS) and SPI-6 (*safABCD*, *sinR*, STM14_0359, and *ybeJ*) encoding type VI (T6SS) secretion system were required for *in vitro* survival in long-term starvation stress. *Salmonella* usually requires SPI-1 genes for the invasion of intestinal epithelial cells (Klein et al., 2000). HilACD regulates SPI-1 invasion gene expression during multiple environmental conditions including stationary phase, pH, osmolarity, oxygen tension, and short chain fatty acids (Olekhnovich and Kadner, 2007). SPI-2 genes are expressed under *in vitro* starvation conditions indicating the use of nutritional deprivation as a signal (Hensel, 2000). T6SS has been

hypothesized to confer a growth advantage to bacteria in environmental niches where bacterial competition for nutrient is critical for survival (Brunet et al., 2015).

Two-component systems (TCs), a basic stimulus-response coupling mechanism, enable microbes to respond to various stimuli such as pH, osmolarity, quorum signals, or nutrient availability and regulate their cellular functions (Freeman et al., 2013). TCs required for fitness during starvation conditions were *envZ*/*OmpR*, *cpxA*/*cpxR*, sensory histidine kinase protein (*phoQ*), and kdpD (Figure S3B). EnvZ/*OmpR* regulates the synthesis of porin proteins (*ompF* and *OmpC*) that are important for the survival of *E. coli* in sea water under starvation stress condition (Darcan et al., 2009). It is believed that carbon starvation causes cell envelope stress. Bacchelor et al. (2005) found *cpxA*/*cpxR* in *E. coli* regulates the expression of prions *ompF* and *ompC*, a major component of the outer membrane. However, Kenyon et al. (2002) showed the starvation stress of *S.* Typhimurium do not require cpxR-regulated extra-cytoplasmic functions (Batchelor et al., 2005; Kenyon et al., 2002). PhoQ and *kdpD* plays a role in Mg^{2+} and K^{+} homeostasis respectively, critical to the virulence and intracellular survival of *S*. Typhimurium (Freeman et al., 2013; Kato and Groisman, 2008).

The outer membrane of Gram-negative bacteria contains phospholipids and lipopolysaccharides (LPS). LPS molecules act as a permeability barrier to prevent the entry of toxic compounds and allow the entry of nutrient molecules (Schakermann et al., 2013). LPS biosynthetic process genes required for fitness in starvation conditions were *rfbABCD*, *rfbUNMKP, galF*, *udg*, *wzxE,* and *wzzB*. Starvation of carbon energy source activates envelope stress response in *S.* Typhimurium (Rowley et al., 2006). Additionally, *pstSCAB* coding for the Pst ABC transporter catalyzes the uptake of inorganic phosphate (Lüttmann et al., 2012). Mutations in the Pst system results in structural modifications of lipid A and an imbalance in unsaturated fatty acids consequently

leading to increase in outer membrane permeability making *E. coli* more vulnerable to environmental stresses including antimicrobial peptide and low pH (Lüttmann et al., 2012).

Additional genes required for starvation stress resistance were *aroGH*, ytfMNP (ytfM - outer membrane protein), *stcB* (putative periplasmic outer chaperone protein). Furthermore, other envelope proteins were outer membrane lipoproteins (*stcD* and *yifL*), putative outer membrane proteins (*stcC*, STM14_0404, and ytfM), and putative inner membrane proteins (STM14_0398, STM14_0402, STM14_2763, STM14_4741, STM14_4742, STM14_4745, STM14_4880, *ydiK* and *yjeT*). Similarly, putative cytoplasmic proteins required for starvation stress were STM14_2759, STM14_4743, STM14_5374, *ydiL,* and *ytfP*.

CEGs required for tolerance to multiple stressors

We found 12 *Salmonella* genes required for stress resistance in either three or four of the *in vitro* host stresses in our study as shown in STRING protein-protein interaction network (http://bit.ly/2btx1zg). The enriched GO biological process / KEGG pathways were ncRNA processing (*gidAB* and *mnmE*), DNA metabolic process (*dam*, *uvrD* (SOS response), *xerC*), and biosynthesis of amino acids (*aroB* and *rpe* - microbial metabolism in diverse environments). In addition, other responsive proteins include ATP synthase subunit protein (*atpI*), putative permease (STM14_4659), inner membrane protein (*damX*), and flavin mononucleotide phosphatase.

DamX, *dam*, *rpe*, *aroB*, *uvrD*, and *yigB* were required for fitness in pH3, Starvation, and H_2O_2. Disruption of *damX* in *S. enterica* causes bile sensitivity (López-Garrido and Casadesús, 2010). DNA adenine methylation gene (*dam*) plays an important role in bacterial gene expression and virulence (Low et al., 2001). Dam mutants of *S. enterica* are extremely attenuated in mouse

(Jakomin et al., 2008). The gene *aroB* encodes dehyroquinate synthase, a part of shikimate pathway, is essential for bacteria and absent in mammals (de Mendonca et al., 2007). In prokaryote species, *uvrD* is involved in maintaining genomic stability and helps DNA lesion repair, mismatch repair, nucleotide excision repair and recombinational repair (Kang and Blaser, 2006). Overproduction of *yigB* produced higher-level persister, cells that exhibit multidrug tolerance, in *E. coli* (Hansen et al., 2008). However, deletion of *gidB* (glucose-inhibited division gene B) confers high-level antimicrobial resistance in *Salmonella* and has compromised overall bacterial fitness compared to wildtype (Mikheil et al., 2012). GidA (together with *mnmE*) is responsible for the proper biosynthesis of 5-methylaminomethtyl-2-thouridine of tRNAs and deletion causes attenuation in bacterial pathogenesis (Shippy and Fadl, 2014b).

ATP synthase genes are obligatory for *Salmonella* fitness during *in vitro* host stressors

ATP synthase (F_1F_0-ATPase) is a ubiquitous enzyme largely conserved across all domains of life. All the eight genes encoding ATP synthase subunit proteins were required for fitness of *S.* Typhimurium in every 5 *in vitro* conditions of our study (Figure 3A and 4A). F_1F_0-ATP synthase complex is required for ATP production from ADP and Pi. ATP synthase also regulates pH homeostasis in bacteria (*Listeria monocytogenes* and *S.* Typhimurium) at the expense of ATP (Balemans et al., 2012). In *Streptococcal faecalis*, upregulation of F_1F_0-ATPase promotes ATP-dependent H+ extrusion under acidic conditions. However, in *E. coli* the expression of ATP synthase is decreased under acidic condition (Krulwich et al., 2011). ATP synthase in *Mycobacterium* and *Staphylococcus* has been validated as a promising target for new antimicrobial drugs (Balemans et al., 2012; Lu et al., 2014).

Mechanistic basis of *Salmonella* to cause enteric and systemic infection

The network diagram shown in Figure 4 and Figure 5 shows all of the genes that are commonly important for fitness under at least one *in vitro* and *in vivo* conditions. The genes that were important only either in the *in vitro* or *in vivo* conditions were excluded in the diagram. This will show the biochemical basis of the genetic requirements for *in vivo* fitness.

Enteric infection

We have identified an overlapping set of 135 CEGs common in at least one of the host [pig, calf, and chicken (Chaudhuri et al., 2013)] that causes enteric infection and one of the *in vitro* host stressors [LB42, Bile (Khatiwara et al., 2012), pH3, PA, and NaCl] encountered during enteric infection (Figure 5, Table S3). Genes in SPI-1 (*invABCEIJ*, *sicAP*, *sipABCD*, *spaOPQRS*, *sptP*) and SPI-3 (*cigR*, *marT*, *mgtBC*, *misL*, *slsA*) were required for fitness in NaCl and all host. However, genes encoding SPI-2 (*sseCG*), SPI-5(*slsA*, *pipC*) and SPI-11(*envEF*) were essential for fitness only one *in vitro* stressor pH3 and intestinal colonization in 3 hosts. Other enriched pathways were lipopolysaccharide biosynthesis (*rfaIJKLQY* and *rfbBDKMNP*), oxidative phosphorylation (ATP synthase genes and *sdhA*), and biosynthesis of amino acids (*aroABD, rpe and metC*) including others as shown in STRIN protein-protein interaction against *S. enterica* LT2 (http://mcaf.ee/wzljud).

High osmolarity, low oxygen, and late log phase induce *hilA* expression *in vitro* that in turn regulates the expression of SPI-1 genes (Lostroh and Lee, 2001). Interestingly, we identified SPI-1 genes as fitness genes required for *in vitro* NaCl stressor. Similarly, lipopolysaccharide (LPS) biosynthetic process genes were enriched in LB42, Bile and in pig, calf, and chicken for fitness during enteric infection. LPS, a critical factor in the virulence of gram-negative bacterial

infection is required for intestinal colonization, resistance to killing by macrophage, swarming motility, serum resistance and bile stress (Khatiwara et al., 2012; Kong et al., 2011). CsgBA (curli subunit protein) mutant of *S*. Typhimurium was attenuated to elicit fluid accumulation in bovine ligated ileal loops (Tükel et al., 2005) and are required for fitness in pH3 including *csgF* and *csgG*. Additionally, putative proteins STM14_1138, STM14_1486, STM14_1981, STM14_3333 and STM14_4826, STM14_4828, STM14_5184, STM14_5185 (hypothetical protein) were required for fitness *in vitro* acidic and osmotic stress respectively and enteric infection in the entire three host.

Systemic infection

We compared the CEGs that are at least shared between the one of the *in vitro* systemic host stressors (H_2O_2, NaCl, pH3, Starvation and dLB (Khatiwara et al., 2012), stress encountered inside MΦ) and *in vivo* systemic infections (MΦ (Chan et al., 2005), A-Mice (Chan et al., 2005), P-Mice (Lawley et al., 2006), Sp-Liv (Silva-Valenzuela et al., 2015)) with an overlapping set of 130 genes (Figure 6, Table S4) shown in protein-protein interaction network using STRING (http://mcaf.ee/p34rjn). SPI-1 genes (*hilACD*, *iacP*, *iagB*, *invABCEFGI*, *orgA*, *prgHIJK*, *sicA*, *sipABC*, *spaOPQRS*) encoding TTSS were essential for fitness in NaCl, Starvation, MΦ survival and systemic infection. Additionally, SPI-2 genes (*ssaBCDEGIJKLMNOPQRSTV*, *orf245*, *orf408*, *sscAB*, *sseCDEF*, *ssrA*, STM14_1706) encoding TTSS were required for fitness in pH3, starvation, MΦ survival and systemic infection. Similarly, SPI-3 genes (*marBCT*) were required for fitness in NaCl, MΦ survival, and persistent infection in mice (P-Mice). SPI-11 genes (envF, pagCD) were required for fitness in pH3, MΦ survival, and P-Mice.

Other than SPI genes, the majorly enriched genes were nucleic acid metabolic process (*dam*, *trpS*, *MnmE*, *truA*, *serc*, *csgD*, *ompR* and *cra*), lipopolysaccharide biosynthetic process (*rfbABCNPU*, *rfaB*, *udg*, *galF*), oxidative phosphorylation (ATP synthase genes, NADH dehydrogenase genes), two component system (*ompR*, *barA*, *phoQ*, *glnDL*, *pagKO*) among others (Figure 6). Gene *dam* was required for fitness in H_2O_2, NaCl, A-Mice, and Sp-Liv. XerC and *rpe* were required for H_2O_2, pH3, Starvation and Sp-Liv. Interestingly, *pagK* were not identified as CEG in A-Mice, P-Mice, Sp-Liv but in pH3, Starvation, and MΦ. Putative genes either essential for one of *in vitro* or *in vivo* systemic infection were STM14_1138, STM14_4880, STM14_4992, STM14_5184, STM14_2759, STM14_2807, STM14_3334, STM14_4825, STM14_5299, and STM14_5300.

CONCLUSION

A recent study by Kroger et. al. (2013) presented transcriptomes of *S*. Typhimurium in 22 distinct infection-relevant environmental conditions *in vitro*. The researchers found induction of *Salmonella* pathogenicity islands *in vitro* conditions such as early stationary phase, anaerobic growth, oxygen shock, nitric oxide shock as well as in pH3, NaCl, bile, and peroxide shock including others (Kröger et al., 2013). However, transcription of a gene does not necessarily indicate the need of that gene in given particular condition. The transcript may be a leaky expression or required for fitness in the upcoming environment in a cost effective way through predictive adaptation, phenomena where microorganisms are able to anticipate and pre-emptively respond to the regular environmental fluctuations (temporally distributed stimuli) that confers a considerable fitness advantage for the survival of an organism (Mitchell et al., 2009; Tagkopoulos et al., 2008). Traditionally, it is believed that “central dogma of life” i.e. flow of information from DNA to RNA to proteins are highly concordant. However, there is a modest

correlation between levels of transcripts and corresponding proteins (Foss et al., 2007; Fu et al., 2009; Ghazalpour et al., 2011). Thus, that functional genomics screening such as Tn-seq is expected to reveal more direct functional aspects of the genes involved in responding to the current stresses.

In this report, we were able to map genotype to phenotype links providing the biochemical basis the genetic requirements for fitness for an overlapping set of 221 virulence genes for *in vivo* fitness (Figure S5). These CEGs were required for fitness in at least one of the *in vitro* host stressors (PA, NaCl, pH3, Starvation, Bile, LB42 and dLB), enteric infection (calf, chicken and pig), or systemic infection (mice including intracellular survival inside macrophage). Forty-four common CEGs were required to cause both systemic and enteric infections (*in vivo* fitness) and *in vitro* fitness (Figure S5 and Table S6). Common SPI genes for *in vivo* and *in vitro* fitness were SPI-1 (*invABCEI*, *sicA*, *sipABD*, *spaOPQRS*), SPI-2(*sseC*), SPI-3(*marT*, *mgtCB*) and SPI-11(*envF*). *Salmonella* genes other than SPI essential for fitness *in vitro* stresses and *in vivo* survival were *atpAEF*, *lepA*, *dam*, *pstB*, *xerC*, *manA*, *phoQ*, *rfaQ*, *rfbBIP*, *rpe*, *trmE*, *rfbIP*, *ompR*, *csgF*, *recG*, *hscA*, *barA*, and putative genes STM14_1138, STM14_3334, STM14_4825, and STM14_5184 (Table S6).

Interestingly, most of the common forty-four genes required for *in vitro* and *in vivo* (enteric and systemic infection) fitness have been implicated in vaccine or drug target development against broad spectrum of bacteria. Such as ATP synthase genes (Balemans et al., 2012; Lu et al., 2014), *dam* (Garcia-Del Portillo et al., 1999), *pstB* (Garmory and Titball, 2004), *phoQ* (Miller and Mekalanos, 1998), *ompR* (Dougan et al., 1996), *xerC* (Hur et al., 2011), and *rfbBPN* (Sturm and Timmis, 1986), *manA* (Amineni et al., 2010), *rpe* (Edwards et al., 2004), *lepA* (Patton, 2007), *csgF* (Cegelski et al., 2008), trmE (Shippy and Fadl, 2014a), and SPI-1 and SPI-2 (Matulová et

al., 2012) have been used as vaccine development or drug target (Table S6). Thus, there lies a great potential to explore genes *marBCT*, *envF*, *barA*, *hscA*, *rfaQ*, *rfbI* and putative proteins STM14_1138, STM14_3334, STM14_485, and STM_5184 as novel therapeutic and intervention strategy to curb *Salmonella* infection. However, there is the possibility that mutation only attenuates the bacterium but do not elicit appropriate immune response, which needs further verification.

MATERIAL AND METHODS

Bacterial strains and growth conditions.

Salmonella enterica serovar Typhimurium 14028S, a spontaneous mutant resistant to nalidixic acid (NA) were grown in Luria-Bertani (LB) plate or LB medium (BD Difco, Sparks, MD) on shaking rack at 225 rpm and incubated at 37°C unless otherwise indicated. Nalidixic acid (NA, ICN Biomedicals Inc., Aurora OH, USA) and kanamycin (Km, Shelton Scientific, Inc. CT, USA) were used at 25 µg/ml and 50 µg/ml respectively. *S.* Typhimurium was stored in 50% glycerol at -80°C.

Construction of Transposon mutant library.

To prepare electrocompetent cells, *S.* Typhimurium was grown overnight in 10 ml LB medium with NA and was diluted 100 fold in 10 ml 2xYT (BD Difco, Sparks, MD, USA) medium with NA and incubated for 3 h on a shaking rack. Bacterial cells were washed 6 times with wash solution (10% glycerol). Centrifugation was done at 8,000 rpm for 1 min at refrigeration temperature (4ºC). The bacterial pellet was mixed gently in 60µl of wash solution preventing aeration. One µl of the EZ-Tn5 <KAN-2> Tnp transposome complex (Epicentre

BioTechnologies, Madison, WI, USA) was added to electrocompetent *S.* Typhimurium cells and incubated on ice for 10 min. Then, the bacteria were gently transferred to ice cold cuvette avoiding the formation of any air bubble and electroporated at 2450 V. Immediately, 500 µl of SOC was added and incubated for 90 minutes on a shaking rack at 37°C. The transformants were plated on LB plate with NA and Km. With three electroporations we were able to collect 700,000 Tn5 mutants and stored them in LB medium with 50% glycerol at -80ºC (Figure 1).

***In vitro* growth assay of transposon mutant library.**

In vitro selection of transposon mutant library was done as described by Opijnen and Camilli, (2010)(van Opijnen et al., 2014) with some modifications. Briefly, transposon mutant library was thawed on ice and 300 µl of aliquot were added to 60 ml LB with NA and Km (OD_{600} = 0.131). The library was incubated at 37°C on a shaking rack for 30 min (OD_{600} = 0.135) and centrifuged at 5500 rpm for 8 min at room temperature. The transposon mutant library pellet was mixed in 50 ml 1X phosphate buffer saline (PBS) (OD_{600} = 0.143) and CFU ($4X10^7$/ml) was measured (t_1). Ten ml aliquot were saved from t_1 as an input pool (IP1). Above procedure were repeated to make a technical replicate of IP1 as input pool 2(IP2). An aliquot of 0.5 ml from t_1 was inoculated to 10 ml LB (LB), LB with 3% NaCl (NaCl), LB with 100mM propionate adjusting to pH 7 (PA), LB with 1mM H_2O_2 (H_2O_2). The initial OD_{600} of inoculated medium was 0.009. We then incubated the library on shaking rack at 37°C with variable incubation time ranging from 3.75 h to 7 h (t_2) to a mid-logarithmic. The final OD_{600} of output pool was 0.64 at time point t_2. Input pool and output pool library were centrifuged and the pellet was stored at -80°C for DNA extraction (Figure 1).

***In vitro* survival assay of transposon mutant library.**

To identify genes negatively selected during starvation, an aliquot of 0.5 ml from t_1 was transferred to 10 ml PBS and incubated at 37°C on shaking rack for 12 days. On the 12th day, the tube was centrifuged and the pellet was dissolved in 1 ml PBS. 100 µl aliquot was incubated on LB plate (NA + Km) overnight at 37°C. The cells were collected in PBS and stored at -80°C for DNA extraction. Whereas for survival in pH3, 0.5 ml from t_1 was exposed to LB medium adjusted at pH3 for 30 min at 37°C and immediately transferred to 40 ml PBS. The cells were centrifuged at 8000 rpm for 8 min and pellet was mixed in 1ml PBS. An aliquot of 250 µl was plated on LB plate (NA + Km) overnight at 37°C. Colonies were collected in PBS and stored at -80°C for DNA extraction (Figure 1).

DNA library preparation for Illumina sequencing

Genomic DNA (gDNA) from the bacterial pellet of input library and output library stored at -80°C was extracted using QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) following manufacturer's protocol. The purity and concentration were checked using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) with Qubit Assay Kits (dsDNA BR Assay) following the manufacturer's manual. The sample for Illumina sequencing was prepared according to Dawoud et al. (2014)(Dawoud et al., 2014) with significant modifications. All the DNA primers (Table S5) used for Tn-seq library were custom designed using Primer3 (v. 0.4.0) (Untergasser et al., 2012) and ordered from Integrated DNA Technologies (Coralville, Iowa). The simplified figure for DNA library preparation is shown Figure S1A**.**

Briefly, Tn5-junctions at the right end of transposon was enriched from gDNA extracted from input and output library. The single primer linear extension was done with EZ-Tn5 primer3 using Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA). The 50 µl linear PCR reaction constituted: Nuclease-free water – 40 µl (volume adjusted according to gDNA volume), Thermopol buffer (10X) – 5 µl, dNTPs (2.5 mM each) – 1 µl, EZ-Tn5 primer3 (20 µM) – 1 µl, gDNA library (50 ng/ul) – 2 µl (~100 ng), and Taq DNA polymerase – 1 µl (added during PCR). The PCR cycle consisted of manual hot start with initial denaturation at 95°C for 2 min, addition of Taq DNA polymerase followed by 50 cycles of 95°C for 30 s, 62°C for 45 s, and 72°C for 10 s, and followed by a hold at 4°C. The linear PCR products were then purified with MinElute PCR purification kit (Qiagen, Valencia, CA, USA) and eluted in 10 µl of elution buffer following manufacturer's protocol.

Then deoxycytosine homopolymer tail (C-tail) was added to the linear extension purified PCR product using Terminal Transferase (TdT, New England Biolabs, Ipswich, MA, USA) enzyme following previous protocol (Lazinski and Camilli, 2013). The C-tailing reaction consisted: DNA (linear extension product from linear PCR) – 10 µl, TdT Buffer (10X) – 2 µl, $CoCl_2$ (2.5 mM) – 2 µl, dCTP (10 mM) – 2.4 µl, ddCTP (1mM) – 1 µl, Nuclease-free H_2O – 1.6 µl, and Terminal Transferase – 1 µl, making a total volume of 20 µl. The reaction was incubated at 37°C for 1 h followed by heat inactivation of the enzyme at 75°C for 20 min on a thermocycler. The C- tailed products were purified using MinElute PCR purification kit and eluted to 10 µl.

Subsequently, C-tailed PCR product was enriched with exponential PCR. PCR reaction constituted: nuclease-free H_2O – 35 µl, Thermopol Buffer (10X) – 5 µl, dNTPs (2.5 mM each) – 4 µl, IR2 BC primer (with Illumina adapter and barcode, 10 µM) – 2 µl, HTM primer (with Illumina adapter, 20 µM) – 1 µl, C-tailed DNA – 2 µl, and Taq DNA Polymerase (NEB) – 1 µl,

making a total volume of 50 µl. The manual hot start PCR cycle comprised of 95°C for 2 min, followed by 25 cycles of 95°C for 30s, 58°C for 45s, and 72°C for 20s, trailed by a final extension at 72°C for 10 min.

Finally, the exponential PCR products were pulse heated at 65°C for 15 min and ran on 1.5% agarose gel. Tn-seq library had smear pattern whereas gDNA of *S*. Typhimurium (-ve control) had almost no amplification (Figure S1B). Gel was excised ranging from 300-500 bp and DNA was extracted using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). The purity and concentration of DNA were measured using Qubit 2.0 Fluorometer. An equal amount (~ 10 ng) of DNA (gel-purified products) from each library were mixed together and sent for next generation sequencing, Illumina HiSeq 2000 single end read 100 cycles (Center for Genome Research and Biocomputing, Oregon State University, Corvallis).

Analysis of Transposon sequencing data.

Raw reads from HiSeq Illumina sequencing were de-multiplexed based on the barcode to their respective library using custom Perl script. The barcode and transposon sequence were trimmed off from 5' end. Consequently, the remaining sequence was genomic DNA (junction sequence) with/without poly C-tail. Only 20 bp from the Tn5-junction were kept discarding most of the poly C-tail.

The reads (junction sequence) mostly with genomic sequence were then aligned against *S.* Typhimurium 14028S complete genome (NC_016856.1) using Bowtie version 0.12.7(Langmead and Salzberg, 2012). The aligned sequence (SAM mapping file) were fed to ARTIST pipeline to identify CEGs using Con-ARTIST(Pritchard et al., 2014). Briefly, Tn5 insertion frequency was assigned to the *S.* Typhimurium 14028S genome divided into 100 bp window size. Uncorrected

raw data (non-normalized) of input and output libraries were used to normalize the control data (IP-1) to account for the random loss of mutants in output pool. Then, reads were compared between input and output pool using a Mann-Whitney U test (MWU). The MWU results were used train hidden Markov model (HMM) to predict the likelihood of loci that were not required for growth in either condition, essential under both conditions, enriched in output library and window depleted in output library ($p < 0.01$). The insertions were only considered in the central 80% of the gene to avoid any polar effect of transposon insertion. The cutoffs for depleted loci and enriched loci were >8 fold and >2 fold respectively.

Comparative analysis of conditionally essential genes (CEGs) between *in vitro* and *in vivo* stressors

We compared the *in vitro* essential genes identified in this study including Khatiwara et al. (2012) with the previously identified *in vivo* fitness genes. CEGs for acute infection of mice (A-Mice), macrophage survival (MΦ) (Chan et al., 2005) and persistent infection of mice (P-Mice) (Lawley et al., 2006) were identified in *S.* Typhimurium strain SL1344 backgound. Additionally, *Salmonella* genes required for gastrointestinal colonization of pig, calf and chicken were studied in *S.* Typhimurium strain ST4/74 (Chaudhuri et al., 2013), and for intraperitoneal infection of mice (Sp-Liv) were reported in *S*. Typhimurium strain 14028S background (Silva-Valenzuela et al., 2015). The CEGs of different strain were looked for the orthologous gene in *S*. Typhimurium strain 14028S background using Prokaryotic Genome Analysis Tool (PGAT) (Brittnacher et al., 2011). To get insight into the biological basis of CEGs required for *in vivo* intestinal colonization of pig, calf and chicken were compared with CEGs of *in vitro* host stressors found in the gut (PA, NaCl, pH3, Bile, and LB42). Similarly, for the biological basis of CEGs for *in vivo*

systemic infection (A-Mice, MΦ, P-Mice and Sp-Liv) were compared to *in vitro* macrophage stressors (H_2O_2, NaCl, Starvation, dLB, and pH3). The CEGs that were common in at least one of the *in vitro* host stressors and at least one of *in vivo* infection were only used for the comparative study.

ACKNOWLEDGEMENTS

We would like to thank Arkansas Biosciences Institute (ABI) for funding support and Arkansas High Performance Computing Center (AHPCC) - University of Arkansas for their computational support.

AUTHOR CONTRIBUTION

YMK- Conceived the study; RKM and TJ- Performed the experiment; RKM- Analyzed the data and drafted the manuscript; YMK- Revised the manuscript.

List of Supplementary Tables

Table S1: List of conditionally essential genes (CEGs) of *S.* Typhimurium 14028s to overcome *in vitro* host stressors.

Table S2: Comparison of conditionally essential genes (CEGs) of *S*. Typhimurium with previously identified essential genes of *S*. Typhimurium.

Table S3: Conditionally essential genes (CEGs) common in at least one of the host (pig, calf, and chicken) to causes enteric infection and at least one of the *in vitro* host stressors (LB42, Bile, pH3, PA, and NaCl) encountered during enteric infection.

Table S4: Conditionally essential genes (CEGs) common in at least one of the *in vitro* systemic host stressors (H_2O_2, NaCl, pH3, Starvation and dLB) and at least one of *in vivo* systemic stressors (MΦ; infections (MΦ, A-Mice, P-Mice, Sp-Liv) to cause infection.

Table S5: Oligonucleotides used in this study.

Table S6: Common genes required for fitness in vitro and in vivo (enteric and systemic) conditions.

List of Figures

Figure 1: Schematic overview of experimental design. A highly saturating Tn5 library was constructed through electroporation Ez-Tn5 transposome complex to electrocompetent *S.* Typhimurium 14028s. More than 350,000 Tn5 mutants were collected on LB (Km + NA) plate. Complex Tn5 mutant library (IP1) were selected based on growth (LB medium, 100 mM Propionate in LB medium (PA), 3% NaCl in LB medium, and 1mM Hydrogen peroxide (H_2O_2) in LB medium) and survival (exposed to pH3 in LB medium (pH3) and incubated for 12 days in 1X PBS (Starvation)). Input pool 2 (IP2) is the technical replicate of input pool 1 (IP1). After selection, DNA was extracted from each sample and stored at -20ºC for DNA library preparation.

Figure 2: Demultiplexing of Illumina sequencing reads and correlation between Tn5 mutant libraries. A) Bar graph shows the number of Illumina sequencing reads distribution in each Tn5 libraries after sorting according to the barcode (blue color), unique insertions (orange color), mean reads per unique insertions (grey color), and median reads for each unique insertions (yellow color). B) Scatter plot display Spearman correlation (R^2) among the Tn5 mutant libraries based on read count per 100 bp window across the genome ($p < 0.0001$).

Figure 3: Conditionally essential genes (CEGs) of *S.* Typhimurium 14028S and cluster of orthologous groups (COG). A) Distribution of CEGs identified in the 5 conditions. Numbers inside the bracket indicate number of CEGs. Red dashed box is the CEG (ATP synthase genes) common to all 5 conditions. X-axis: Overlapping set of 339 CEGs B) Functional assignments of CEGs into COG category. Overall is the COG assigned to all the 339 CEGs. (Red asterisk (*): Abundance of COG C in PA was 57.89 %).

Figure 4: Read coverage of select genes in 7 conditions. Y-axis: Numbers is bracket indicates the raw read coverage. A) ATP synthase genes conditionally essential in all the 5 conditions (PA, NaCl, H_2O_2, pH3 and Starvation. B) Gene *mrp* essential in pH3 and Starvation. C) Gene *yiiD* essential in PA only.

Figure 5: Mechanistic basis of enteric infection [*In vitro* vs *in vivo* (Enteric)]. Large square nodes indicates various conditions (studies) and small nodes are fitness genes. Each node (gene) is at least shared by one of the in vitro condition i.e. stressors encountered by *Salmonella* during enteric infection (PA, pH3, NaCl, Bile, and LB42) and at least one of the *in vivo* enteric condition (Pig, Calf, and Chicken).

Figure 6: Mechanistic basis of systemic infection [*In vitro* vs *In vivo* (Systemic)]. Large square nodes indicates various conditions (studies) and small nodes are fitness genes. Each node (gene) is at least shared by one of the *in vitro* condition i.e. stressors encountered by *Salmonella* inside macrophage (NaCl, H_2O_2, pH3, Starvation, and dLB) and at least one of the *in vivo* systemic condition (Pig, Calf, and Chicken).

List of Supplementary Figures

Figure S1: DNA library preparation protocol for Illumina sequencing. A) DNA of Tn5 mutant library was amplified using Tn-specific primer 1 (Ez-Tn5 primer3, Table S1). Then after C-tail was ligated to the 3' end of purified DNA. The C-tailed product was purified and exponential PCR was performed using Tn-specific primer 2 (Barcoded primers, Table S1) and C-tail specific primer (HTM-Primer, Table S1) with Illumina adapter attached to primers. B) Exponentially amplified DNA was than run on 1.5% agarose gel. DNA from 300bp to 500bp was extracted from the gel and sent for Illumina sequencing. [M: Hi-Lo DNA marker; 1, 2, 3, 4: Tn5 mutant libraries; and C: negative control (gDNA of *S*. Typhimurium 14028S)].

Figure S2: Overlay plot displays global view of genome-wide distribution of Tn5 insertion read count for all samples. X-axis: Position on genome; and Y-axis: Number of read count per 100 bp scaled to log10.

Figure S3: Read coverage of select genes. A) Tn5 insertion in Salmonella pathogenicity island 1 (SPI-1) genes encoding type III secretion system (TTSS). Screen shot image produced using Integrative Genomics Viewer (IGV) showing raw read coverage [100-600] in seven conditions. (Blue asterisk: conditionally essential in NaCl and Starvation; and Red asterisk: conditionally essential in Starvation only). B) CpxAR were conditionally essential in starvation only.

Figure S4: Comparison of overlapping set of conditionally essential genes of *S*. Typhimurium 14028S (this study) with essential genome of *S*. Typhimurium SL3261 and *S*. Typhimurium LT2.

Figure S5: Venn diagram shows the number of common genes between *in vitro* vs *in vivo* (Enteric) (135 CEGs, Figure 4) and *in vitro* vs *in vivo* (Systemic) (130 CEGs, Figure 5). Common list of genes (44) are shown in Table S6.

Figure 1: Schematic overview of experimental design. A highly saturating Tn5 library was constructed through electroporation Ez-Tn5 transposome complex to electrocompetent *S.* Typhimurium 14028s. More than 350,000 Tn5 mutants were collected on LB (Km + NA) plate. Complex Tn5 mutant library (IP1) were selected based on growth (LB medium, 100 mM Propionate in LB medium (PA), 3% NaCl in LB medium, and 1mM Hydrogen peroxide (H_2O_2) in LB medium) and survival (exposed to pH3 in LB medium (pH3) and incubated for 12 days in 1X PBS (Starvation)). Input pool 2 (IP2) is the technical replicate of input pool 1 (IP1). After selection, DNA was extracted from each sample and stored at -20ºC for DNA library preparation.

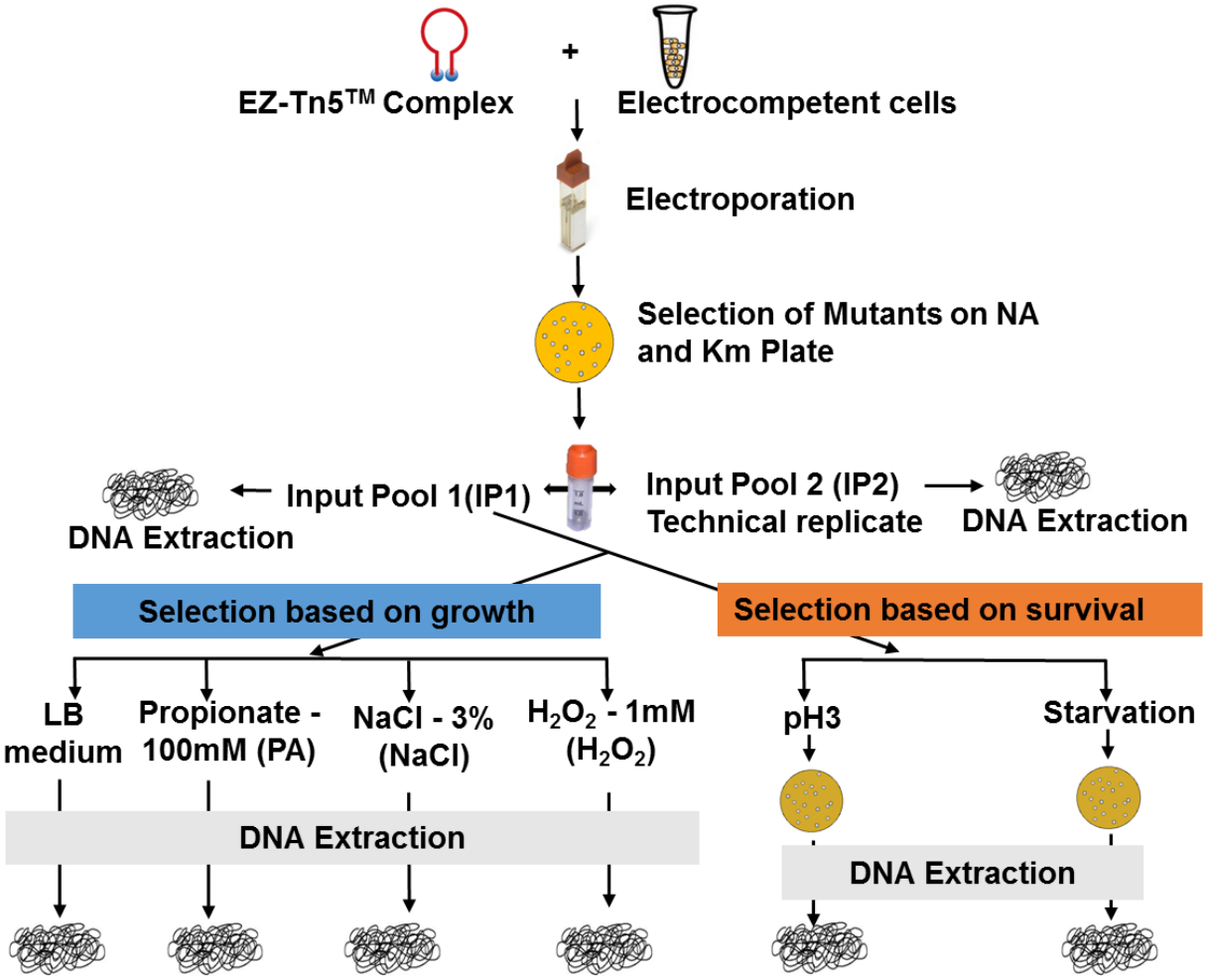


Figure 2: Demultiplexing of Illumina sequencing reads and correlation between Tn5 mutant libraries. A) Bar graph shows the number of Illumina sequencing reads distribution in each Tn5 libraries after sorting according to the barcode (blue color), unique insertions (orange color), mean reads per unique insertions (grey color), and median reads for each unique insertions (yellow color). B) Scatter plot display Spearman correlation (R^2) among the Tn5 mutant libraries based on read count per 100 bp window across the genome ($p < 0.0001$).

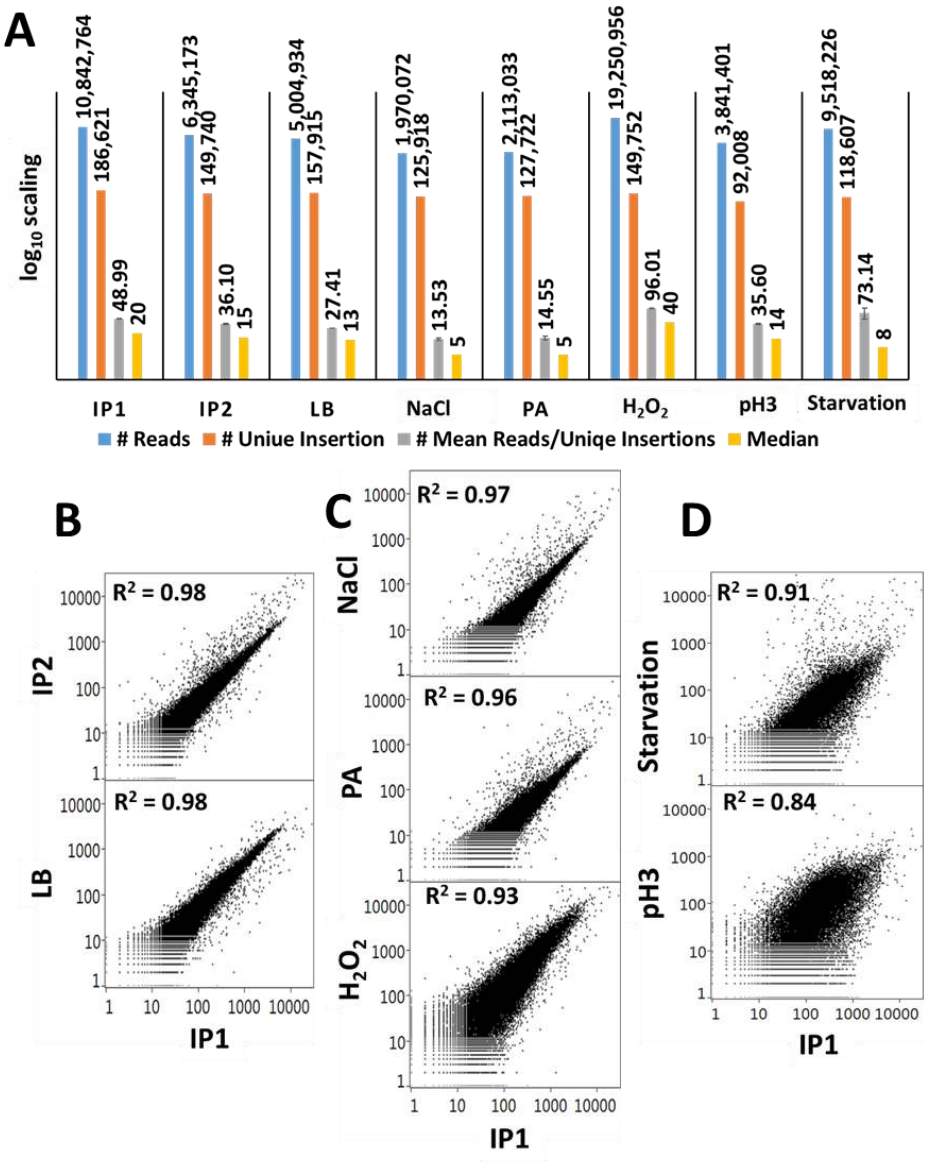


Figure 3: Conditionally essential genes (CEGs) of *S.* Typhimurium 14028S and cluster of orthologous groups (COG). A) Distribution of CEGs identified in the 5 conditions. Numbers inside the bracket indicate number of CEGs. Red dashed box is the CEG (ATP synthase genes) common to all 5 conditions. X-axis: Overlapping set of 339 CEGs B) Functional assignments of CEGs into COG category. Overall is the COG assigned to all the 339 CEGs. (Red asterisk (*): Abundance of COG C in PA was 57.89 %).

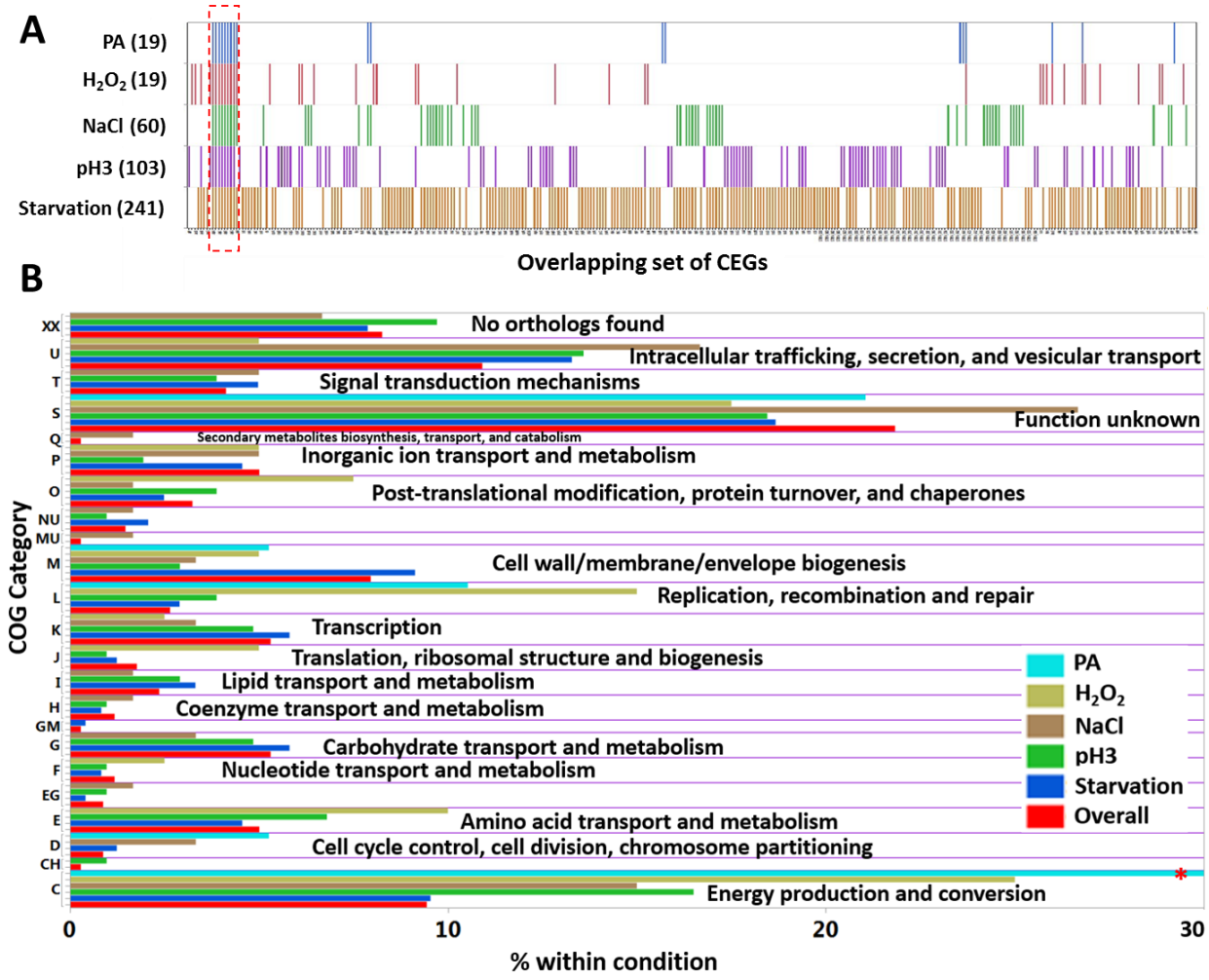


Figure 4: Read coverage of select genes in 7 conditions. Y-axis: Numbers is bracket indicates the raw read coverage. A) ATP synthase genes conditionally essential in all the 5 conditions (PA, NaCl, H_2O_2, pH3 and Starvation. B) Gene *mrp* is essential in pH3 and Starvation only. C) Gene *yiiD* essential in PA only.

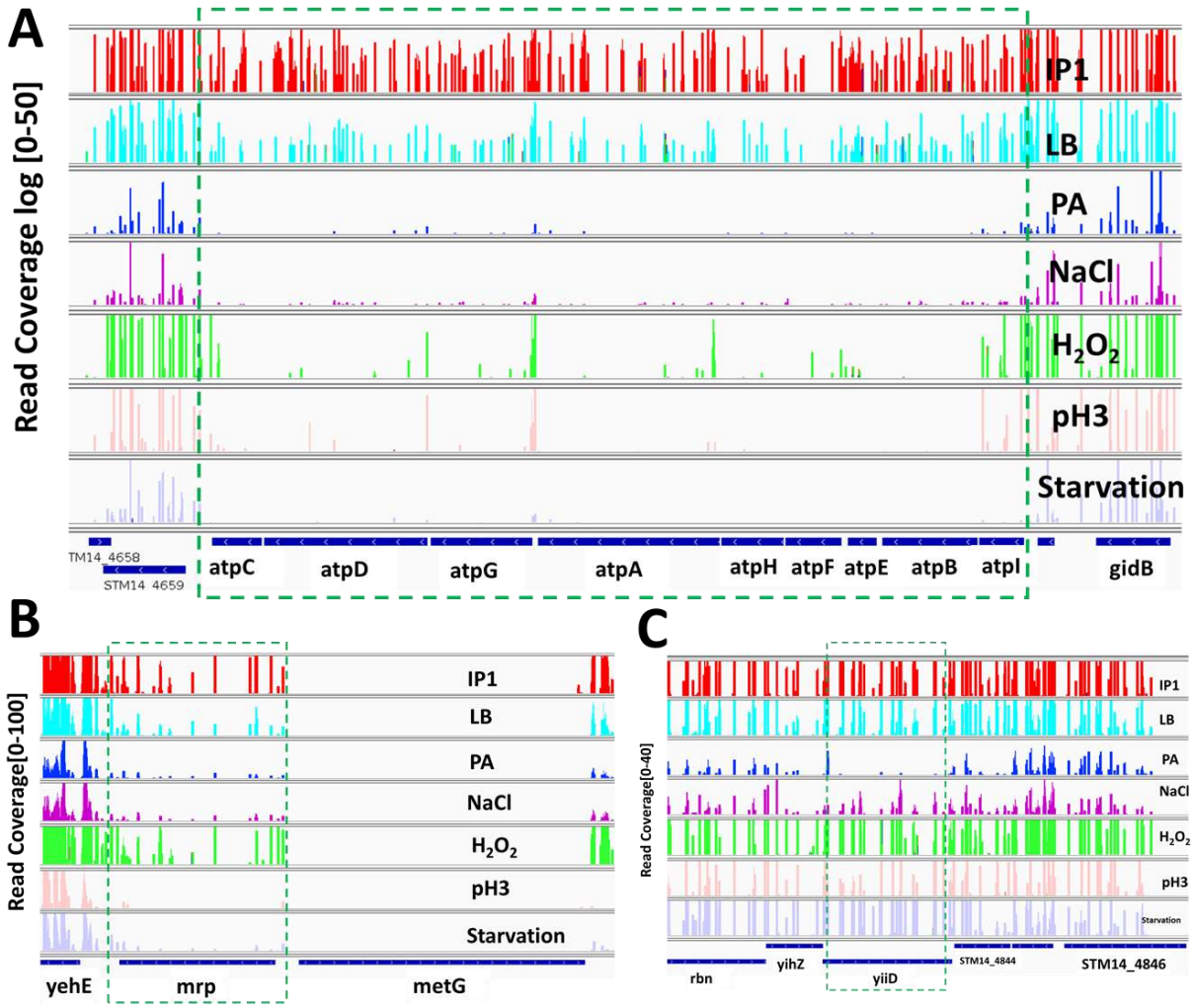


Figure 5: Mechanistic basis of enteric infection [*In vitro* vs *in vivo* (Enteric)]. Large square nodes indicates various conditions (studies) and small nodes are fitness genes. Each node (gene) is at least shared by one of the *in vitro* condition i.e. stressors encountered by *Salmonella* during enteric infection (PA, pH3, NaCl, Bile, and LB42) and at least one of the *in vivo* enteric condition (Pig, Calf, and Chicken).

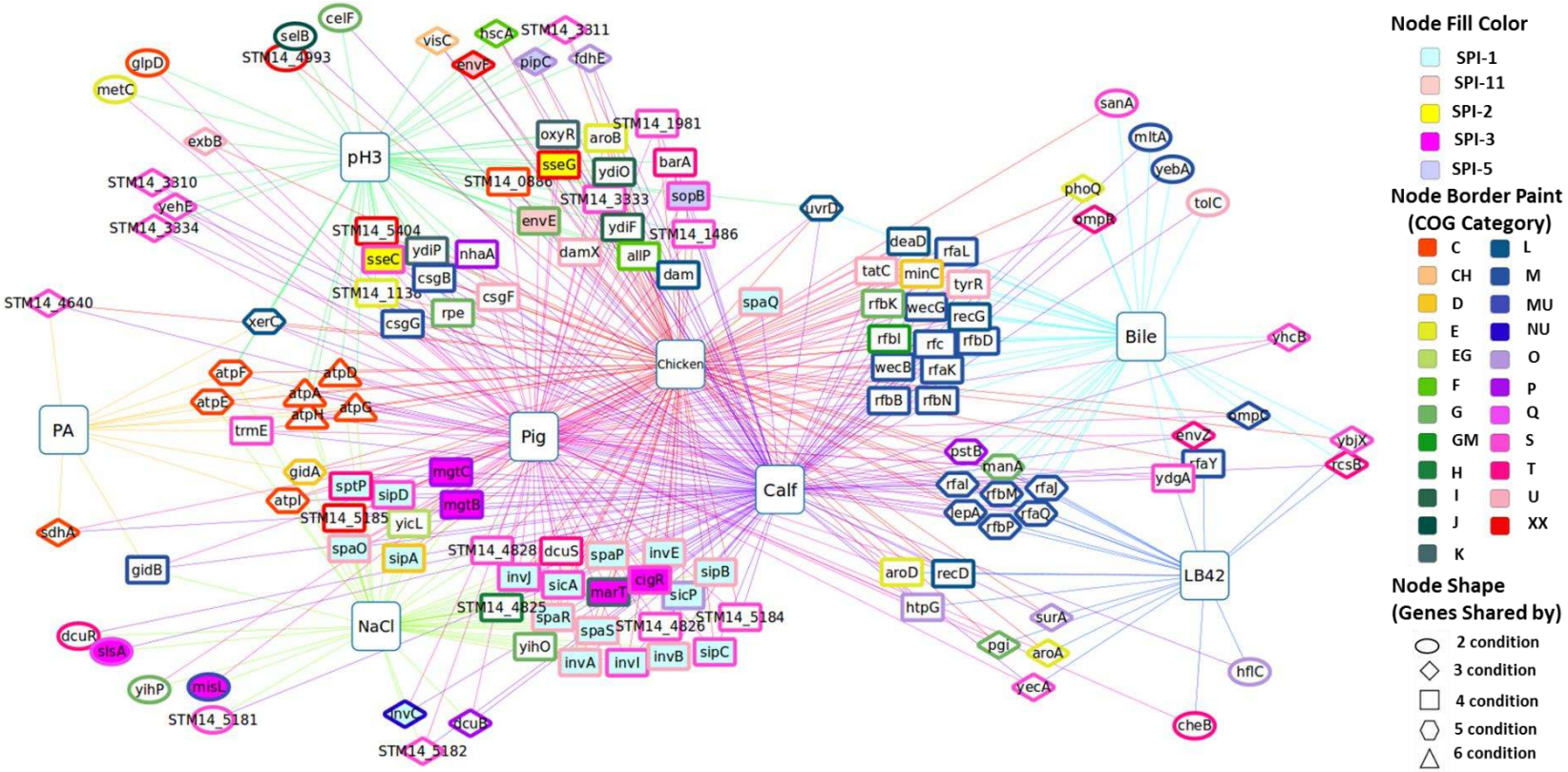


Figure 6: Mechanistic basis of systemic infection [*In vitro* vs *In vivo* (Systemic)]. Large square nodes indicates various conditions (studies) and small nodes are fitness genes. Each node (gene) is at least shared by one of the *in vitro* condition i.e. stressors encountered by *Salmonella* inside macrophage (NaCl, H_2O_2, pH3, Starvation, and dLB) and at least one of the *in vivo* systemic condition (Pig, Calf, and Chicken).

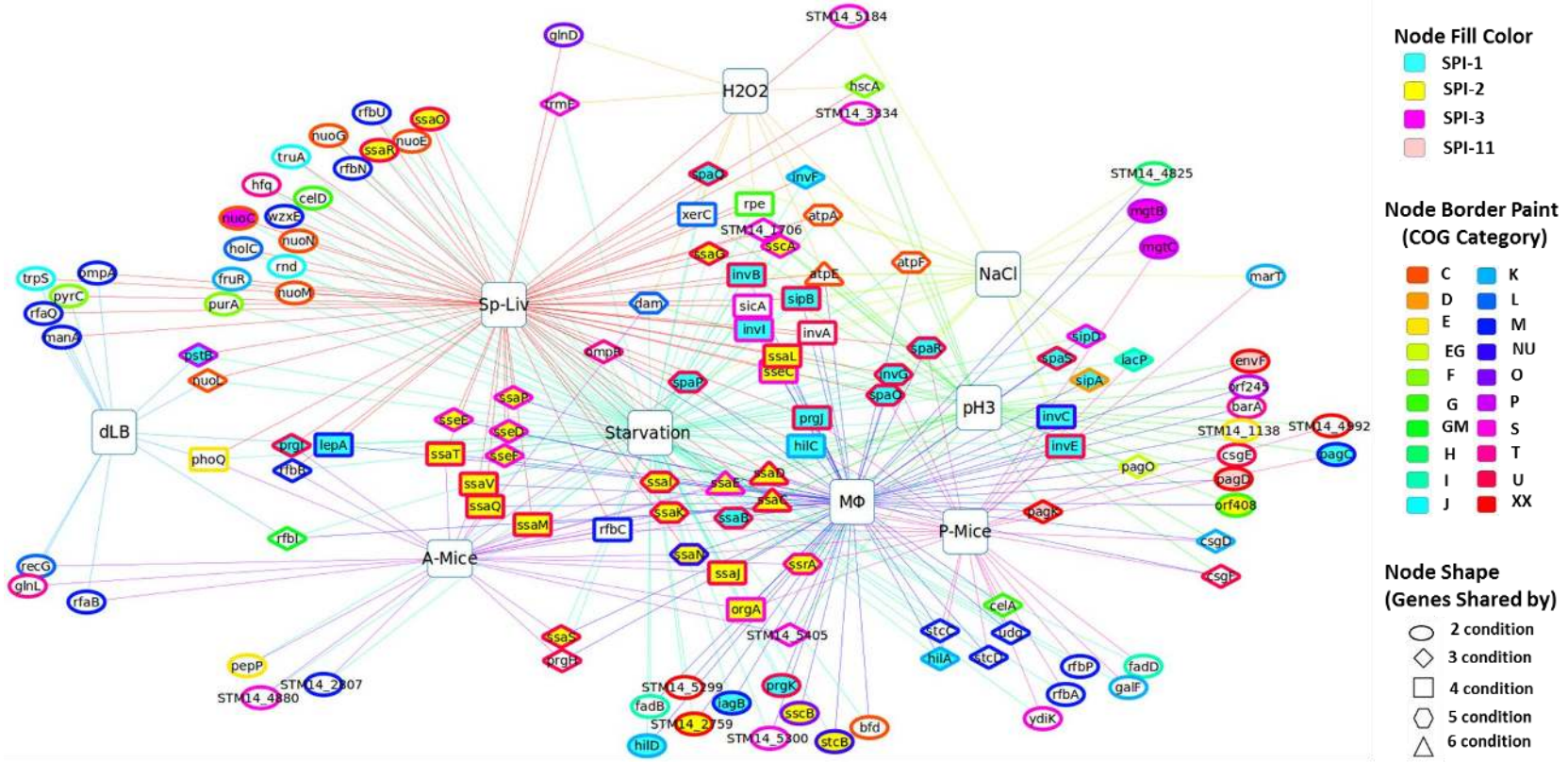


Figure S1: DNA library preparation protocol for Illumina sequencing. A) DNA of Tn5 mutant library was amplified using Tn-specific primer 1 (Ez-Tn5 primer3, Table S1). Then after C-tail was ligated to the 3' end of purified DNA. The C-tailed product was purified and exponential PCR was performed using Tn-specific primer 2 (Barcoded primers, Table S1) and C-tail specific primer (HTM-Primer, Table S1) with Illumina adapter attached to primers. B) Exponentially amplified DNA was than run on 1.5% agarose gel. DNA from 300bp to 500bp was extracted from the gel and sent for Illumina sequencing. [M: Hi-Lo DNA marker; 1, 2, 3, 4: Tn5 mutant libraries; and C: negative control (gDNA of *S*. Typhimurium 14028S)].

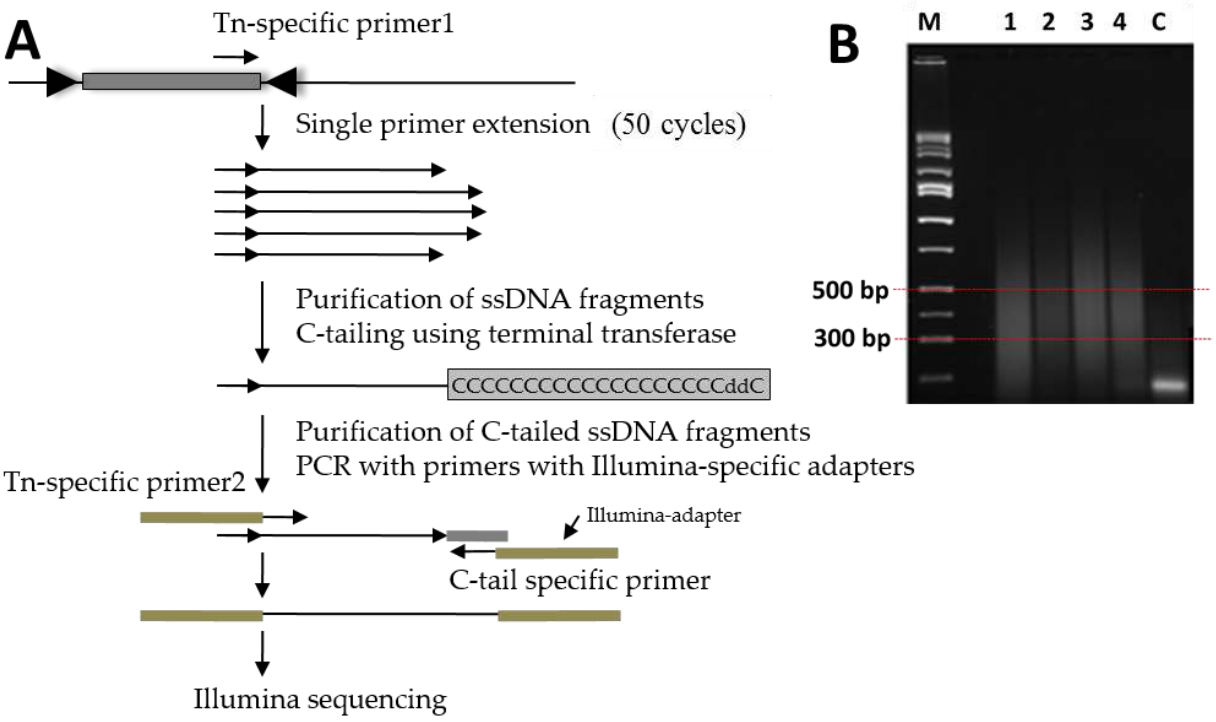


Figure S2: Overlay plot displays global view of genome-wide distribution of Tn5 insertion read count for all samples. X-axis: Position on genome; and Y-axis: Number of read count per 100 bp scaled to log10.

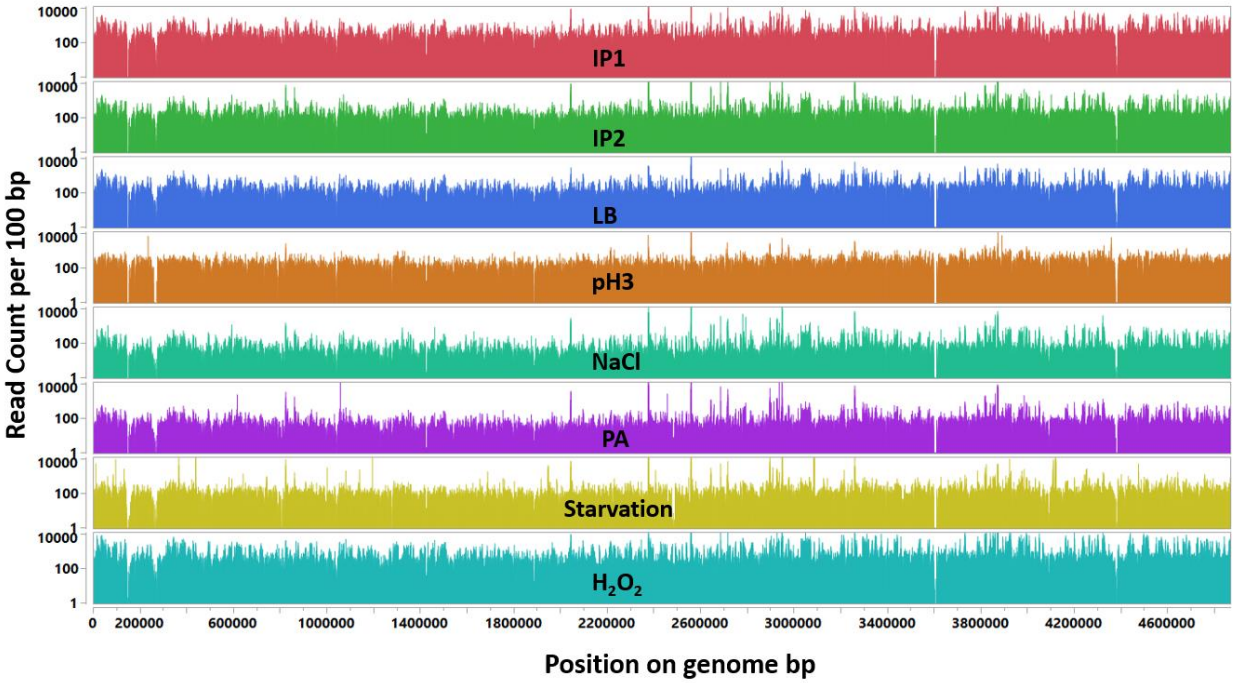


Figure S3: Read coverage of select genes. A) Tn5 insertion in *Salmonella* pathogenicity island 1 (SPI-1) genes encoding type III secretion system (TTSS). Screen shot image produced using Integrative Genomics Viewer (IGV) showing raw read coverage [100-600] in seven conditions. (Blue asterisk: conditionally essential in NaCl and Starvation; and Red asterisk: conditionally essential in Starvation only). B) CpxAR were conditionally essential in starvation only.

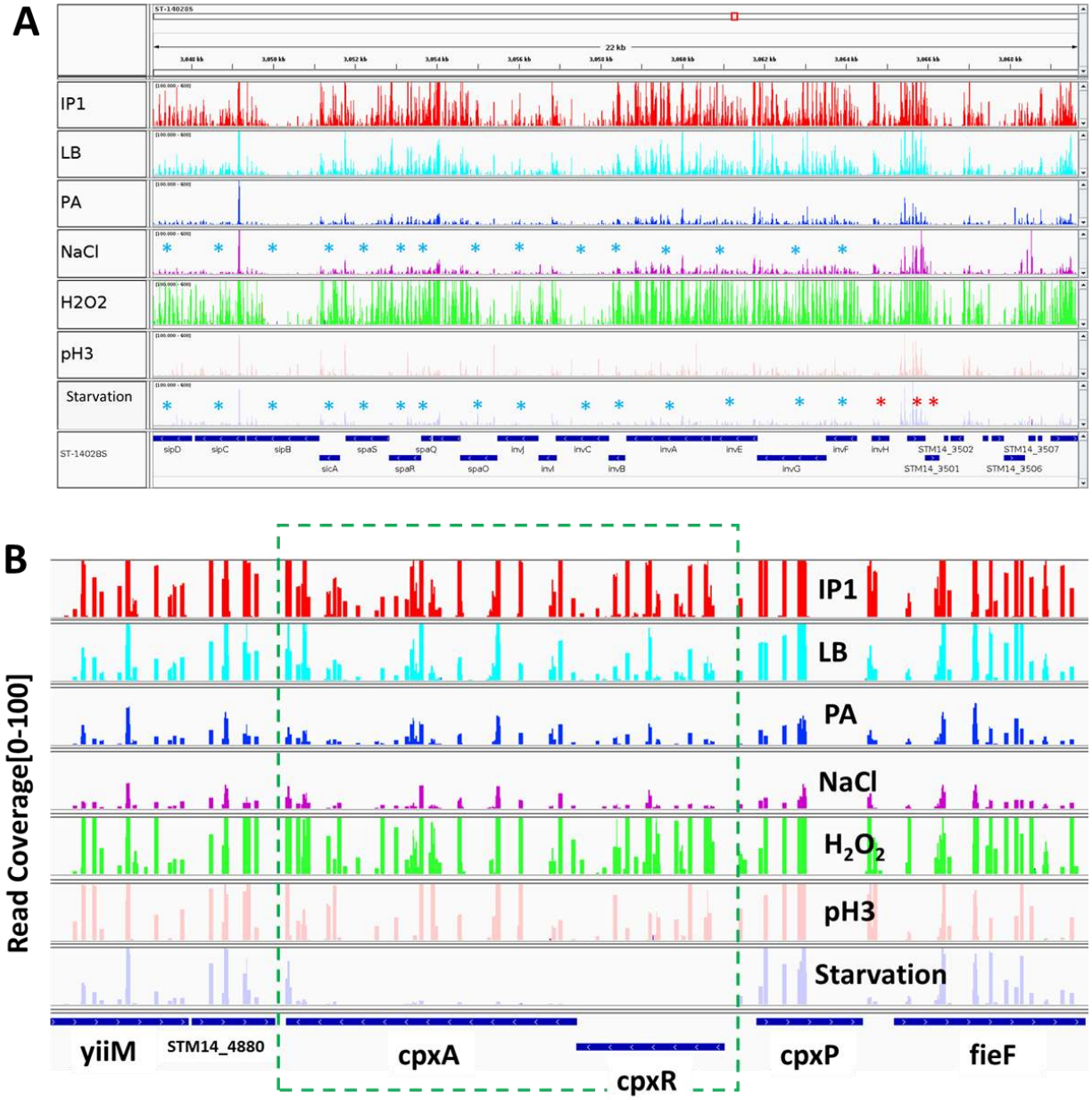


Figure S4: Comparison of overlapping set of conditionally essential genes of *S*. Typhimurium 14028S (this study) with essential genome of *S*. Typhimurium SL3261 and *S*. Typhimurium LT2.

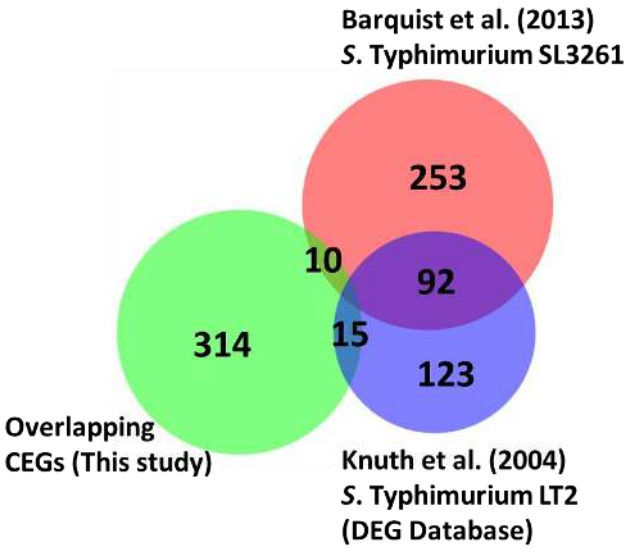


Figure S5: Venn diagram shows the number of common genes between *in vitro* vs *in vivo* (Enteric) (135 CEGs, Figure 4) and *in vitro* vs *in vivo* (Systemic) (130 CEGs, Figure 5). Common list of genes (44) are shown in Table S6.

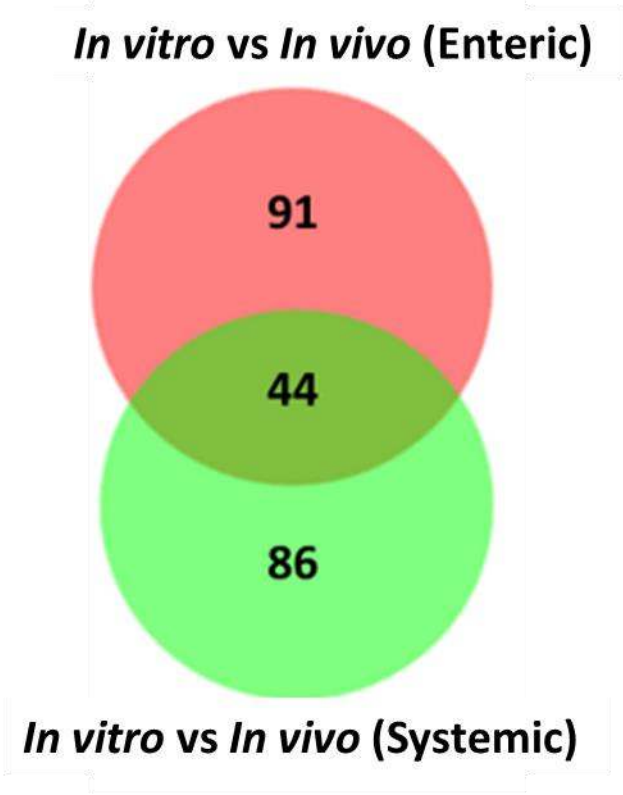


Table S5: Oligonucleotides used in this study.

Primer	Sequence (5'-3')	Barcode	Library
Ez-Tn5 primer3	5’-GATCCTCTAGAGTCGACCTGCAGG CATGCA-3’		
IR2-IS-B7	5’-AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNAGcagatc**C**TCAGGGTT GAGATGTGTATAAGAGACAG-3’	CAGATC	IP1
IR2-IS-B15	5’-AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNAGttcggcTCAGGGTTG AGATGTGTATAAGAGACAG-3’	TTCGGC	IP2
IR2-IS-B8	5’-AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNAGacttgaTCAGGGTTG AGATGTGTATAAGAGACAG-3’	ACTTGA	LB
IR2-IS-B12	5’-AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNAGcttgtaTCAGGGTTGA GATGTGTATAAGAGACAG-3’	CTTGTA	PA
IR2-IS-B10	5’-AATGATACGGCGACCACCGAGATC TACACTCTTTCCCTACACGACGCTCTT CCGATCTNNNNAGtagcttTCAGGGTTGA GATGTGTATAAGAGACAG-3’	TAGCTT	NaCl
IR2-IS-B9	5’-AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNAGgatcagTCAGGGTTGAGA TGTGTATAAGAGACAG-3’	GATCAG	H2O2
IR2-IS-B11	5’-AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNAGggctagTCAGGGTTGAGA TGTGTATAAGAGACAG-3’	GGCTAG	pH3
IR2-IS-B13	5’-AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNAGataacgTCAGGGTTGAGA TGTGTATAAGAGACAG-3’	ATAACG	Starvation
HTM-Primer	5’-CAAGCAGAAGACGGCATACGAGCTC TTCCGATCTGGGGGGGGGGGGGGG-3’		

NNNN: random sequence for efficient cluster analysis. XXXXXX: 6nt sample barcode sequences. For exponential PCR, barcoded and HTM-primer were used. Few wild characters (.C.G.G.T.A.A.G.G.A.A.GAGACAG in Perl script, where “.” is any character) were used in mosaic end reads for identification of the reads containing Tn5-junction sequences.

Table S6: Common genes required for fitness *in vitro* and *in vivo* (enteric and systemic) conditions.

Category	Genes
Salmonella pathogenicity island genes (SPI)	
SPI-1*	*invABCEI, sicA, sipABD,* and *spaOPQRS*
SPI-2*	*sseC*
SPI-3	*marBCT*
SPI-11	*envF*
Non pathogenicity island genes	
Two-component system	*ompR*, phoQ*,* and *barA*
O antigen biosynthetic process	*rfbBPN*
ATP synthase genes	*atpAEF**
Mismatch repair	*dam**
chromosome segregation	*xerC**
Fructose and Mannose Metabolism	*manA**
Carbon metabolism	*rpe**
Homologous recombination	*recG*
ABC transporter	*pstB**
Translational elongation	*lepA**
Iron-sulfur cluster assembly	*hscA*
Others	*csgF*, rfaQ, rfbI,* and *trmE**
Putative Protein	STM14_1138, STM14_3334, STM14_485, and STM_5184

Genes marked with asterisk (*) have been implicated in vaccine development or drug target against a wide range of bacteria.

REFERENCES

Ahmer, B.M., Thomas, M.G., Larsen, R.A., and Postle, K. (1995). Characterization of the exbBD operon of escherichia coli and the role of ExbB and ExbD in TonB function and stability. J. Bacteriol. *177*, 4742-4747.

Amineni, U., Pradhan, D., and Marisetty, H. (2010). In silico identification of common putative drug targets in leptospira interrogans. Journal of Chemical Biology. *3*, 165-173.

Ao, T.T., Feasey, N.A., Gordon, M.A., Keddy, K.H., Angulo, F.J., and Crump, J.A. (2015). Global burden of invasive nontyphoidal salmonella disease, 2010(1). Emerg. Infect. Dis. *21*, 10.3201/eid2106.140999.

Archer, C.D., Wang, X., and Elliott, T. (1993). Mutants defective in the energy-conserving NADH dehydrogenase of salmonella typhimurium identified by a decrease in energy-dependent proteolysis after carbon starvation. Proc. Natl. Acad. Sci. U. S. A. *90*, 9877-9881.

Balemans, W., Vranckx, L., Lounis, N., Pop, O., Guillemont, J., Vergauwen, K., Mol, S., Gilissen, R., Motte, M., Lancois, D*., et al.* (2012). Novel antibiotics targeting respiratory ATP synthesis in gram-positive pathogenic bacteria. Antimicrob. Agents Chemother. *56*, 4131-4139.

Barnhart, M.M., and Chapman, M.R. (2006). Curli biogenesis and function. Annu. Rev. Microbiol. *60*, 131-147.

Barquist, L., Langridge, G.C., Turner, D.J., Phan, M.D., Turner, A.K., Bateman, A., Parkhill, J., Wain, J., and Gardner, P.P. (2013). A comparison of dense transposon insertion libraries in the salmonella serovars typhi and typhimurium. Nucleic Acids Res. *41*, 4549-4564.

Batchelor, E., Walthers, D., Kenney, L.J., and Goulian, M. (2005). The escherichia coli CpxA-CpxR envelope stress response system regulates expression of the porins ompF and ompC. J. Bacteriol. *187*, 5723-5731.

Brittnacher, M.J., Fong, C., Hayden, H.S., Jacobs, M.A., Radey, M., and Rohmer, L. (2011). PGAT: A multistrain analysis resource for microbial genomes. Bioinformatics. *27*, 2429-2430.

Brunet, Y.R., Khodr, A., Logger, L., Aussel, L., Mignot, T., Rimsky, S., and Cascales, E. (2015). H-NS silencing of the salmonella pathogenicity island 6-encoded type VI secretion system limits salmonella enterica serovar typhimurium interbacterial killing. Infect. Immun. *83*, 2738-2750.

Carden, S., Okoro, C., Dougan, G., and Monack, D. (2015). Non-typhoidal salmonella typhimurium ST313 isolates that cause bacteremia in humans stimulate less inflammasome activation than ST19 isolates associated with gastroenteritis. Pathog. Dis. *73*, 10.1093/femspd/ftu023. Epub 2014 Dec 24.

Cegelski, L., Marshall, G.R., Eldridge, G.R., and Hultgren, S.J. (2008). The biology and future prospects of antivirulence therapies. Nature Reviews Microbiology. *6*, 17-27.

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., Morgan, E., Peters, S.E., Pleasance, S.J., Hudson, D.L., Davies, H.M., Wang, J., van Diemen, P.M., Buckley, A.M., and Bowen, A.J. (2013). Comprehensive assignment of roles for salmonella typhimurium genes in intestinal colonization of food-producing animals. PLoS Genet. *9*, e1003456.

Cotter, P.D., and Hill, C. (2003). Surviving the acid test: Responses of gram-positive bacteria to low pH. Microbiol. Mol. Biol. Rev. *67*, 429-53, table of contents.

Crim, S.M., Griffin, P.M., Tauxe, R., Marder, E.P., Gilliss, D., Cronquist, A.B., Cartter, M., Tobin-D'Angelo, M., Blythe, D., and Smith, K. (2015). Preliminary incidence and trends of infection with pathogens transmitted commonly through food—Foodborne diseases active surveillance network, 10 US sites, 2006–2014. MMWR Morb. Mortal. Wkly. Rep. *64*, 495-499.

Darcan, C., Ozkanca, R., Idil, O., and Flint, K.P. (2009). Viable but non-culturable state (VBNC) of escherichia coli related to EnvZ under the effect of pH, starvation and osmotic stress in sea water. Pol. J. Microbiol. *58*, 307-317.

Dawoud, T.M., Jiang, T., Mandal, R.K., Ricke, S.C., and Kwon, Y.M. (2014). Improving the efficiency of transposon mutagenesis in salmonella enteritidis by overcoming host-restriction barriers. Mol. Biotechnol. *56*, 1004-1010.

de Mendonca, J.D., Ely, F., Palma, M.S., Frazzon, J., Basso, L.A., and Santos, D.S. (2007). Functional characterization by genetic complementation of aroB-encoded dehydroquinate synthase from mycobacterium tuberculosis H37Rv and its heterologous expression and purification. J. Bacteriol. *189*, 6246-6252.

Denoncin, K., Vertommen, D., Arts, I.S., Goemans, C.V., Rahuel-Clermont, S., Messens, J., and Collet, J.F. (2014). A new role for escherichia coli DsbC protein in protection against oxidative stress. J. Biol. Chem. *289*, 12356-12364.

Dorsey, C.W., Laarakker, M.C., Humphries, A.D., Weening, E.H., and Bäumler, A.J. (2005). Salmonella enterica serotype typhimurium MisL is an intestinal colonization factor that binds fibronectin. Mol. Microbiol. *57*, 196-211.

Dougan, G., Chatfield, S.N., Higgins, C.F., and Dorman, C.J. (1996). Vaccines Comprising Attenuated Salmonella having Mutations in the Ompr Genes.

D'Urzo, N., Martinelli, M., Pezzicoli, A., De Cesare, V., Pinto, V., Margarit, I., Telford, J.L., Maione, D., and Members of the DEVANI Study Group. (2014). Acidic pH strongly enhances in vitro biofilm formation by a subset of hypervirulent ST-17 streptococcus agalactiae strains. Appl. Environ. Microbiol. *80*, 2176-2185.

Edwards, A., Dharamsi, A., Vedadi, M., Domagala, M., Pinder, B., Alam, M.Z., Nethery, K., Houston, S., Kimber, M., and Clarke, T. (2004). Crystal structures of bacterial ribulose-phosphate 3-epimerases.

Feasey, N.A., Dougan, G., Kingsley, R.A., Heyderman, R.S., and Gordon, M.A. (2012). Invasive non-typhoidal salmonella disease: An emerging and neglected tropical disease in africa. The Lancet. *379*, 2489-2499.

Figueira, R., Watson, K.G., Holden, D.W., and Helaine, S. (2013). Identification of salmonella pathogenicity island-2 type III secretion system effectors involved in intramacrophage replication of S. enterica serovar typhimurium: Implications for rational vaccine design. MBio. *4*, e00065-13.

Foss, E.J., Radulovic, D., Shaffer, S.A., Ruderfer, D.M., Bedalov, A., Goodlett, D.R., and Kruglyak, L. (2007). Genetic basis of proteome variation in yeast. Nat. Genet. *39*, 1369-1375.

Freeman, Z.N., Dorus, S., and Waterfield, N.R. (2013). The KdpD/KdpE two-component system: Integrating K homeostasis and virulence. PLoS Pathog. *9*, e1003201.

Fu, J., Keurentjes, J.J., Bouwmeester, H., America, T., Verstappen, F.W., Ward, J.L., Beale, M.H., De Vos, R.C., Dijkstra, M., and Scheltema, R.A. (2009). System-wide molecular evidence for phenotypic buffering in arabidopsis. Nat. Genet. *41*, 166-167.

Garcia-Del Portillo, F., Pucciarelli, M.G., and Casadesus, J. (1999). DNA adenine methylase mutants of salmonella typhimurium show defects in protein secretion, cell invasion, and M cell cytotoxicity. Proc. Natl. Acad. Sci. U. S. A. *96*, 11578-11583.

Garmory, H.S., and Titball, R.W. (2004). ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. Infect. Immun. *72*, 6757-6763.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., and Furlotte, N. (2011). Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genet. *7*, e1001393.

Gibson, D.L., White, A.P., Snyder, S.D., Martin, S., Heiss, C., Azadi, P., Surette, M., and Kay, W.W. (2006). Salmonella produces an O-antigen capsule regulated by AgfD and important for environmental persistence. J. Bacteriol. *188*, 7722-7730.

Gogol, E.B., Rhodius, V.A., Papenfort, K., Vogel, J., and Gross, C.A. (2011). Small RNAs endow a transcriptional activator with essential repressor functions for single-tier control of a global stress regulon. Proc. Natl. Acad. Sci. U. S. A. *108*, 12875-12880.

Gunn, J.S., Alpuche-Aranda, C.M., Loomis, W.P., Belden, W.J., and Miller, S.I. (1995). Characterization of the salmonella typhimurium pagC/pagD chromosomal region. J. Bacteriol. *177*, 5040-5047.

Ha, S., Maciorowski, K., Kwon, Y., Jones, F., and Ricke, S. (1998). Survivability of indigenous microflora and a salmonella typhimurium marker strain in poultry mash treated with buffered propionic acid. Anim. Feed Sci. Technol. *75*, 145-155.

Hansen, S., Lewis, K., and Vulic, M. (2008). Role of global regulators and nucleotide metabolism in antibiotic tolerance in escherichia coli. Antimicrob. Agents Chemother. *52*, 2718-2726.

Hartman, C.E., Samuels, D.J., and Karls, A.C. (2016). Modulating salmonella typhimurium's response to a changing environment through bacterial enhancer-binding proteins and the RpoN regulon. Frontiers in Molecular Biosciences. *3*, 41.

Hasan, C.M.M., and Shimizu, K. (2008). Effect of temperature up-shift on fermentation and metabolic characteristics in view of gene expressions in escherichia coli. Microbial Cell Factories. *7*, 1.

Henderson, M.L., and Kreuzer, K.N. (2015). Functions that protect escherichia coli from tightly bound DNA-protein complexes created by mutant EcoRII methyltransferase. PLoS One. *10*, e0128092.

Hensel, M. (2000). Salmonella pathogenicity island 2. Mol. Microbiol. *36*, 1015-1023.

Hur, J., Xiang, Z., Feldman, E.L., and He, Y. (2011). Ontology-based brucella vaccine literature indexing and systematic analysis of gene-vaccine association network. BMC Immunology. *12*, 1.

Hwang, S., Jeon, B., Yun, J., and Ryu, S. (2011). Roles of RpoN in the resistance of campylobacter jejuni under various stress conditions. BMC Microbiol. *11*, 207-2180-11-207.

Imlay, J.A., and Linn, S. (1986). Bimodal pattern of killing of DNA-repair-defective or anoxically grown escherichia coli by hydrogen peroxide. J. Bacteriol. *166*, 519-527.

Jakomin, M., Chessa, D., Baumler, A.J., and Casadesus, J. (2008). Regulation of the salmonella enterica std fimbrial operon by DNA adenine methylation, SeqA, and HdfR. J. Bacteriol. *190*, 7406-7413.

Jensen, L.J., Julien, P., Kuhn, M., von Mering, C., Muller, J., Doerks, T., and Bork, P. (2008). eggNOG: Automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. *36*, D250-4.

Jitprasutwit, S., Ong, C., Juntawieng, N., Ooi, W.F., Hemsley, C.M., Vattanaviboon, P., Titball, R.W., Tan, P., and Korbsrisate, S. (2014). Transcriptional profiles of burkholderia pseudomallei reveal the direct and indirect roles of sigma E under oxidative stress conditions. BMC Genomics. *15*, 1.

Kang, J., and Blaser, M.J. (2006). UvrD helicase suppresses recombination and DNA damage-induced deletions. J. Bacteriol. *188*, 5450-5459.

Kato, A., and Groisman, E.A. (2008). The PhoQ/PhoP regulatory network of Salmonella enterica. In Bacterial Signal Transduction: Networks and Drug Targets, Springer), pp. 7-21.

Kenyon, W.J., Sayers, D.G., Humphreys, S., Roberts, M., and Spector, M.P. (2002). The starvation-stress response of salmonella enterica serovar typhimurium requires σE-, but not CpxR-regulated extracytoplasmic functions. Microbiology. *148*, 113-122.

Khatiwara, A., Jiang, T., Sung, S.S., Dawoud, T., Kim, J.N., Bhattacharya, D., Kim, H.B., Ricke, S.C., and Kwon, Y.M. (2012). Genome scanning for conditionally essential genes in salmonella enterica serotype typhimurium. Appl. Environ. Microbiol. *78*, 3098-3107.

Klein, J.R., Fahlen, T.F., and Jones, B.D. (2000). Transcriptional organization and function of invasion genes within salmonella enterica serovar typhimurium pathogenicity island 1, including the prgH, prgI, prgJ, prgK, orgA, orgB, and orgC genes. Infect. Immun. *68*, 3368-3376.

Knuth, K., Niesalla, H., Hueck, C.J., and Fuchs, T.M. (2004). Large-scale identification of essential salmonella genes by trapping lethal insertions. Mol. Microbiol. *51*, 1729-1744.

Kong, Q., Yang, J., Liu, Q., Alamuri, P., Roland, K.L., and Curtiss, R.,3rd. (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.

Kröger, C., Colgan, A., Srikumar, S., Händler, K., Sivasankaran, S.K., Hammarlöf, D.L., Canals, R., Grissom, J.E., Conway, T., and Hokamp, K. (2013). An infection-relevant transcriptomic compendium for salmonella enterica serovar typhimurium. Cell Host & Microbe. *14*, 683-695.

Kroupitski, Y., Brandl, M., Pinto, R., Belausov, E., Tamir-Ariel, D., Burdman, S., and Sela, S. (2013). Identification of salmonella enterica genes with a role in persistence on lettuce leaves during cold storage by recombinase-based in vivo expression technology. Phytopathology. *103*, 362-372.

Krulwich, T.A., Sachs, G., and Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. Nature Reviews Microbiology. *9*, 330-343.

Kwon, Y.M., Ricke, S.C., and Mandal, R.K. (2016). Transposon sequencing: Methods and expanding applications. Appl. Microbiol. Biotechnol. *100*, 31-43.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with bowtie 2. Nature Methods. *9*, 357-359.

Lawhon, S.D. (2002). Genetic and environmental regulation of virulence genes in salmonella enterica serovar typhimurium.

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.

Lazinski, D.W., and Camilli, A. (2013). Homopolymer tail-mediated ligation PCR: A streamlined and highly efficient method for DNA cloning and library construction. BioTechniques. *54*, 25-34.

Lee, E., and Groisman, E.A. (2012). Tandem attenuators control expression of the salmonella mgtCBR virulence operon. Mol. Microbiol. *86*, 212-224.

Lee, Y.H., Kim, S., Helmann, J.D., Kim, B., and Park, Y.K. (2013). RaoN, a small RNA encoded within salmonella pathogenicity island-11, confers resistance to macrophage-induced stress. Microbiology. *159*, 1366-1378.

Lee, E.J., Choi, J., and Groisman, E.A. (2014). Control of a salmonella virulence operon by proline-charged tRNA(pro). Proc. Natl. Acad. Sci. U. S. A. *111*, 3140-3145.

Leonhartsberger, S., Korsa, I., and Bock, A. (2002). The molecular biology of formate metabolism in enterobacteria. J. Mol. Microbiol. Biotechnol. *4*, 269-276.

Liu, Y., Wang, Y., Walsh, T.R., Yi, L., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., and Huang, X. (2015). Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in china: A microbiological and molecular biological study. The Lancet Infectious Diseases.

López-Garrido, J., and Casadesús, J. (2010). The DamX protein of escherichia coli and salmonella enterica. Gut Microbes. *1*, 285-288.

Lostroh, C.P., and Lee, C.A. (2001). The salmonella pathogenicity island-1 type III secretion system. Microb. Infect. *3*, 1281-1291.

Low, D.A., Weyand, N.J., and Mahan, M.J. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect. Immun. *69*, 7197-7204.

Lu, P., Lill, H., and Bald, D. (2014). ATP synthase in mycobacteria: Special features and implications for a function as drug target. Biochimica Et Biophysica Acta (BBA)-Bioenergetics. *1837*, 1208-1218.

Lüttmann, D., Göpel, Y., and Görke, B. (2012). The phosphotransferase protein EIIANtr modulates the phosphate starvation response through interaction with histidine kinase PhoR in escherichia coli. Mol. Microbiol. *86*, 96-110.

Ma, Q., and Wood, T.K. (2011). Protein acetylation in prokaryotes increases stress resistance. Biochem. Biophys. Res. Commun. *410*, 846-851.

Majowicz, S.E., Musto, J., Scallan, E., Angulo, F.J., Kirk, M., O'Brien, S.J., Jones, T.F., Fazil, A., Hoekstra, R.M., and International Collaboration on Enteric Disease 'Burden of Illness' Studies. (2010). The global burden of nontyphoidal salmonella gastroenteritis. Clin. Infect. Dis. *50*, 882-889.

Marcus, E.A., Sachs, G., and Scott, D.R. (2013). The role of ExbD in periplasmic pH homeostasis in helicobacter pylori. Helicobacter. *18*, 363-372.

Matulová, M., Havlickova, H., Sisak, F., and Rychlik, I. (2012). Vaccination of chickens with salmonella pathogenicity island (SPI) 1 and SPI2 defective mutants of salmonella enterica serovar enteritidis. Vaccine. *30*, 2090-2097.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., and Du, F. (2001). Complete genome sequence of salmonella enterica serovar typhimurium LT2. Nature. *413*, 852-856.

McKenna, M. (2013). CDC threat report:'We will soon be in a post-antibiotic era'. Wired Magazine.

Medini, D., Serruto, D., Parkhill, J., Relman, D.A., Donati, C., Moxon, R., Falkow, S., and Rappuoli, R. (2008). Microbiology in the post-genomic era. Nature Reviews Microbiology. *6*, 419-430.

Michalik, S., Liebeke, M., Zühlke, D., Lalk, M., Bernhardt, J., Gerth, U., and Hecker, M. (2009). Proteolysis during long-term glucose starvation in staphylococcus aureus COL. Proteomics. *9*, 4468-4477.

Mikheil, D.M., Shippy, D.C., Eakley, N.M., Okwumabua, O.E., and Fadl, A.A. (2012). Deletion of gene encoding methyltransferase (gidB) confers high-level antimicrobial resistance in salmonella. J. Antibiot. *65*, 185-192.

Miller, S.I., and Mekalanos, J.J. (1998). Salmonella Vaccines.

Mitchell, A., Romano, G.H., Groisman, B., Yona, A., Dekel, E., Kupiec, M., Dahan, O., and Pilpel, Y. (2009). Adaptive prediction of environmental changes by microorganisms. Nature. *460*, 220-224.

Murinova, S., and Dercova, K. (2014). Response mechanisms of bacterial degraders to environmental contaminants on the level of cell walls and cytoplasmic membrane. Int. J. Microbiol. *2014*, 873081.

Nava, G., Bielke, L., Callaway, T., and Castaneda, M. (2005). Probiotic alternatives to reduce gastrointestinal infections: The poultry experience. Animal Health Research Reviews. *6*, 105-118.

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., and Wong, A. (2011). Phenotypic landscape of a bacterial cell. Cell. *144*, 143-156.

Olekhnovich, I.N., and Kadner, R.J. (2007). Role of nucleoid-associated proteins hha and H-NS in expression of salmonella enterica activators HilD, HilC, and RtsA required for cell invasion. J. Bacteriol. *189*, 6882-6890.

Patton, B.S. (2007). Applications and mechanisms of colicin E1 ProQuest).

Pritchard, J.R., Chao, M.C., Abel, S., Davis, B.M., Baranowski, C., Zhang, Y.J., Rubin, E.J., and Waldor, M.K. (2014). ARTIST: High-resolution genome-wide assessment of fitness using transposon-insertion sequencing. PLoS Genet. *10*, e1004782.

Reid, A.N., Pandey, R., Palyada, K., Whitworth, L., Doukhanine, E., and Stintzi, A. (2008). Identification of campylobacter jejuni genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis. Appl. Environ. Microbiol. *74*, 1598-1612.

Riordan, J.T., Tietjen, J.A., Walsh, C.W., Gustafson, J.E., and Whittam, T.S. (2010). Inactivation of alternative sigma factor 54 (RpoN) leads to increased acid resistance, and alters locus of enterocyte effacement (LEE) expression in escherichia coli O157: H7. Microbiology. *156*, 719-730.

Rosenberger, C.M., and Finlay, B.B. (2003). Phagocyte sabotage: Disruption of macrophage signalling by bacterial pathogens. Nature Reviews Molecular Cell Biology. *4*, 385-396.

Rowley, G., Spector, M., Kormanec, J., and Roberts, M. (2006). Pushing the envelope: Extracytoplasmic stress responses in bacterial pathogens. Nature Reviews Microbiology. *4*, 383-394.

Schakermann, M., Langklotz, S., and Narberhaus, F. (2013). FtsH-mediated coordination of lipopolysaccharide biosynthesis in escherichia coli correlates with the growth rate and the alarmone (p)ppGpp. J. Bacteriol. *195*, 1912-1919.

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

Shippy, D.C., and Fadl, A.A. (2014a). Bacterial tRNA modification enzymes: Potential role in biology and virulence. Biology and Medicine. *2014*.

Shippy, D.C., and Fadl, A.A. (2014b). tRNA modification enzymes GidA and MnmE: Potential role in virulence of bacterial pathogens. International Journal of Molecular Sciences. *15*, 18267-18280.

Silva-Valenzuela, C.A., Molina-Quiroz, R.C., Desai, P., Valenzuela, C., Porwollik, S., Zhao, M., Hoffman, R.M., Andrews-Polymenis, H., Contreras, I., and Santiviago, C.A. (2015). Analysis of two complementary single-gene deletion mutant libraries of salmonella typhimurium in intraperitoneal infection of BALB/c mice. Frontiers in Microbiology. *6*.

Sleator, R.D., and Hill, C. (2002). Bacterial osmoadaptation: The role of osmolytes in bacterial stress and virulence. FEMS Microbiol. Rev. *26*, 49-71.

Smith, J.L. (2003). The role of gastric acid in preventing foodborne disease and how bacteria overcome acid conditions. Journal of Food Protection®. *66*, 1292-1303.

Spencer, B. (2015). Superbug resistant to ALL antibiotics reaches europe: Danish patient becomes infected with untreatable form of salmonella that is 'probably already in britain'. *2016*.

Sturm, S., and Timmis, K. (1986). Cloning of the rfb gene region of shigella dysenteriae 1 and construction of an rfb-rfp gene cassette for the development of lipopolysaccharide-based live anti-dysentery vaccines. Microb. Pathog. *1*, 289-297.

Tagkopoulos, I., Liu, Y.C., and Tavazoie, S. (2008). Predictive behavior within microbial genetic networks. Science. *320*, 1313-1317.

Tükel, Ç, Raffatellu, M., Humphries, A.D., Wilson, R.P., Andrews-Polymenis, H.L., Gull, T., Figueiredo, J.F., Wong, M.H., Michelsen, K.S., and Akçelik, M. (2005). CsgA is a pathogen-associated molecular pattern of salmonella enterica serotype typhimurium that is recognized by toll-like receptor 2. Mol. Microbiol. *58*, 289-304.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., and Rozen, S.G. (2012). Primer3—new capabilities and interfaces. Nucleic Acids Res. *40*, e115-e115.

van der Heijden, J., Bosman, E.S., Reynolds, L.A., and Finlay, B.B. (2015). Direct measurement of oxidative and nitrosative stress dynamics in salmonella inside macrophages. Proc. Natl. Acad. Sci. U. S. A. *112*, 560-565.

van Opijnen, T., Lazinski, D.W., and Camilli, A. (2014). Genome-Wide fitness and genetic interactions determined by tn-seq, a High-Throughput massively parallel sequencing method for microorganisms. Current Protocols in Molecular Biology. , 7.16. 1-7.16. 24.

van Opijnen, T., and Camilli, A. (2012). A fine scale phenotype-genotype virulence map of a bacterial pathogen. Genome Res. *22*, 2541-2551.

Wakamatsu, T., Kim, K., Uemura, Y., Nakagawa, N., Kuramitsu, S., and Masui, R. (2011). Role of RecJ-like protein with 5'-3' exonuclease activity in oligo(deoxy)nucleotide degradation. J. Biol. Chem. *286*, 2807-2816.

Weber, A., and Jung, K. (2002). Profiling early osmostress-dependent gene expression in escherichia coli using DNA macroarrays. J. Bacteriol. *184*, 5502-5507.

World Health Organization (WHO). WHO Antimicrobial Resistance: Global Report on Surveillance, April 2014.

Yin, J., Xia, J., Tao, M., Xu, L., Li, Q., Geng, S., and Jiao, X. (2016). Construction and characterization of a cigR deletion mutant of salmonella enterica serovar pullorum. Avian Pathol. , 1-22.

Zhang, R., Ou, H.Y., and Zhang, C.T. (2004). DEG: A database of essential genes. Nucleic Acids Res. *32*, D271-2.

Zhou, L., Zhao, M., Wolf, R.Z., Graham, D.E., and Georgiou, G. (2009). Transcriptional regulation of the escherichia coli gene rraB, encoding a protein inhibitor of RNase E. J. Bacteriol. *191*, 6665-6674.

CHAPTER 3:

Global screening of *Salmonella* genes for desiccation survival

Rabindra K. Mandal[1] and Young Min Kwon[1, 2*]

[1]Department of Poultry Science, University of Arkansas, Fayetteville, AR, 72701

[2]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, 72701

*Corresponding author:

Department of Poultry Science

College of Agricultural, Food and Life Sciences

University of Arkansas

Fayetteville, AR 72701

Phone: (479) 575-4935

Email ykwon@uark.edu

ABSTRACT

Salmonella spp., one of the most common foodborne bacterial pathogens, has the ability to survive under desiccation condition in foods and food processing facilities for years and is the most notable and frequent cause of contamination in low water activity foods. The response of *Salmonella* to desiccation stress is complex involving immediate physiological actions as well as coordinated genetic responses. However, the exact mechanisms of *Salmonella* to resist desiccation stress remain to be fully elucidated. For the first time to our knowledge, we screened a genome-saturating transposon (Tn5) library of *Salmonella* Typhimurium (*S*. Typhimurium) 14028S under the *in vitro* desiccation stress using transposon sequencing (Tn-seq). We identified 61 genes and 6 intergenic regions required to overcome desiccation stress. *Salmonella* desiccation resistance genes were mostly related to energy production and conversion; cell wall/membrane/envelope biogenesis, inorganic ion transport and metabolism; regulation of biological process; DNA metabolic process; ABC transporters; and two component system. More than 20% of the *Salmonella* desiccation resistance genes encode either putative or hypothetical proteins. Phenotypic evaluation of 12 single gene knockout mutants during desiccation survival showed 3 mutants (*atpH*, *atpG*, and *corA*) showed significantly ($p < 0.02$) reduced survival as compared to the wild type. Thus, our study was able to provide new insights on the molecular mechanisms utilized by *Salmonella* for survival against desiccation stress. The findings might be further exploited to develop effective control strategies against *Salmonella* contamination in low water activity foods and food processing facilities.

Keywords: *Salmonella*; Desiccation stress; Genetic determinants; Tn-seq

INTRODUCTION

Salmonella is one of the most common causes of foodborne illness worldwide. It is capable of withstanding a spectrum of hostile milieus such as desiccation found in natural and food industry settings (1). *Salmonella* can persist in the low water activity (low-a_w) environment for extended periods of time. *Salmonella* are able to survive from several weeks and months, even to years in dry foods (chocolate, hard cheese, dried eggs, infant dried milk, salami, halva, almonds kernels, pecans, dry confectionery raw materials, and peanut-flavored candy) (2-8) and dry surfaces (desiccated paper discs, plastics , and eggshells) (9). Globally, there were 7,315 cases of bacterial outbreak illness and 63 deaths due to consumption of contaminated low-water activity foods and spices during the period from 2007 to 2012. *Salmonella* alone was accountable for 94% of the low water activity food recalls in the U.S. and 53% of outbreaks worldwide in above six years (10).

Additionally, exposure of *Salmonella* to low water activity increases cross-protection against other stresses including heat, ethanol, sodium hypochlorite, dodecyl dimethyl ammonium chloride, hydrogen peroxide, NaCl, bile salts, and UV irradiation (11), which ultimately makes the prevention and control strategies less effective. Food industry faces a significant challenge to rein *Salmonella* burden from dry foods and spices without damaging the organoleptic properties. Control of *Salmonella* contamination in the low water activity foods might be improved marginally through improvement in the hygiene and rapid and sensitive detection of *Salmonella* in food and food processing environment. However, more critical is to understand the genetic mechanism of *Salmonella* resistance in low water activity environment for improvement of food safety and public health (12).

In the last few years, a considerable progress has been made to unveil the underlying mechanism of *Salmonella* tolerance against desiccation using transcriptome analysis (13-16). Immediate response of bacteria to low moisture environment involves balancing the internal osmotic pressure to keep it viable. Commonly believed features for desiccation tolerance in *Salmonella* include the followings: increased potassium influx by *kdp* transporter; increased expression of osmoprotectant transport (*proPU* and *osmU*), glutamate and trehalose synthesis; and up-regulation of fatty acid catabolism, Fe-S cluster, sigma factors (*rpoE* and *rpoS*) and *ompC*. Additionally, cellulose and curli fimbrae may play important role in desiccation resistance in *Salmonella*. However, Finn et al. (2015) reported that *S*. Typhimurium response to different humectants, agents that reduce water content of food products, does not simply reflect low water activity but rather are linked to specific humectants (17). In addition, transcripts present at a given time in a cell do not necessarily reflect the functional role of the genes at the given moment. The presence of transcripts can be a reflection of the predictive adaptation of bacteria, thus expression of the transcripts may not have any functional role in the current conditions (18, 19). Furthermore, there can be posttranscriptional regulation of genes and constitutively expressed genes. The limitations of transcriptional analysis can be circumvented by a more direct functional screening approach such as Tn-seq screening employed in the current study.

In this study, for the first time to our knowledge, we used the transposon sequencing (Tn-seq) approach to undermine the genetic mechanisms of desiccation survival in *S*. Typhimurium 14028S. We screened a genome-saturating Tn5 mutant library of *S*. Typhimurium, and identified 61 fitness genes required for survival of *S.* Typhimurium during desiccation stress.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Salmonella enterica serovar Typhimurium 14028S, a spontaneous mutant resistant to nalidixic acid (NA) was used for the transposon insertional mutagenesis. Bacteria were grown in Luria-Bertani (LB) medium or LB agar plates at 37°C and stored at -80°C unless indicated otherwise. NA (ICN Biomedicals Inc., Aurora OH, USA) and kanamycin (Km, Shelton Scientific, Inc. CT, USA) were used at 25 µg/ml and 50 µg/ml, respectively. Bacteria were incubated on shaking rack at 225 rpm when required. Polystyrene disposable petri dish (60 x 15 mm; (VWR International, USA) were used for screening of the mutant library under desiccation stress.

Construction of Tn5 mutant library

Electrocompetent *S.* Typhimurium cells were prepared and transformed with EZ-Tn5 <KAN-2> Tnp transposome complex (Epicentre BioTechnologies, Madison, WI, USA). Electroporation was performed using 0.1-cm cuvettes in a Micropulser electroporator (Bio-Rad Laboratories, Inc., Mississauga, Ontario, Canada) with a field strength of 2450 V. The electroporated cells were immediately resuspended in 500 µl of SOC medium (Quality Biological Inc., Gaithersburg, MD) and incubated for 1.5 h at 37°C on shaking rack (225 rpm). Then, the Tn5 mutant cells were plated on LB plate with double antibiotics (NA and Km) and incubated overnight at 37°C. We collected and combined ~370,000 Tn5 mutants from three transformations, making it a highly complex library. The mutant cells were scrapped off LB plates in 1X phosphate buffered saline (PBS) and stored in 50% glycerol at -80°C.

Screening of Tn5 mutant library during desiccation stress

The Tn5 complex library stored at -80°C was thawed on ice and 300 µl was mixed in 60 ml LB and incubated at 37°C on a shaking rack (OD_{600}=0.135) for 30 min. Then, bacteria were collected by centrifugation at 5,500 rpm for 8 min at room temperature (RT) and resuspended in 50 ml PBS (OD_{600}=0.143). Ten ml from this mutant suspension (t_0 time point) were centrifuged and the bacterial pellet was saved for DNA extraction (Input pool, IP).

For negative selection of Tn5 mutants during desiccation survival, 10 ml of suspension from t_0 was centrifuged and resuspended in 1 ml PBS (~8.0×10^8 CFU/ml). Then aliquots of 100 µl were placed at the center of 10 petri plates (60 x 15 mm size) and air-dried with the lid open inside a bio-safety hood with blower on for 4 h. Then, the plates were covered with the lids and incubated at the RT for 24 h. The desiccated cells were collected from all the 10 petri plates by resuspending them in one ml PBS buffer on each petri plate (in total10 ml PBS) and concentrated to in one ml PBS (8.7×10^6 CFU/ml) with recovery rate of 1.08 % in reference to t_0. Bacterial cells (100 µl aliquot) were plated on 10 LB plates (Km + NA) and incubated overnight. The cells were collected from all the plates in PBS, centrifuged and the pellet was stored at -20°C [output pool (OP) – Desiccation].

DNA library preparation for Illumina sequencing

Genomic DNA (gDNA) were extracted from the input and output pools (100 µl aliquot) using QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) following manufacturer's protocol. The purity and concentration were checked using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA). DNA library were prepared following the protocol developed in our lab (20) with some modifications with details in Supplementary File 1. Briefly, linear extension PCR was done

to enrich the Tn5-juction sequences using a single primer specific to Tn5 transposon (7 bp upstream of invert repeat 2, IR2). The linear extension products were purified and C-tail was attached to it. Then, C-tailed products were purified and exponential PCR was performed using barcoded forward primer and poly G primer with Illumina adapter attached to it (HTM primer; Table S1). The PCR product were purified on 1.5 % agarose gel ranging from 300 – 500 bp. Equal quantity of DNA (10 ng) were mixed and sent for Illumina sequencing using HiSeq 2000 single end read option with 100 cycles (Center for Genome Research and Biocomputing, Oregon State University, Corvallis).

Data analysis

Sequencing reads obtained from the Illumina HiSeq 2000 single end read were analyzed using Analysis of high-Resolution Transposon-Insertion Sequences Technique (ARTIST) (21). Briefly, demultiplexed reads with 20 bp transposon junction sequence were aligned against *S.* Typhimurium 14028S complete genome (NC_016856.1) using Bowtie version 0.12.7 (22). The aligned SAM mapping file were fed to ARTIST pipeline to identify conditionally essential genes (CEGs) using Con-ARTIST (21). Tn5 insertion reads were assigned to 100bp window size of *S.* Typhimurium genome. Uncorrected raw data were used to normalize input data and then reads were compared between input pool and output pool using Mann-Whitney U test (MWU). The MWU results were used to train hidden Markov model (HMM) to predict the likelihood of loci to be conditionally essential or non-essential in the output pool ($p < 0.01$). Only the insertions in the mid 80% (excluding 10% at both 5' and 3' ends) of the protein-coding genes were considered to inactivate the protein functions and thus included in the analysis with cutoff >8 fold and >2 fold for depleted and enriched loci, respectively.

Phenotypic evaluation of single gene knockout mutants

Single-gene knockout mutants of *S*. Typhimurium 14028S were ordered from the BEI Resources (www.beiresources.org) (23). The mutants from 96 well plates were streaked on LB plate (Km) and grown overnight on standard conditions. A single colony was picked and grown in LB broth (Km) for each mutant, and the strains were stored at -80°C in 50% glycerol. Twelve mutants were chosen based on the availability in our strain collection to represent the wide range of fold reduction in read numbers after the selection. Desiccation experiment was performed as described by Gruzdev et al. (2011) with some modifications (11). Single colony of bacteria was picked from LB plate with appropriate antibiotics (wild type with NA and mutant with Km) and incubated in 10 ml LB broth with appropriate antibiotics aerobically overnight in standard conditions (37ºC and shaking rack @225 rpm). Overnight grown bacteria were washed 3 times in 1X PBS at 4ºC with centrifugation at 8,000rpm for 2 min. $O.D_{600}$ was adjusted to 0.4 (±0.05) in 1X PBS. Colony forming units (CFUs) were measured for the wild type and mutants. Fifty µl of bacteria with adjusted OD_{600} (0.4) was then transferred to 96 microtiter well plates. The microtiter plate (with lid open) was placed inside a laminar flow hood with blower on for 10 h for desiccation. The bottom of the microtiter well was completely opaque after drying. After 10 h of drying the microtiter plate was covered with a lid and placed on bench for additional 14 h. Then 200 µl of 1X PBS was added to each well and the plate was shaken for 30 minute at RT. The desiccated bacteria were then released from the wells and resuspended in PBS by vigorously pipetting 15 times and collected in 1.5 ml micro centrifuge tubes. The CFU of the recovered viable cells was measured by plating the suspension on LB plates, followed by overnight incubation for each of the wild type and mutants. Three replications were performed for each

strain. Survival (%) for each strain was calculated as [Total bacteria recovered after desiccation / Total bacteria added to microtiter well plate]*100.

RESULTS AND DISCUSSION

Overview of the selection process and Illumina Sequencing

In this experiment, we subjected a complex Tn5 library of *S*. Typhimurium with more than 350,000 mutants to desiccation stress. The complex Tn5 library (input pool) was air-dried for 4 hours on petri plates inside a laminar flow hood and incubated at the room temperature for 24 h. The number Tn5 mutants before and after the desiccation selection was 8×10^8 CFU/ml and 8.7×10^6 CFU/ml, respectively, indicating only 1.08% of the mutants were able to survive following the desiccation stress. The desiccated Tn5 mutant were resuscitated on LB agar plates (Na + Km) with standard growth conditions. DNA was extracted from the input and output pool and DNA library was prepared for HiSeq Illumina sequencing as described in Material and Methods. Illumina sequencing reads were first demultiplexed based on perfect matching to sample barcodes and then Tn5-junction sequences (20 bp) were extracted allowing some ambiguities in the Tn5 mosaic end of transposon as summarized in Table S1. IP and OP (Desiccation) had 10,842,764 and 5,516,907 reads, respectively and more than 186,000 and 132,000 unique insertions, respectively (Table 1). The number of genomic sites disrupted by Tn5 transposon in the input pool was unexpectedly low (10,842,762/186,000≅58), considering the number mutants collected after electroporation of transposome complex. This might be due to replication of Tn5 mutants during 1.5 h of incubation in SOC medium immediately after electroporation. Moderate Spearman's correlation was observed between input pool and desiccation pool (R^2=0.85, $p<0.0001$) with Tn5 insertion frequency at nucleotide level (Fig. 1A;

Table S2). Additionally, Tn5 transposon were randomly inserted throughout the entire genome without any genomic hot spots and amplification bias (Fig. 1B). This reflects the good quality of Tn5 mutant library and DNA library used for Illumina sequencing.

Identification of desiccation resistance genes

We used Con-ARTIST pipeline to identify the resistance genes required for the desiccation stress tolerance in *S*. Typhimurium. Con-ARTIST identifies transposon mutants at the single-insertion level and normalizes bottleneck effect enabling discovery of conditionally essential mutants at subgenic level (21). We identified 37 entirely conditionally essential and 24 domain essential genes that were required for survival during desiccation stress (Table 2). Among them, ten genes encode putative proteins, and six genes hypothetical proteins.

Further, we assigned desiccation resistance gene to the cluster of orthologous groups (COG) using EggNOG 4.5 (http://eggnogdb.embl.de/#/app/home) with target taxa *Salmonella* (Table 2, Fig 1C). The desiccation resistance genes having no orthologous were assigned to 'No orhtho group'. Equally highly abundant COG were energy production and conversion (C), Cell wall/membrane/envelope biogenesis (M) (14.45%), followed by post-translational modification, protein turnover, and chaperones (O) and inorganic ion transport and metabolism (P) (both 11. 48%). Additionally, desiccation resistance genes belonging to no orthologous group and function unknown were also relatively higher (11.48 and 9.84 % respectively). The moderately abundant COGs were replication, recombination and repair (L), intracellular trafficking, secretion, and vesicular transport (U), translation, ribosomal structure and biogenesis (J) and transcription (K) (4.92 % - 3.28%). Furthermore, COGs with only one genes were amino acid transport and metabolism (E), carbohydrate metabolism and transport (G), nucleotide transport and

metabolism (F), signal transduction mechanisms (T), and defense mechanism (V). PagO belonged to both amino acid and carbohydrate metabolism and transport (Table 2; Fig. 1C).

Additionally, we performed gene enrichment analysis using STRIN database. The KEGG pathways and gene ontology (GO) process enriched for desiccation stress survival were searched in *S. enterica* LT2 (http://bit.ly/2cBK2e6). Genes (STM14_1487, STM14_3165, STM14_4725) that do not have orthologous genes in *S. enterica* LT2 background were not considered. The abundant enriched categories were oxidative phosphorylation (ATP synthase genes), ABC transporters (*fepCDG*, *siiF* and *pstB*), two component system (*glnD*, *rpoN*, and *pagO*), regulation of biological process (*hfq*, *rpoN*, *lepA*, *dsbC*, *dam*, and *glnD*), DNA metabolic process (*dam*, *dnaJK*, and *xerCD*) and O antigen biosynthetic process (*rfbU*, and *rfbA*) plus others.

ATP synthase

All genes encoding the 9 subunits of ATP synthase were shown to be important for dessication survival of *S.* Typhimurium (Table 2). ATP synthase is a highly conserved enzyme across the kingdoms of life with pivotal role in chemiosmotic energy conversion. Bacteria when exposed to desiccation stress, also suffers osmotic stress. Nouri and Komatsu (2010) found that during osmotic stress in soybean plant, H^+-ATPases were the most prominent upregulated proteins, which help the plant maintain membrane potential for energy production, cell turgidity and intracellular pH (24). Also, ATP synthase were one of the dominant proteins expressed over dehydration stress in chickpeas (25). Additionally, in *Plectus murrayi*, a bacteria feeding nematode, ATP synthase subunit transcripts were among the abundantly expressed under desiccated condition (26).

Cell wall/membrane/envelope biogenesis

Genes involved in cell wall/membrane/envelope biogenesis required by *S.* Typhimurium for desiccation survival were *rfbAU*, *wzxE*, *yaeL*, *pal*, *lepA*, *glmS*, STM14_0838, and STM14_0839. RfbAU are essential for O antigen (O polysaccharide) biosynthetic process. Polysaccharides in bacteria may act as water reservoir in dry terrestrial environments. Garmiri *et al*. (2008) found that *Salmonella* spp. lacking O antigen are more sensitive to desiccation (27). Additionally, *wzxE* is involved in translocation of O antigen. STM14_0838 (putative UDP-galactopyranose mutase) generates UDP-alpha-D-galactofuranose required for synthesis of cell wall in bacteria, fungi and protozoa (http://www.genome.jp/dbget-bin/www_bget?ec:5.4.99.9). Furthermore, *Escherichia* coli *yaeL*, a membrane-bound zinc metalloprotease involved in regulated intramembrane proteolysis, is required for activation sigma factor E (σE) encoded by *rpoE* gene in response to an envelope stress (28). Mutation in *pal* gene (peptidoglycan associated lipoprotein) causes a severe defect in the cell envelope of gram-negative bacteria (29). *LepA*, ribosomal elongation factor 4 (EF4), has two opposing function – promoting survival during moderate stress by allowing stress-paused translation to resume and death during severe stress through self-destruction in *E. coli* (30).

Post-translational modification, protein turnover, and chaperones

Salmonella desiccation resistance genes belonging to COG "O" category were *dnaJK*, *dsbC*, *glnD*, STM14_3328, STM14_2014, and STM14_2258. DnaK/DnaJ chaperone machinery are required for protein folding and DNA replication process and essential for protein repair under all stressful conditions including heat shock stress (31, 32). *E.coli* DsbC, a protein required for disulfide bond isomerization in periplasm, assist in folding of several envelope proteins containing disulfides formed between cysteine residues and involved in the defense mechanism against oxidative stress (33). *E. coli* GlnD, a bifunctional uridylyltransferase/uridylyl-removing

signal-transduction enzyme and the primary sensor of nitrogen status in cell, has a critical role in response to growth in either nitrogen limitation or excess. Commonly, nitrogen is an essential chemical for all living being, which is irreplaceable constituent of proteins, DNA and RNA (34, 35). Probably, *S.* Typhimurium faces nitrogen limitation stress during desiccation in PBS making *glnD* a survival fitness gene. STM14_2258 (STM1864), a putative inner membrane protein, is regulated by RcsCDB system, which responds to envelope stress (36).

Inorganic ion transport and metabolism

Salmonella genes involved in inorganic ion transport and metabolism involved in desiccation stress survival were *fepCDG*, *pstB*, *corA*, *nhaA*, and *phoU*. The *fepCDG* and *pstB* encode ATP-binding cassette (ABC) transporters. In *Rhizobium leguminosaurm*, a soil bacterium with ability to fix nitrogen, mutation in an uncharacterized ABC transporter operon (RL2975–RL2977) is highly sensitive to desiccation stress due to significantly lower accumulation of exopolysaccharide (37). FepCDG are iron-enterobactin transporter, a high affinity siderophore that acquires iron for microbial systems (38). Virtually, iron is a vital nutrient for all forms of life and is required for energy generation, DNA replication, oxygen transport and protection against oxidative stress (39). Finn et al. (2013) showed a number of genes involved in Fe-S clusters formation were upregulated during desiccation on a stainless steel surface which are induced under iron-limiting conditions (16). Majority of bacteria regulate the uptake of inorganic orthophosphate (P_i) by a negative regulatory protein PhoU via ABC phosphate-specific transporter (Pst). Phosphorous is an essential element in all cells with roles in ranging from structural and metabolic biological processes to the composition of nucleic acids, phospholipids, and energy intermediates. However, we found only a cytoplasmic ATPase (*pstB*) and *phoU* required for survival during desiccation in *S.* Typhimurium. The other three genes involved in P_i

uptake systems includes extracellular P_i binding proteins (*pstS*) and two transmembrane channel proteins (*pstCA*) (40). This might indicate the possibility of other redundant pathways doing the job of these three genes. Nonetheless, phosphate transport genes (pstACS) were differentially upregulated for survival of desiccated *S.* Typhimurium on s stainless steel surface (16).

CorA encodes magnesium/nickel/cobalt transporter. CorA mutant of *Salmonella* shows a range of phenotype including altered expression of *Salmonella* pathogenicity island 1(SPI-1) genes; decreased tolerance to heat shock and peroxide; defective invasion, survival, and proliferation inside macrophage and epithelial cells; decreased virulence and decreased tolerance to lactoperoxidase enzyme (41). NhaA is pH-dependent sodium/proton (H^+) antiporter that plays a critical role in intracellular pH regulation under alkaline conditions, cell volume regulation, and maintenance of electrochemical potential of Na+ across cytoplasmic membrane plus other (42).

Transcription (K) and Replication, recombination and repair (L)

Salmonella desiccation tolerance gene related to transcription (K) were *rpoN* and *wecD*; and replication, recombination and repair (L) were *dam* and *xerCD*. Alternative sigma factor 54 (*rpoN*, σ54) plays important role in regulation of stress resistance in many bacteria species. *E. coli* RpoN controls more than 14 operons/regulators during nitrogen-limiting conditions and protects the cells from alkaline pH during stationary-phase growth (43, 44). Deletion of *rpoN* in *Listeria monocytogenes* affects ability to grow under osmotic stress (45). Importantly, *Salmonella* Typhi RpoN regulates the expression of O-antigen, a water reservoir, during nitrogen limitation via transcriptional control of *rfaH* gene (46). Additionally, in *Bradyrhizobium japonicum*, a nitrogen-fixing bacteria, deletion of σ54 (*rpoN1*, *rpoN2*, and both) led to significant decrease in viability during desiccation stress (47). WecD, TDP-fucosamine acetyltransferase, is

required in the final step for the synthesis of 4-acetamido-4,6-dideoxy-d-galactose, a sugar unit of polysaccharide (O antigen) which is composed of repeating unit of trisaccharide (48).

Dam, DNA adenine methylase, plays important role in DNA replication, DNA mismatch repair and SOS response (a genome-wide response to DNA damage where cell cycle is arrested and DNA repair and mutagenesis is active) (49). Dam plays a protective role during oxidative stress in *S*. Typhimurium (50). XerCD are site-specific tyrosine recombinase genes that resolves chromosome dimer (and is lethal if not resolved) at a *dif* site (51). Mutation of *xerC* mutant in *Staphylococcus aureus* limited biofilm formation and attenuated virulence in murine bacteremia model (52).

Other desiccation survival genes

Hfq, an RNA chaperone protein, has diverse role in bacterial physiology including growth-dependent metabolism, stress resistance, virulence and drug resistance through post-transcriptional control of gene expression. The most prominent role of Hfq protein in bacteria is in facilitating the interactions between noncoding sRNAs and their cognitive target mRNA molecules (53). Although no sRNA genes implicated in desiccation survival has been reported, the importance of *hfq* gene in desiccation survival of *S.* Typhimurium may suggest the involvement of unknown sRNAs in the process. The 6 intergenic regions identified in this study to be involved in desiccation survival may support this hypothesis (see the next section). In *Francisella novicida*, Hfq protein have important role in stress like osmotic change, low pH, heat shock and oxidative stress. *Salmonella* Hfq protein positively regulates virulence by targeting *hilD* mRNA that affect secretion of type III secretion system (T3SS) encoded by *Salmonella* pathogenicity island 1 (SPI-1) (54, 55). SPI-2 genes (*ssaHIJ*) encoding T3SS were also required

for desiccation survival of *Salmonella*. TolB, a translocation periplasmic protein, is involved in maintaining the integrity of outer membrane via Tol/Pal system in *E. coli* (56).

MiaA, a tRNA delta (2)-isopentenylpyrophosphate transferase gene, is required for the efficient translation of the *rpoS* (σ^S) mRNA. σ^S factor is necessary for the stationary phase/general stress response and required during nutrient starvation and presence of toxic metabolite in *E. coli* (57). However, *rpoS* was not identified as desiccation survival gene in this study. Pnp, a polynucleotide phosphorylase/polyadenylase, provides protection against lactic acid exposure in *S*. Typhimurium. Moreover, *pnp* mutant in *E. coli* have a decrease in RpoS-regulated transcripts (58). Null mutations of *wecE* gene (TDP-4-oxo-6-deoxy-D-glucose transaminase) in *E. coli* responsible for synthesis of enterobacterial common antigen (ECA), a glycolipid found in the outer leaflet of the outer membrane in all species of family *Enterobacteriaceae*, confer sensitivity to bile (59). PagO, an integral inner membrane protein, is activated by phoPQ regulon. *Salmonella* phoPQ is a two-component regulatory system that provides protection against host cationic antimicrobial peptides and intracellular survival within acidic phagosomes by regulating outer membrane (OM) acidic glycerophospholipids with lipid A structure (60). *Salmonella* lacking tpiA gene, a glycolytic enzyme triosephosphate isomerase plays a key role in the central carbon metabolism, have an altered morphology with an elongated shape compared with wild type and is required for full *in vivo* fitness (61).

Additionally, putative genes required for desiccation survival included inner membrane protein (STM14_3329, STM14_2258, and STM14_3328); glycosyl transferase (STM14_0839 and STM14_0845); STM14_1490 (*envF*, putative envelope protein); STM14_1486 (putative cytoplasmic protein); STM14_2014 (putative thiol peroxidase); and STM14_0838 (putative UDP-galactopyranose mutase). Similarly, *S.* Typhimurium desiccation tolerance genes encoding

hypothetical proteins were STM14_0872 (*ybgF*), STM14_1487, STM14_2015, STM14_3164 (*gogB*), STM14_3165, and STM14_4907 (*yiiQ*). ybgF (ybgC-tolQRAB-pal-ybgF operon) is involved in maintenance of cell envelope integrity. GogB, a phage-encoded effector protein, is an anti-inflammatory effector, which regulates inflammation-enhanced colonization and limits tissue damage during *Salmonella* infection (62).

Desiccation resistance *Salmonella* intergenic regions

We identified six entirely essential intergenic regions of *S*. Typhimurium required for survival in desiccation stress on petri plate (Table 3). To determine if any of these intergenic regions encode noncoding sRNA,genomic DNA sequence was extracted for these intergenic region and blasted for the presence of small RNA (sRNA) against sRNATarBase 2.0, a database for bacterial sRNA targets verified by experiment (63). However, we could not find hit for any known sRNA. There might be novel genetic elements in these intergenic regions of *Salmonella* genome yet to be explored.

Furthermore, we searched for the presence of coding region in the desiccation resistance intergenic region using GeneMark (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/genemark.cgi). There was no coding sequence in the 5 intergenic regions. Strikingly, IG_STM14_1490 had a coding sequence with start at 1337639 bp and end at 1337842bp of 204 bp. The result corroborates with PATRIC annotation that contains a hypothetical protein (fig|588858.6.peg.1457) on negative strand (http://mcaf.ee/7pj21l). Moreover, we looked for the promoter regions in the desiccation resistance intergenic regions using PePPER (http://genome2d.molgenrug.nl/index.php/prokaryote-promoters). Interestingly, only

IG_STM14_3165 had the predicted promoter starting at 2782101 bp with 28 bp length on plus strand.

Phenotypic evaluation of single gene knockout mutants

We preformed phonotypic evaluation of 12 single knockout mutants to validate the involvement of the genes in desiccation survival. Six *S*. Typhimurium knockout mutants were entirely essential (Δ*nahA*, Δ*atpG*, Δ*atpH*, Δ*ssaj*, Δ*lepA* and Δ*pagO*) and six were domain essential (Δ*corA*, Δ*pstB*, ΔSTM14_2014, ΔSTM14_5120, ΔSTM14_2258, and ΔSTM14_5122). Tn-seq analysis showed read fold change [$\log_2$(Output pool reads/Input pool reads)] of mutant strains varied from -7.98 to -0.6 with difference of unique insertion count (DUIC; unique insertions of Desiccation – unique insertions of Input pool) ranging from -86 to -10 (Table 2) calculated using Tn-Seq Explorer (64). Unique insertion count is the number of genomic locus disrupted by Tn5 insertion. Mutant survival (%) was calculated as described in MATERIALS AND METHODS. The result showed that only five of the mutants have reduced survival as compared to the wild type and seven strains had higher survival rate than the wile type (Fig. 2A). Among the five mutants with reduced desiccation survival, only three mutants (Δ*atpH*, Δ*atpG* and Δ*corA*) showed statistically significant reduction in survival than wild type ($p < 0.02$, unpaired t-test). To surprise, four mutants had significantly increased survival compared to wild type (ΔSTM14_5122, ΔSTM14_2258, Δ*pagO*, and Δ*pstB*) contrary to the results of Tn-seq analysis (Fig. 2A). To understand the discrepancy between Tn-seq result and phenotypic data, we inspected Tn5 insertion profiles in the input pool and desiccation. The profiles show significantly reduced read numbers after the selection for each identified gene as shown in Fig. 2 (B, C, D, E, and F) and Fig. S1 (A, B, C, D, and E), corroborating well with the genes identified by the

analysis of Tn-seq data. Spearman correlation analysis also indicated that there was significant correlation between survival (%) and log_2FC ($R^2 = 0.62$, $p = 0.0307$) as shown in Fig. S2.

However, the result of phenotypic study did not well substantiate the result of Tn-seq analysis for all mutants tested. We speculate that the disagreement is partially due to the differences in the assay conditions for the library selection and phenotypic assay for single mutants. They differ in terms of the context of experimental vessel (petri plate vs. 96 well plate) and cells (library vs. single mutant), drying method, duration of desiccation stress etc. During the process of optimizing the condition for phenotypic assay, we found that the survival rate of the wild type cells fluctuates greatly depending on the parameters used in the assays. Also, if the phenotype is influenced by the factors secreted into media, the phenotypic outcome of a mutant can be different depending on whether it exist in the context of a library or the pure culture of the same mutant cells. Therefore, we expect that the use of further optimized assay condition may provide the results more consistent with the result of Tn-seq analysis for all mutants tested.

Comparative study

We have searched for *Salmonella* genes in literature, which have been associated with desiccation resistance. Major genes involved in desiccation resistance were K^+ transport channel *kdpFABC* transporter, isocitrate-lyase *aceA*, lipid A biosynthesis palmitoleoyl-acyltransferase *ddg*, iron-sulfur cluster scaffolding protein *nifU*, global regulator *fnr*, alternative sigma factor *rpoE* (13), specialized sigma factor *rpoS* (65), osmoprotectant transporters (*proUP* and *osmU*) (16), and trehalose biosynthesis genes (*ostAB*) (15). All of these genes were disrupted by Tn5 in both input pool and output pool (Table S3), making them non-essential for survival during

desiccation in our experimental setting. Furthermore, these mutants were not sensitive to desiccation stress in our study (Fig. S3). We speculate that the discrepancy may be due to the differences in multiple factors, including the experimental settings (stainless steel surface, sterile filter paper or plasticware), variable desiccation period (couple of hours to weeks), genomic techniques (transcriptome vs Tn-seq), and/or sensitivity of Con-ARTIST pipeline to identify conditionally essential genes. Interestingly, trehalose-negative strains of *Cronobacter* spp., was shown to survive dry stress as well as the wild type strains, suggesting that the factors for desiccation survival could vary in different genetic backgrounds (66).

Additionally, we compared the desiccation resistance genes with related environmental stress like starvation, osmotic, and oxidative stress encountered during the infection cycle by *S.* Typhimurium from our recent study (unpublished data). To note, same input pool (Tn5 complex library of *S.* Typhimurium 14028S) was used for all the stress conditions. Interestingly, we found a more than 50% of desiccation resistance genes were shared to the genes for starvation survival (*Salmonella* starved for 12 days) and more than 30% of desiccation resistance genes were shared to the genes required for resistance to hydrogen peroxide (H_2O_2, 1mM) stress. Hence, this may indicate that *S.* Typhimurium experience starvation as well as oxidative stress during desiccation. Additionally, only ATP synthase genes (9 subunit proteins) were shared between desiccation and osmotic stress (3% NaCl) that were also required for fitness during starvation and hydrogen peroxide insult (Fig. 3). Thus, osmotic stress imposed by 3% NaCl is distinct from the osmotic stress incurred by *Salmonella* during desiccation stress.

CONCLUSION

For the first time to our knowledge, we performed a genome-wide screening of a transposon mutant library to identify desiccation survival gene in *S.* Typhimurium. The precision and accuracy for the identification of conditionally essential genes depends on complexity of input and output library, experimental design and downstream bioinformatics analysis. Tn-seq bioinformatics analysis depends on several factors like library normalization (bottleneck, positional read bias, differences in sequencing depth and stochastic difference in library complexity), annotation dependent analysis and annotation-independent analysis (67). In this study, we used Con-ARIST pipeline that enables the characterization of transposon mutant with annotation-independent approach for discovery of genetic elements at a sub-genic level. We identified 61 protein coding genes and six intergenic regions required for the survival of *S.* Typhimurium during desiccation stress. The important resistance genes to survive the desiccation stress by *S*. Typhimurium were related to energy production and conversion required to maintain basal metabolism; cell wall/membrane/envelope biogenesis required for production of extracellular polysaccharide ; post-translational modification, protein turnover, and chaperones; inorganic ion transport and metabolism for transport of magnesium, nickel, cobalt, sodium, iron and phosphate; replication, recombination and repair to overcome DNA damage; intracellular trafficking, secretion, and vesicular transport ;translation, ribosomal structure and biogenesis and transcription. More than 20% of were either putative or hypothetical genes, thus indicating that this study assigned novel functions to previously unknown genes. Few genes related to amino acid, nucleotide and carbohydrate transport and metabolism were also required to survive desiccation stress encountered by *Salmonella*. Thus, our study was able to provide novel insights into the underlying mechanisms of desiccation survival of *Salmonella*. We expect that our

findings can be further exploited to develop effective control strategies to control the *Salmonella* contamination from low water activity foods and food processing facilities.

ACKNOWLEDGEMENTS

We would like to thank Arkansas Biosciences Institute (ABI) for funding support and Arkansas High Performance Computing Center (AHPCC) - University of Arkansas for their support on bioinformatics analysis.

AUTHOR CONTRIBUTION

YMK- Conceived the study; RKM- performed the experiment, analyzed the data, and drafted the manuscript; YMK- revised the manuscript.

List of Tables

List of Supplementary Tables

List of Figures

Fig 1: Overview of transposon sequencing. A) Spearman correlation (R^2) of Tn5 insertion raw reads frequency distribution between input pool and desiccation at nucleotide level. X- and Y-axis are log transformed. B) Overlay plot displays genome wide Tn5 insertion distribution in input pool and desiccation at nucleotide level (Table S1). C) Cluster of orthologous group (COG) assigned to *S.* Typhimurium desiccation resistance genes using EggNOG 4.5 database. X-axis: Percentage of genes into each COG category. Y-axis: COG categroy. (C- Energy production and conversion; E- Amino acid transport and metabolism; EG- Amino acid transport and metabolism, Carbohydrate transport and metabolism; F -Nucleotide transport and metabolism; G- Carbohydrate transport and metabolism; J- Translation, ribosomal structure and biogenesis; K-Transcription; L - Replication, recombination and repair; M- Cell wall/membrane/envelope biogenesis; O- Post-translational modification, protein turnover, and chaperones; P-Inorganic ion transport and metabolism; S- Function unknown; T- Signal transduction mechanisms; U-Intracellular trafficking, secretion, and vesicular transport; V- Defense mechanisms and No ortho: No orthologous found).

Fig 2: Phenotypic study of null mutants. A) Box plot displays survival (%) of WT (S. Typhimurium, yellow) and mutants (red and green). Red and green color boxplot has mean survival (%) lower and greater than WT respectively. Box represents first and third quartile, line inside box is median and whisker shows minimum and maximum. Strain marked with asterisk (*) have significantly different survival than wild type. (Δ2014 - ΔSTM14_2014, Δ5122 – ΔSTM14_5122, and Δ5120- ΔSTM14_5120). B,C,D,E, and F) Presentation of Tn5 read coverage in input pool (red) and desiccation (green) produced using Integrative Genomics Viewer (IGV)(68). Numbers in square is read coverage.

Fig 3: Venn diagram showing comparison of *S*. Typhimurium desiccation resistance genes with other environmental stress resistance genes. Desiccation resistance genes compared with previously identified resistance genes (unpublished data) during: A- osmotic stress (3% NaCl in LB medium); B - Starvation (starved for 12 days in PBS); C - oxidative stress (1 mm hydrogen peroxide (H_2O_2) in LB medium); and D- all the four stressors.

List of Supplementary Figures

Figure S1: Read coverage in input pool (red) and desiccation in genes of interest. Graph is produced using Integrative Genomics Viewer (IGV). Numbers in square shows read coverage. Dashed line across the gene indicate domain essential gene.

Figure S2: Spearman correlation between survival (%) of null mutants of this study and log_2 fold change (log_2FC). log_2FC is calculated for each gene after normalization of input pool read count based on desiccation read count. More –ve log_2FC of a gene implies more reduced fitness and is lower survival (%) during desiccation. Y-axis: Average survival (%) of 12 mutants and X-axis: log_2FC of read distribution in central 80% of gene based on Tn-seq analysis. Line of fit shows linear regression with confidence intervals. * is survival (%) of wild type (WT) with hypothetical log_2FC of zero (There would be no change in the read count for WT after desiccation).

Figure S3: Tn5 insertion read coverage in the previously identified S. Typhimurium desiccation resistance genes. Graphs were generated using IGV. Tn5 insertion in K+ transport channel kdpABCDE transporter (A), isocitrate-lyase aceA (B), lipid A biosynthesis palmitoleoyl-acyltransferase ddg (C), iron-sulfur cluster scaffolding protein nifU (D) and osmoprotectant transporters proP (E).

Figure S4: Schematic diagram of DNA library preparation of transposon library for Illumina sequencing. Tn-specific primer 1 (Ez-Tn5 primer3, Table S1) is used for linear extension and Tn-specific primer 2 (library specific barcoded primer) is used for exponential PCR in conjunction with C-tail specific primer (HTM-Primer, Table S1).

Figure S5: Agarose gel (1.5%) run of DNA library after exponential PCR (Step 4). DNA from 300 bp – 500 bp is extracted from gel using QIAquick Gel Extraction Kit.

List of Supplementary File

Supplementary File 1: Transposon sequencing (Tn-seq) DNA library preparation protocol. (Figure S4 and Figure S5).

Table 1: Summary of Illumina Sequencing Reads

Library	Total Reads	Reads Mapped (% Aligned)	Unque Insertions	Mean Reads/UnqIns (±SE)	Median Reads
Input pool (IP)	10,842,764	8,867,116 (81%)	186,621	48.99 ±0.99	20
Desiccation	5,516,907	4,248,156 (77%)	132,631	33.18±0.13	18

Table 2: The protein coding genes of *S.* Typhimurium 14028S required for desiccation survival.

Locus_Tag (Gene)	Protein annotation	COG Symbol	Ess [log_2FC] {DUIC}
STM14_4660 (*atpC*)	F0F1 ATP synthase subunit epsilon	C	1 [-5.16] {-10}
STM14_4661 (*atpD*)	F0F1 ATP synthase subunit beta	C	2 [-3.58] {-39}
STM14_4662 (*atpG*)	F0F1 ATP synthase subunit gamma	C	2 [-6.51] {-21}
STM14_4663 (*atpA*)	F0F1 ATP synthase subunit alpha	C	2 [-7.25] {-27}
STM14_4664 (*atpH*)	F0F1 ATP synthase subunit delta	C	2 [-5.97] {-10}
STM14_4665 (*atpF*)	F0F1 ATP synthase subunit B	C	2 [-4.32] {-5}
STM14_4666 (*atpE*)	F0F1 ATP synthase subunit C	C	2 [-5.11] {-13}
STM14_4667 (*atpB*)	F0F1 ATP synthase subunit A	C	2 [-7.01] {-25}
STM14_4668 (*atpI*)	F0F1 ATP synthase subunit I	C	2 [-0.44] {-4}
STM14_4723 (*wecE*)	TDP-4-oxo-6-deoxy-D-glucose transaminase	E	2 [-2.94] {-9}
STM14_2256 (*pagO*)	integral membrane protein	EG	2 [-2.16] {-23}
STM14_3075 (*guaA*)	bifunctional GMP synthase/glutamine amidotransferase protein	F	2 [-4.74] {-17}
STM14_4906 (*tpiA*)	triosephosphate isomerase	G	1 [-5.39] {-10}
STM14_3964 (*pnp*)	polynucleotide phosphorylase/polyadenylase	J	2 [-2.68] {-32}
STM14_5241 (*miaA*)	tRNA delta(2)-isopentenylpyrophosphate transferase	J	1 [-5.64] {-22}
STM14_4008 (*rpoN*)	RNA polymerase factor sigma-54	K	1 [-5.08] {-18}
STM14_4722 (*wecD*)	TDP-fucosamine acetyltransferase	K	1 [-5.43] {-8}
STM14_3676 (*xerD*)	site-specific tyrosine recombinase XerD	L	2 [-3.7] {-11}
STM14_4196 (*dam*)	DNA adenine methylase	L	1 [-4.2] {-20}
STM14_4750 (*xerC*)	site-specific tyrosine recombinase XerC	L	2 [-7.81] {-8}
STM14_0265 (*yaeL*)	zinc metallopeptidase	M	1 [-7.81] {-18}
STM14_0838	putative UDP-galactopyranose mutase	M	2 [-3.78] {-31}
STM14_0839	putative glycosyl transferase	M	1 [-1.93] {-12}
STM14_0871 (*pal*)	peptidoglycan-associated outer membrane lipoprotein	M	2 [-3.8] {-4}
STM14_2580 (*rfbU*)	mannosyl transferase	M	2 [-6.95] {-40}
STM14_2589 (*rfbA*)	dTDP-glucose pyrophosphorylase	M	1 [-6.81] {-49}
STM14_3163 (*lepA*)	GTP-binding protein LepA	M	2 [-2.33] {-17}
STM14_4656 (*glmS*)	D-fructose-6-phosphate amidotransferase	M	2 [-5.37] {-26}
STM14_4724 (*wzxE*)	O-antigen translocase	M	2 [-6.17] {-38}
STM14_0013 (*dnaK*)	molecular chaperone DnaK	O	2 [-5.39] {-12}
STM14_0014 (*dnaJ*)	chaperone protein DnaJ	O	1 [-1.26] {-17}
STM14_0254 (*glnD*)	PII uridylyl-transferase	O	2 [-3.96] {-30}
STM14_2014	putative thiol peroxidase	O	1 [-3.03] {-16}

Table 2. (Cont.) The protein coding genes of *S*. Typhimurium 14028S for desiccation survival.

Locus_Tag (Gene)	Protein annotation	COG Symbol	Ess [log$_2$FC] {DUIC}
STM14_2258	putative inner membrane protein	O	1 [-1.15] {-23}
STM14_3328	putative inner membrane protein	O	2 [-2.11] {-25}
STM14_3675 (*dsbC*)	thiol:disulfide interchange protein DsbC	O	1 [-0.13] {-4}
STM14_0048 (*nhaA*)	pH-dependent sodium/proton antiporter	P	2 [-7.98] {-15}
STM14_0688 (*fepC*)	iron-enterobactin transporter ATP-binding protein	P	1 [-4.21] {-10}
STM14_0689 (*fepG*)	iron-enterobactin transporter permease	P	2 [-4.67] {-5}
STM14_0690 (*fepD*)	iron-enterobactin transporter membrane protein	P	2 [-4.46] {-4}
STM14_4648 (*phoU*)	transcriptional regulator PhoU	P	2 [-3] {-8}
STM14_4649 (*pstB*)	phosphate transporter subunit	P	1 [-2.9] {-17}
STM14_4754 (*corA*)	magnesium/nickel/cobalt transporter CorA	P	1 [-4.94] {-25}
STM14_0845	putative glycosyl transferase	S	1 [-0.41] {-18}
STM14_0872 (*ybgF*)	hypothetical protein STM14_0872	S	1 [0.64] {-4}
STM14_1486	putative cytoplasmic protein	S	2 [-3.08] {-20}
STM14_3164 (*gogB*)	hypothetical protein STM14_3164	S	2 [-0.63] {-34}
STM14_3329	putative inner membrane protein	S	1 [-2.64] {-8}
STM14_4907 (*yiiQ*)	hypothetical protein STM14_4907	S	1 [0.77] {-3}
STM14_5242 (*hfq*)	RNA-binding protein Hfq	T	1 [-0.57] {-6}
STM14_0870 (*tolB*)	translocation protein TolB	U	1 [-3.35] {-15}
STM14_1703 (*ssaH*)	type III secretion system apparatus protein	U	2 [-0.58] {-1}
STM14_1705 (*ssaJ*)	needle complex inner membrane lipoprotein	U	2 [-3.47] {-10}
STM14_5122	putative ABC-type bacteriocin/lantibiotic exporter	V	1 [-0.6] {-86}
STM14_1487	hypothetical protein STM14_1487	No ortho	2 [-1.59] {-3}
STM14_1490 (*envF*)	putative envelope lipoprotein	No ortho	2 [-0.58] {-7}
STM14_1704 (*ssaI*)	type III secretion system apparatus protein	No ortho	2 [-5.47] {-10}
STM14_2015	hypothetical protein STM14_2015	No ortho	2 [-0.55] {-8}
STM14_3165	hypothetical protein STM14_3165	No ortho	2 [NA] {NA}
STM14_4725	4-alpha-L-fucosyltransferase	No ortho	2 [-2.36] {-23}
STM14_5120	cation efflux pump	No ortho	1 [-1.56] {-23}

Ess: Essentiality based on Con-ARTIST; 1- Domain essential; 2- Entirely Essential; COG-Cluster of Orthologous Groups; log$_2$FC: log$_2$ fold change after read normalization in central 80% of gene; DUIC: difference of unique insertion count between input pool and desiccation in core 80% of gene; the more –ve more reduced is the fitness. COG annotation are similar as Fig. 1.

Table 3: The intergenic regions of *S.* Typhimurium 14028S required for desiccation survival.

Intergenic region	Start	End	Length	Essentiality
IG_STM14_3329	2923580	2923838	259	2
IG_STM14_3165	2782023	2782225	203	2
IG_STM14_3164	2780125	2780528	404	2
IG_STM14_2257	1971827	1971958	132	2
IG_STM14_1490	1337373	1338163	791	2
IG_STM14_0255	253521	253756	236	2

2- Entirely essential as classified by Con-ARTIST pipeline

Fig 1: Overview of transposon sequencing. A) Spearman correlation (R^2) of Tn5 insertion raw reads frequency distribution between input pool and desiccation at nucleotide level. X- and Y-axis are log transformed. B) Overlay plot displays genome wide Tn5 insertion distribution in input pool and desiccation at nucleotide level (Table S1). C) Cluster of orthologous group (COG) assigned to *S.* Typhimurium desiccation resistance genes using EggNOG 4.5 database. X-axis: Percentage of genes into each COG category. Y-axis: COG category. (C- Energy production and conversion; E - Amino acid transport and metabolism; EG- Amino acid transport and metabolism, Carbohydrate transport and metabolism; F -Nucleotide transport and metabolism; G- Carbohydrate transport and metabolism; J- Translation, ribosomal structure and biogenesis; K-Transcription; L - Replication, recombination and repair; M- Cell wall/membrane/envelope biogenesis; O- Post-translational modification, protein turnover, and chaperones; P-Inorganic ion transport and metabolism; S- Function unknown; T- Signal transduction mechanisms; U- Intracellular trafficking, secretion, and vesicular transport; V- Defense mechanisms and No ortho: No orthologous found).

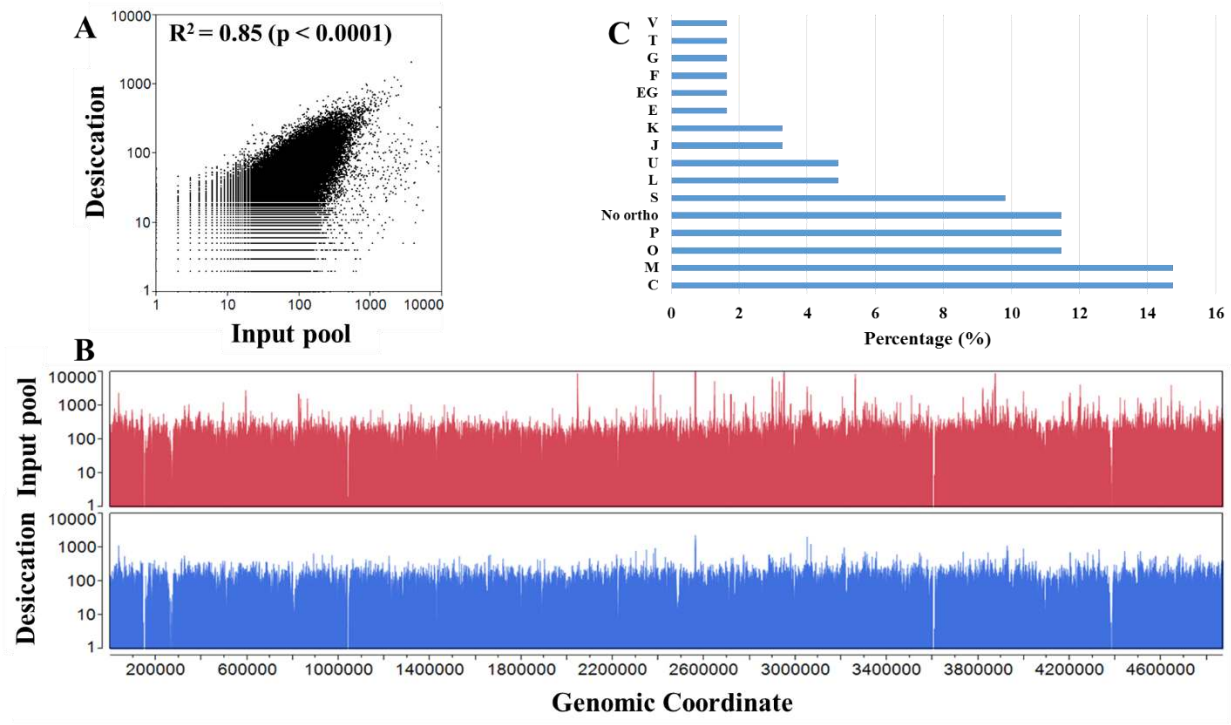


Fig 2: Phenotypic study of null mutants. A) Box plot displays survival (%) of WT (S. Typhimurium, yellow) and mutants (red and green). Red and green color boxplot has mean survival (%) lower and greater than WT respectively. Box represents first and third quartile, line inside box is median and whisker shows minimum and maximum. Strain marked with asterisk (*) have significantly different survival than wild type. (Δ2014 - ΔSTM14_2014, Δ5122 – ΔSTM14_5122, and Δ5120- ΔSTM14_5120). B,C,D,E, and F) Presentation of Tn5 read coverage in input pool (red) and desiccation (green) produced using Integrative Genomics Viewer (IGV)(68). Numbers in square is read coverage.

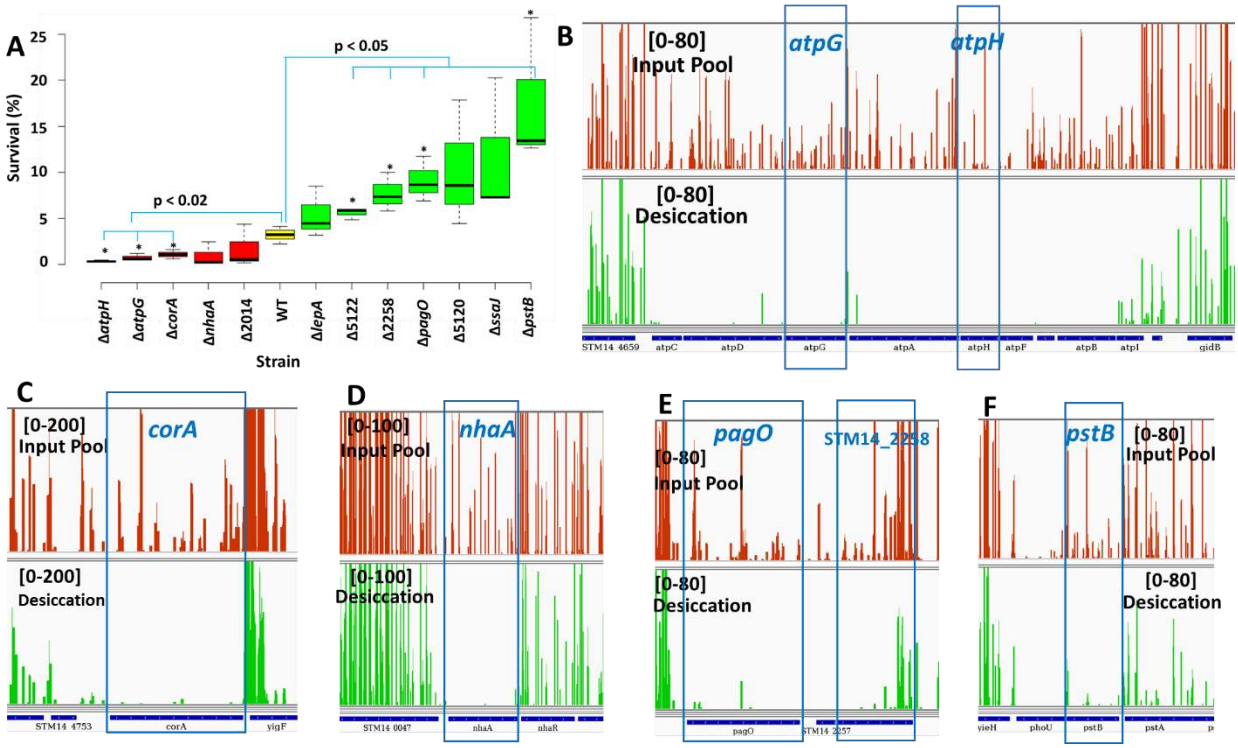


Fig 3: Venn diagram showing comparison of *S*. Typhimurium desiccation resistance genes with other environmental stress resistance genes. Desiccation resistance genes compared with previously identified resistance genes (unpublished data) during: A- osmotic stress (3% NaCl in LB medium); B - Starvation (starved for 12 days in PBS); C - oxidative stress (1 mm hydrogen peroxide (H_2O_2) in LB medium); and D- all the four stressors.

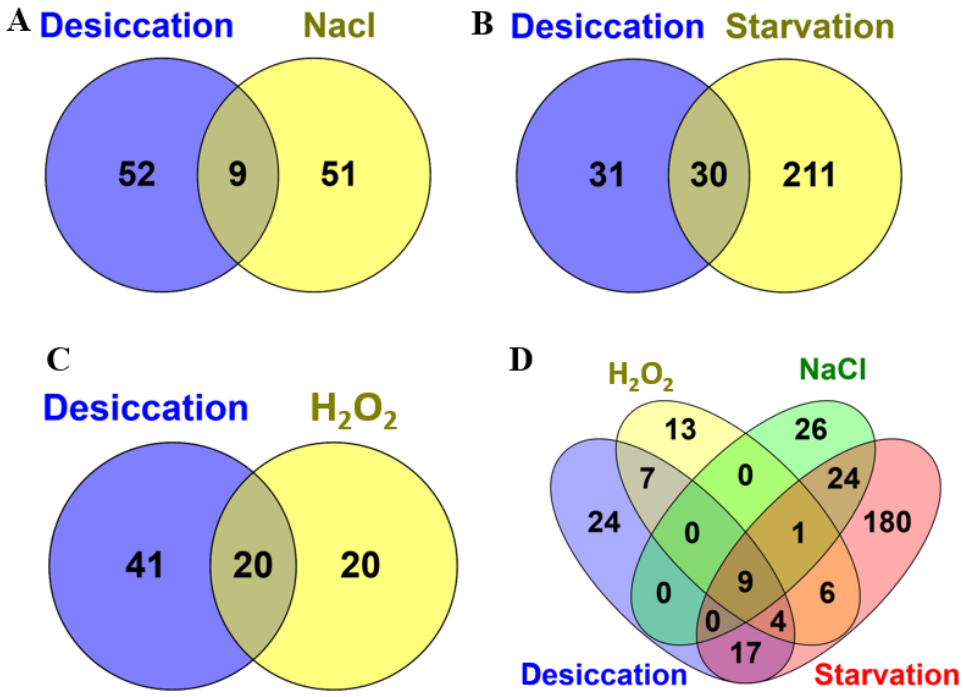


Figure S1: Read coverage in input pool (red) and desiccation in genes of interest. Graph is produced using Integrative Genomics Viewer (IGV). Numbers in square shows read coverage. Dashed line across the gene indicate domain essential gene.

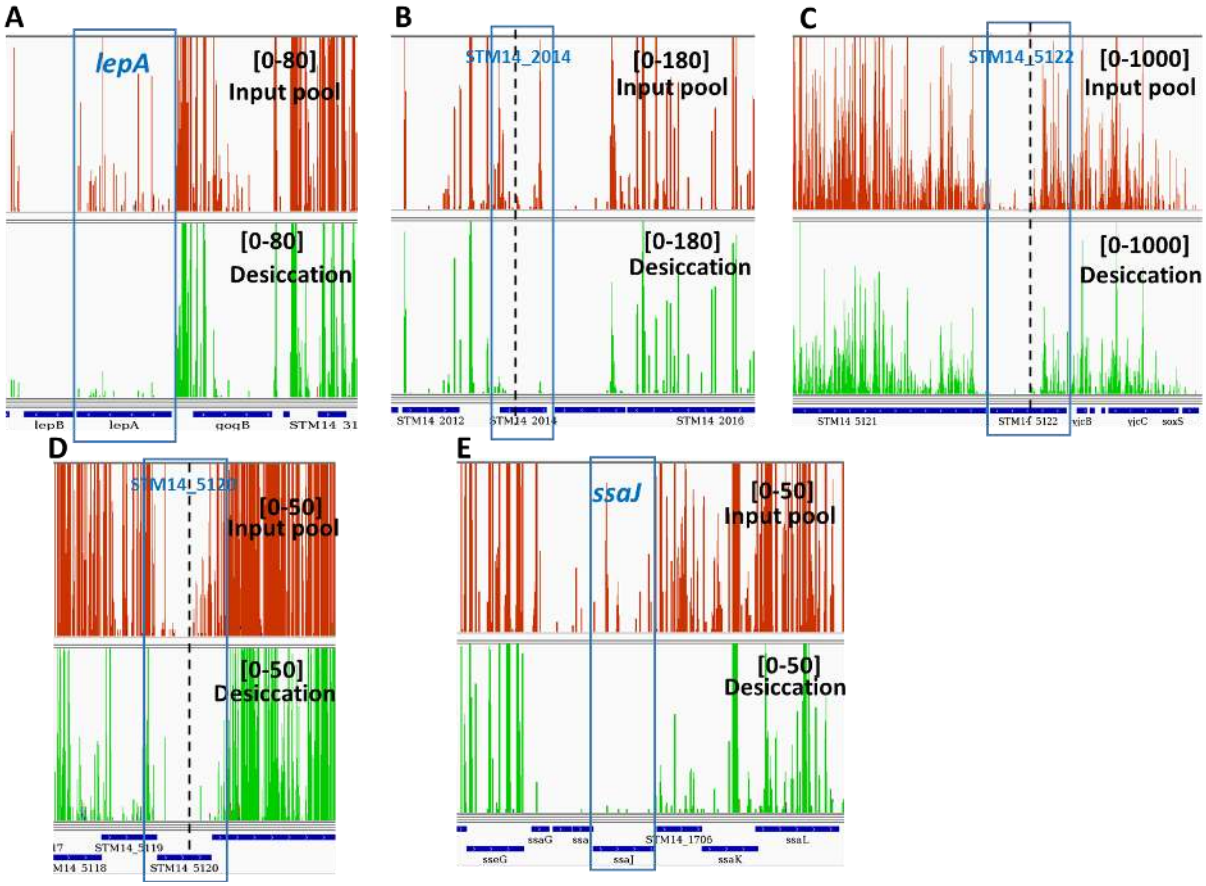


Figure S2: Spearman correlation between survival (%) of null mutants of this study and log_2 fold change (log_2FC). log_2FC is calculated for each gene after normalization of input pool read count based on desiccation read count. More –ve log_2FC of a gene implies more reduced fitness and is lower survival (%) during desiccation. Y-axis: Average survival (%) of 12 mutants and X-axis: log_2FC of read distribution in central 80% of gene based on Tn-seq analysis. Line of fit shows linear regression with confidence intervals. * is survival (%) of wild type (WT) with hypothetical log_2FC of zero (There would be no change in the read count for WT after desiccation).

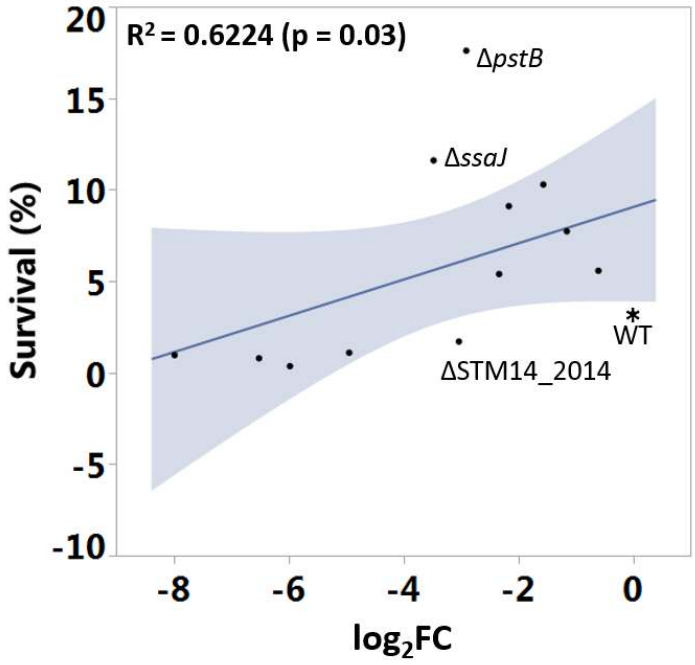


Figure S3: Tn5 insertion read coverage in the previously identified *S*. Typhimurium desiccation resistance genes. Graphs were generated using IGV. Tn5 insertion in K+ transport channel *kdpABCDE* transporter (A), isocitrate-lyase *aceA* (B), lipid A biosynthesis palmitoleoyl-acyltransferase *ddg* (C), iron-sulfur cluster scaffolding protein *nifU* (D) and osmoprotectant transporters *proP* (E).

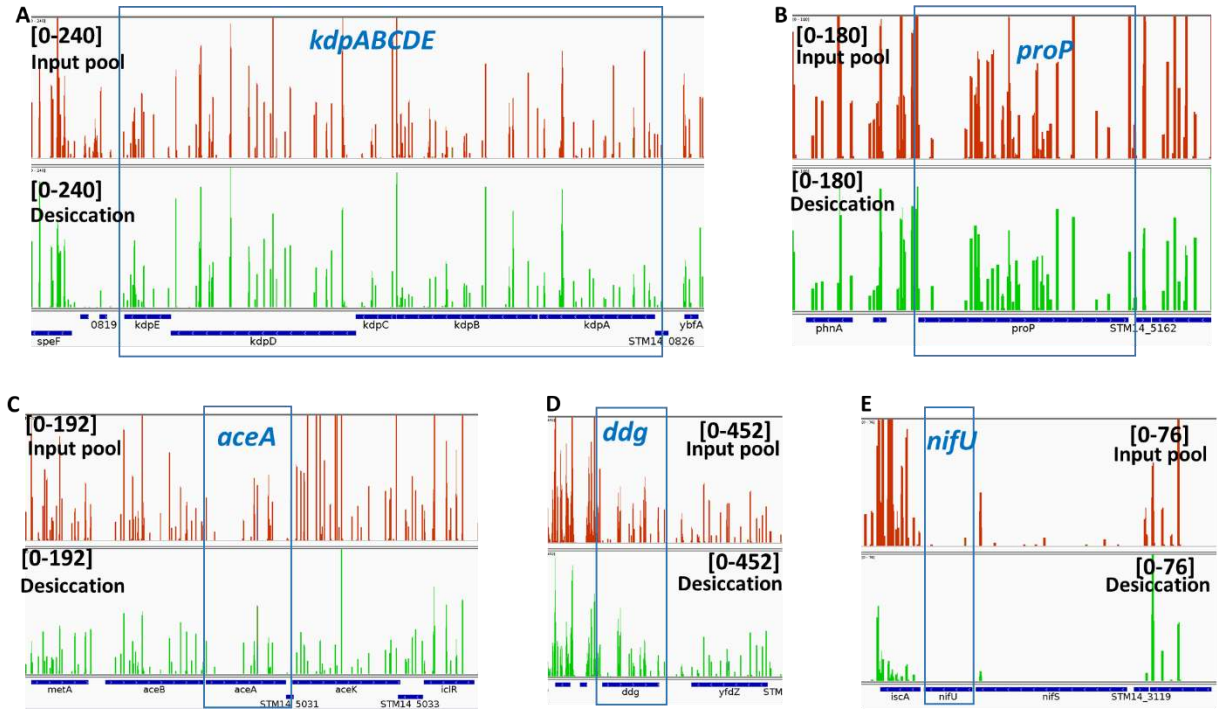


Table S1: Oligonucleotide used for Tn-seq PCR

Tn-seq PCR	Oligonucleotide	Sequence	Barcode
Linear Extension	Ez-Tn5 primer3	5’-GATCCTCTAGAGTCGACCTGCAGGCATGCA-3’	
Exponential PCR	IR2-IS-B7	5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGXXXXXXTCAGGGTTGAGATGTGTATAAGAGACAG-3’	CAGATC
	IR2-IS-B14	5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNAGXXXXXXTCAGGGTTGAGATGTGTATAAGAGACAG-3’	CGTTGT
	HTM-Primer	5’-CAAGCAGAAGACGGCATACGAGCTCTTCCGATCTGGGGGGGGGGGGGGG-3’	

NNNN: random sequence for efficient cluster analysis. XXXXXX: 6nt sample barcode sequences. For exponential PCR, IR2-IS-B7 and HTM-primer were used for input pool and IR2-IS-B14 and HTM-primer was used for desiccation. Few wild characters (.C.G.G.T.A.A.G.G.A.A.GAGACAG in Perl script, where “.” is any character) were used in mosaic end reads for identification of the reads containing Tn5-junction sequences.

Supplementary File 1: DNA library preparation protocol for transposon sequencing (Tn-seq) using regular primer

This protocol for DNA library preparation of transposon mutant library is the improved version of previously developed methodology in our laboratory (1). This approach is based on addition of poly deoxycytosine (C) tails to 3' end of either single or double-stranded DNA (2). Addition of C tail is controlled effectively by the mixture of deoxycytidine triphosphate (dCTP) and dideoxy CTP (ddCTP). A single-primer is used for the linear extension of transposon junction sequences with a transposon-specific primer. The linear extension product is purified and subsequently C-tail is attached to it. Next, the transposon junction sequence is easily amplified by transposon specific primer and poly G primer (3). This improved version uses only one polymearase and minimizes the use of oligonucleotide and PCR amplification. The simplified diagram for improved version is shown in Figure S4. DNA from 300 bp – 500 bp is extracted from 1.5% agarose gel as shown in Figure S5.

MATERIALS

DNA extracted from Tn5 mutant library

Wild type DNA (Control DNA)

QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA)

Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA)

Oligonucleotide (Table S1)

Taq DNA polymerase (New England Biolabs, Ipswich, MA, USA)

Thermopol buffer

2.5 mM dNTP mix

Nuclease-free water

MinElute PCR purification kit (Qiagen, Valencia, CA, USA)

Terminal Transferase (TdT, New England Biolabs, Ipswich, MA, USA)

TdT Reaction Buffer (10X)

CoCl2 (2.5 mM) (TdT, New England Biolabs, Ipswich, MA, USA)

dCTP (100 mM) (Promega, Madison, WI, USA)

ddCTP (10 mM) (Promega Madison, WI, USA)

1.5% Agarose gel

QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA)

Thermocycler (PCR machine)

STEPS

Step1: Linear extension PCR

1. Reaction mixture

ddH_2O	40 µl
Thermopol Buffer (10X)	5 µl
dNTPs (2.5 mM each)	1 µl
EZ-Tn5 primer 3 (20 µM) (T_m=65.6ºC)	1 µl
Genomic DNA of Tn library	2 µl
Taq DNA polymerase (NEB)	1 µl
Total	50 µl

2. PCR cycle

	95°C	2 min (manual hot-start)
50 cycles	[95°C	30 sec]
	[62°C	45 sec]
	[72°C	10 sec]
	4°C	hold

3. Purify the linear extension PCR products using Qiagen MinElute PCR-purification kit. Elute DNA in 10 µl EB buffer and store at -20°C.

Step 2: C-tailing reaction

1. Preparation of dNTP working stock

Dilute 100 mM dCTP to 10 mM dCTP with ddH_2O (nuclease-free)
Dilute 10 mM ddCTPto 1 mM ddCTP with ddH_2O (nuclease-free)

2. Reaction mixture

DNA (linear extension products)	10.0 µl
TdT Buffer (10X)	2.0 µl
2.5 mM $CoCl_2$	2.0 µl
10 mM dCTP	2.4 µl
1 mM ddCTP	1.0 µl
ddH_2O	1.6 µl
Terminal transferase	1.0 µl
Total	20.0 µl

3. Incubate the reaction tube at 37°C for 1 hr.

4. Incubate the reaction tube at 75°C for 20 min for heat inactivation of TdT.

5. Purify the C-tailed products using Qiagen MinElute PCR-purification kit. Elute DNA in 10 µl EB buffer and store at -20°C.

Step 3: PCR to amplify Tn-flanking sequences

1. Reaction mixture

ddH_2O	35 µl
Thermopol Buffer (10X)	5 µl
dNTPs (2.5 mM each)	4 µl
IR2-IS-BC primer (10 µM)	2 µl
HTM primer (20 µM)	1 µl
C-tailed DNA	2 µl
Taq DNA polymerase (NEB)	1 µl
Total	50 µl

Note- IR2-IS-BC primer is barcoded primer used during exponential PCR.

4. PCR cycle

	95°C	2 min (manual hot-start)
36 cycles	[95°C	30 sec]
	[58°C	45 sec]
	[72°C	20 sec]
	72°C	10 min
	4°C	hold

Step 4: Gel-purification of PCR products

1. Mix the sample with loading buffer and heat at 65ºC for 15 min.
2. Run 10 µl/sample on 1.5% agarose gel.
3. Cut 300-500bp bands and gel-purify DNA fragments.

References (Supplementary File 1):

1. **Dawoud TM, Jiang T, Mandal RK, Ricke SC, Kwon YM.** 2014. Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers. Mol Biotechnol **56:**1004-1010.

2. **Lazinski DW, Camilli A.** 2013. Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. BioTechniques **54:**25.

3. **Kwon YM, Ricke SC, Mandal RK.** 2016. Transposon sequencing: methods and expanding applications. Appl Microbiol Biotechnol **100:**31-43.

Figure S4: Schematic diagram of DNA library preparation of transposon library for Illumina sequencing. Tn-specific primer 1 (Ez-Tn5 primer3, Table S1) is used for linear extension and Tn-specific primer 2 (library specific barcoded primer according) is used for exponential PCR in conjunction with C-tail specific primer (HTM-Primer, Table S1).

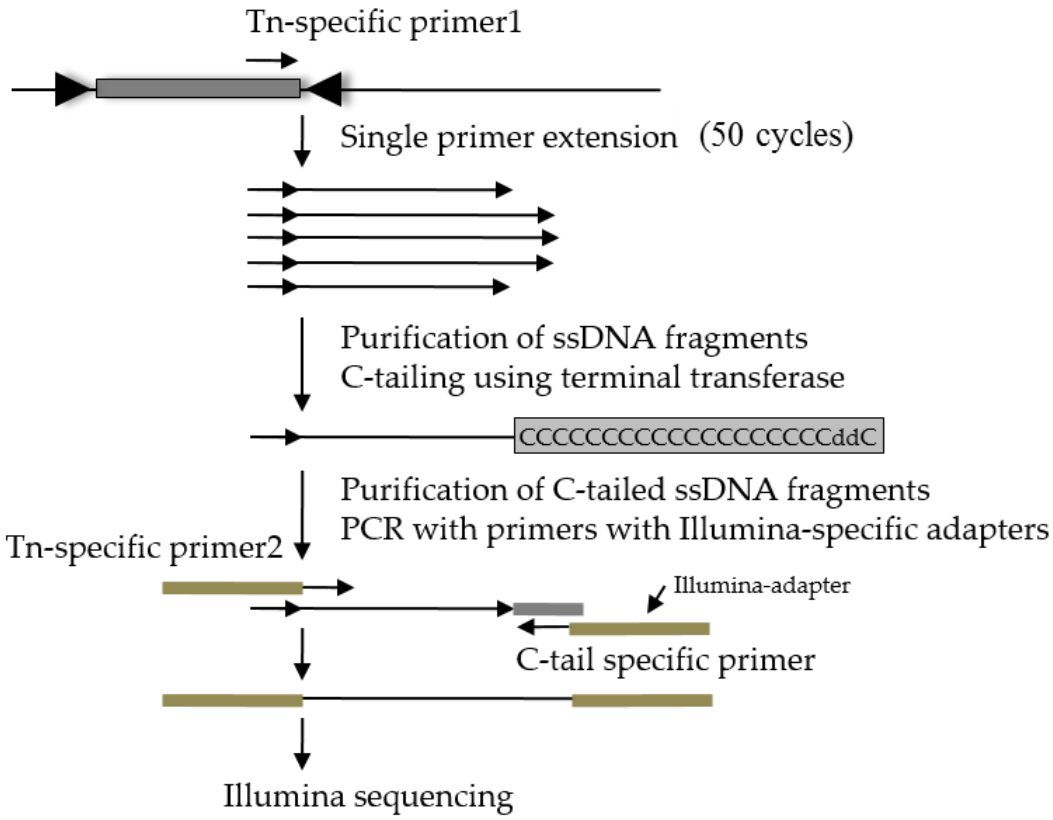
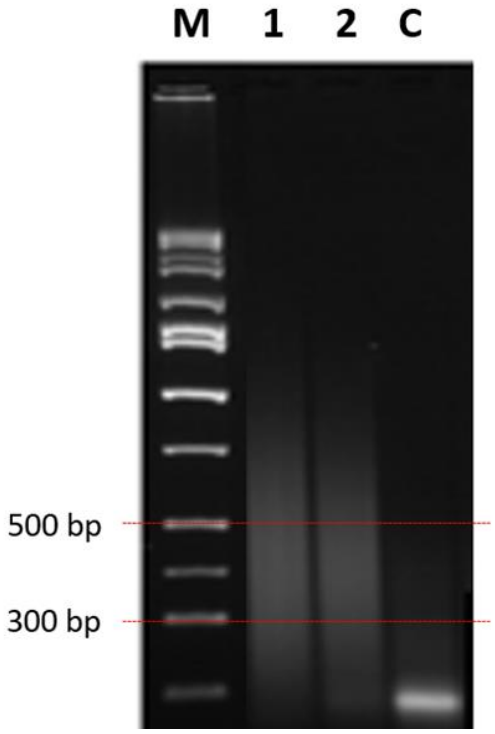


Figure S5: Agarose gel (1.5%) run of DNA library after exponential PCR (Step 4). DNA from 300 bp – 500 bp is extracted from gel using QIAquick Gel Extraction Kit. M: Hi-Lo Marker, 1: Input pool, 2: Desiccation, C: Control (Wild Type DNA).



REFERENCES

1. **Spector, MP, Kenyon, WJ.** 2012. Resistance and survival strategies of Salmonella enterica to environmental stresses. Food Res. Int. **45:**455-481.

2. **Bell, C, Kyriakides, A.** 2008. Salmonella: a practical approach to the organism and its control in foods. John Wiley & Sons.

3. **Hiramatsu, R, Matsumoto, M, Sakae, K, Miyazaki, Y.** 2005. Ability of Shiga toxin-producing Escherichia coli and Salmonella spp. to survive in a desiccation model system and in dry foods. Appl. Environ. Microbiol. **71:**6657-6663. doi: 71/11/6657 [pii].

4. **Kotzekidou, P.** 1998. Microbial stability and fate of Salmonella Enteritidis in halva, a low-moisture confection. Journal of Food Protection®. **61:**181-185.

5. **Uesugi, AR, Danyluk, MD, Harris, LJ.** 2006. Survival of Salmonella Enteritidis phage type 30 on inoculated almonds stored at− 20, 4, 23, and 35 C. Journal of Food Protection®. **69:**1851-1857.

6. **Beuchat, LR, Heaton, EK.** 1975. Salmonella survival on pecans as influenced by processing and storage conditions. Appl. Microbiol. **29:**795-801.

7. **Komitopoulou, E, Peñaloza, W.** 2009. Fate of Salmonella in dry confectionery raw materials. J. Appl. Microbiol. **106:**1892-1900.

8. **Gruzdev, N, Pinto, R, Sela, S.** 2012. Persistence of Salmonella enterica during dehydration and subsequent cold storage. Food Microbiol. **32:**415-422.

9. **Braun, P, Fehlhaber, K, Wicke, A.** 1999. Salmonella enteritidis invades the egg through the shell. World Poult Special. 23-24.

10. **Farakos, SMS, Frank, JF.** 2014. Challenges in the Control of Foodborne Pathogens in Low-Water Activity Foods and Spices, p. 15-34. *In* Anonymous The Microbiological Safety of Low Water Activity Foods and Spices. Springer.

11. **Gruzdev, N, Pinto, R, Sela, S.** 2011. Effect of desiccation on tolerance of salmonella enterica to multiple stresses. Appl. Environ. Microbiol. **77:**1667-1673. doi: 10.1128/AEM.02156-10 [doi].

12. **Beuchat, LR, Komitopoulou, E, Beckers, H, Betts, RP, Bourdichon, F, Fanning, S, Joosten, HM, Ter Kuile, BH.** 2013. Low–water activity foods: increased concern as vehicles of foodborne pathogens. Journal of Food Protection®. **76:**150-172.

13. **Gruzdev, N, McClelland, M, Porwollik, S, Ofaim, S, Pinto, R, Saldinger-Sela, S.** 2012. Global transcriptional analysis of dehydrated Salmonella enterica serovar Typhimurium. Appl. Environ. Microbiol. **78:**7866-7875. doi: 10.1128/AEM.01822-12 [doi].

14. **Deng, X, Li, Z, Zhang, W.** 2012. Transcriptome sequencing of Salmonella enterica serovar Enteritidis under desiccation and starvation stress in peanut oil. Food Microbiol. **30:**311-315.

15. **Li, H, Bhaskara, A, Megalis, C, Tortorello, ML.** 2012. Transcriptomic analysis of Salmonella desiccation resistance. Foodborne Pathogens and Disease. **9:**1143-1151.

16. **Finn, S, Handler, K, Condell, O, Colgan, A, Cooney, S, McClure, P, Amezquita, A, Hinton, JC, Fanning, S.** 2013. ProP is required for the survival of desiccated Salmonella enterica serovar typhimurium cells on a stainless steel surface. Appl. Environ. Microbiol. **79:**4376-4384. doi: 10.1128/AEM.00515-13 [doi].

17. **Finn, S, Rogers, L, Handler, K, McClure, P, Amezquita, A, Hinton, JC, Fanning, S.** 2015. Exposure of Salmonella enterica Serovar Typhimurium to Three Humectants Used in the Food Industry Induces Different Osmoadaptation Systems. Appl. Environ. Microbiol. **81:**6800-6811. doi: 10.1128/AEM.01379-15 [doi].

18. **Mitchell, A, Romano, GH, Groisman, B, Yona, A, Dekel, E, Kupiec, M, Dahan, O, Pilpel, Y.** 2009. Adaptive prediction of environmental changes by microorganisms. Nature. **460:**220-224.

19. **Tagkopoulos, I, Liu, YC, Tavazoie, S.** 2008. Predictive behavior within microbial genetic networks. Science. **320:**1313-1317. doi: 10.1126/science.1154456 [doi].

20. **Dawoud, TM, Jiang, T, Mandal, RK, Ricke, SC, Kwon, YM.** 2014. Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers. Mol. Biotechnol. **56:**1004-1010.

21. **Pritchard, JR, Chao, MC, Abel, S, Davis, BM, Baranowski, C, Zhang, YJ, Rubin, EJ, Waldor, MK.** 2014. ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. PLoS Genet. **10:**e1004782.

22. **Langmead, B, Trapnell, C, Pop, M, Salzberg, SL.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. **10:**1.

23. **Porwollik, S, Santiviago, CA, Cheng, P, Long, F, Desai, P, Fredlund, J, Srikumar, S, Silva, CA, Chu, W, Chen, X.** 2014. Defined single-gene and multi-gene deletion mutant collections in Salmonella enterica sv Typhimurium. PloS One. **9:**e99820.

24. **Nouri, M, Komatsu, S.** 2010. Comparative analysis of soybean plasma membrane proteins under osmotic stress using gel-based and LC MS/MS-based proteomics approaches. Proteomics. **10:**1930-1945.

25. **Jaiswal, DK, Mishra, P, Subba, P.** 2014. Membrane-associated proteomics of chickpea identifies Sad1/UNC-84 protein (CaSUN1), a novel component of dehydration signaling. Nature (Scientific Reports). **4:**1-10.

26. **Adhikari, BN, Wall, DH, Adams, BJ.** 2009. Desiccation survival in an Antarctic nematode: molecular analysis using expressed sequenced tags. BMC Genomics. **10:**1.

27. **Garmiri, P, Coles, KE, Humphrey, TJ, Cogan, TA.** 2008. Role of outer membrane lipopolysaccharides in the protection of Salmonella enterica serovar Typhimurium from desiccation damage. FEMS Microbiol. Lett. **281:**155-159. doi: 10.1111/j.1574-6968.2008.01093.x [doi].

28. **KANEHARA, K, ITO, K, AKIYAMA, Y.** 2002. YaeL (EcfE) activates the σE pathway of stress response through a site-2 cleavage of anti-σe, RseA. Genes Dev. **16:**2147-2155.

29. **Vines, ED, Marolda, CL, Balachandran, A, Valvano, MA.** 2005. Defective O-antigen polymerization in tolA and pal mutants of Escherichia coli in response to extracytoplasmic stress. J. Bacteriol. **187:**3359-3368. doi: 187/10/3359 [pii].

30. **Li, L, Hong, Y, Luan, G, Mosel, M, Malik, M, Drlica, K, Zhao, X.** 2014. Ribosomal elongation factor 4 promotes cell death associated with lethal stress. MBio. **5:**e01708-14. doi: 10.1128/mBio.01708-14 [doi].

31. **Takaya, A, Tomoyasu, T, Matsui, H, Yamamoto, T.** 2004. The DnaK/DnaJ chaperone machinery of Salmonella enterica serovar Typhimurium is essential for invasion of epithelial cells and survival within macrophages, leading to systemic infection. Infect. Immun. **72:**1364-1373.

32. **Rychlik, I, Barrow, PA.** 2005. Salmonella stress management and its relevance to behaviour during intestinal colonisation and infection. FEMS Microbiol. Rev. **29:**1021-1040. doi: S0168-6445(05)00039-2 [pii].

33. **Denoncin, K, Vertommen, D, Arts, IS, Goemans, CV, Rahuel-Clermont, S, Messens, J, Collet, JF.** 2014. A new role for Escherichia coli DsbC protein in protection against oxidative stress. J. Biol. Chem. **289:**12356-12364. doi: 10.1074/jbc.M114.554055 [doi].

34. **Tondervik, A, Torgersen, HR, Botnmark, HK, Strom, AR.** 2006. Transposon mutations in the 5' end of glnD, the gene for a nitrogen regulatory sensor, that suppress the osmosensitive phenotype caused by otsBA lesions in Escherichia coli. J. Bacteriol. **188:**4218-4226. doi: 188/12/4218 [pii].

35. **Yurgel, SN, Rice, J, Kahn, ML.** 2013. Transcriptome analysis of the role of GlnD/GlnBK in nitrogen stress adaptation by Sinorhizobium meliloti Rm1021. PloS One. **8:**e58028.

36. **Mariscotti, JF, Garcia-del Portillo, F.** 2009. Genome expression analyses revealing the modulation of the Salmonella Rcs regulon by the attenuator IgaA. J. Bacteriol. **191:**1855-1867. doi: 10.1128/JB.01604-08 [doi].

37. **Vanderlinde, EM, Harrison, JJ, Muszynski, A, Carlson, RW, Turner, RJ, Yost, CK.** 2010. Identification of a novel ABC transporter required for desiccation tolerance, and biofilm

formation in Rhizobium leguminosarum bv. viciae 3841. FEMS Microbiol. Ecol. **71:**327-340. doi: 10.1111/j.1574-6941.2009.00824.x [doi].

38. **Porcheron, G, Garénaux, A, Proulx, J, Sabri, M, Dozois, CM.** 2015. Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. Metal Economy in Host-Microbe Interactions. .

39. **Skaar, EP.** 2010. The battle for iron between bacterial pathogens and their vertebrate hosts. PLoS Pathog. **6:**e1000949.

40. **Zheng, JJ, Sinha, D, Wayne, KJ, Winkler, ME.** 2016. Physiological Roles of the Dual Phosphate Transporter Systems in Low and High Phosphate Conditions and in Capsule Maintenance of Streptococcus pneumoniae D39. Frontiers in Cellular and Infection Microbiology. **6:**63.

41. **Sermon, J, Wevers, EM, Jansen, L, De Spiegeleer, P, Vanoirbeek, K, Aertsen, A, Michiels, CW.** 2005. CorA affects tolerance of Escherichia coli and Salmonella enterica serovar Typhimurium to the lactoperoxidase enzyme system but not to other forms of oxidative stress. Appl. Environ. Microbiol. **71:**6515-6523. doi: 71/11/6515 [pii].

42. **Vimont, S, Berche, P.** 2000. NhaA, an Na(+)/H(+) antiporter involved in environmental survival of Vibrio cholerae. J. Bacteriol. **182:**2937-2944.

43. **Reitzer, L, Schneider, B.** 2001. Metabolic context and possible physiological themes of sigma (54)-dependent genes in Escherichia coli. Microbiology and Molecular Biology Reviews: MMBR. **65:**422-44, table of contents.

44. **Model, P, Jovanovic, G, Dworkin, J.** 1997. The Escherichia coli phage-shock-protein (psp) operon. Mol. Microbiol. **24:**255-261.

45. **Okada, Y, Okada, N, Makino, S, Asakura, H, Yamamoto, S, Igimi, S.** 2006. The sigma factor RpoN (sigma54) is involved in osmotolerance in Listeria monocytogenes. FEMS Microbiol. Lett. **263:**54-60. doi: FML405 [pii].

46. **Bittner, M, Saldías, S, Estévez, C, Zaldívar, M, Marolda, CL, Valvano, MA, Contreras, I.** 2002. O-antigen expression in Salmonella enterica serovar Typhi is regulated by nitrogen availability through RpoN-mediated transcriptional control of the rfaH gene. Microbiology. **148:**3789-3799.

47. **Cytryn, EJ, Sangurdekar, DP, Streeter, JG, Franck, WL, Chang, WS, Stacey, G, Emerich, DW, Joshi, T, Xu, D, Sadowsky, MJ.** 2007. Transcriptional and physiological responses of Bradyrhizobium japonicum to desiccation-induced stress. J. Bacteriol. **189:**6751-6762. doi: JB.00533-07 [pii].

48. **Hung, MN, Rangarajan, E, Munger, C, Nadeau, G, Sulea, T, Matte, A.** 2006. Crystal structure of TDP-fucosamine acetyltransferase (WecD) from Escherichia coli, an enzyme required for enterobacterial common antigen synthesis. J. Bacteriol. **188:**5606-5617. doi: 188/15/5606 [pii].

49. **Stephenson, SA, Brown, PD.** 2016. Epigenetic Influence of Dam Methylation on Gene Expression and Attachment in Uropathogenic Escherichia coli. Frontiers in Public Health. **4:**.

50. **Chatti, A, Messaoudi, N, Mihoub, M, Landoulsi, A.** 2012. Effects of hydrogen peroxide on the motility, catalase and superoxide dismutase of dam and/or seqA mutant of Salmonella typhimurium. World Journal of Microbiology and Biotechnology. **28:**129-133.

51. **Dörr, T, Lewis, K, Vulić, M.** 2009. SOS response induces persistence to fluoroquinolones in Escherichia coli. PLoS Genet. **5:**e1000760.

52. **Atwood, DN, Beenken, KE, Loughran, AJ, Meeker, DG, Lantz, TL, Graham, JW, Spencer, HJ, Smeltzer, MS.** 2016. XerC Contributes to Diverse Forms of Staphylococcus aureus Infection via agr-Dependent and agr-Independent Pathways. Infect. Immun. **84:**1214-1225. doi: 10.1128/IAI.01462-15 [doi].

53. **De Lay, N, Schu, DJ, Gottesman, S.** 2013. Bacterial small RNA-based negative regulation: Hfq and its accomplices. J. Biol. Chem. **288:**7996-8003. doi: 10.1074/jbc.R112.441386 [doi].

54. **Yang, G, Wang, L, Wang, Y, Li, P, Zhu, J, Qiu, S, Hao, R, Wu, Z, Li, W, Song, H.** 2015. hfq regulates acid tolerance and virulence by responding to acid stress in Shigella flexneri. Res. Microbiol. **166:**476-485.

55. **Shakhnovich, EA, Davis, BM, Waldor, MK.** 2009. Hfq negatively regulates type III secretion in EHEC and several other pathogens. Mol. Microbiol. **74:**347-363.

56. **Walburger, A, Lazdunski, C, Corda, Y.** 2002. The Tol/Pal system function requires an interaction between the C-terminal domain of TolA and the N-terminal domain of TolB. Mol. Microbiol. **44:**695-708.

57. **Thompson, KM, Gottesman, S.** 2014. The MiaA tRNA modification enzyme is necessary for robust RpoS expression in Escherichia coli. J. Bacteriol. **196:**754-761. doi: 10.1128/JB.01013-13 [doi].

58. **Bearson, SM, Bearson, BL, Rasmussen, MA.** 2006. Identification of Salmonella enterica serovar Typhimurium genes important for survival in the swine gastric environment. Appl. Environ. Microbiol. **72:**2829-2836. doi: 72/4/2829 [pii].

59. **Danese, PN, Oliver, GR, Barr, K, Bowman, GD, Rick, PD, Silhavy, TJ.** 1998. Accumulation of the enterobacterial common antigen lipid II biosynthetic intermediate stimulates degP transcription in Escherichia coli. J. Bacteriol. **180:**5875-5884.

60. **Dalebroux, ZD, Matamouros, S, Whittington, D, Bishop, RE, Miller, SI.** 2014. PhoPQ regulates acidic glycerophospholipid content of the Salmonella Typhimurium outer membrane. Proc. Natl. Acad. Sci. U. S. A. **111:**1963-1968. doi: 10.1073/pnas.1316901111 [doi].

61. **Paterson, GK, Cone, DB, Northen, H, Peters, SE, Maskell, DJ.** 2009. Deletion of the gene encoding the glycolytic enzyme triosephosphate isomerase (tpi) alters morphology of Salmonella enterica serovar Typhimurium and decreases fitness in mice. FEMS Microbiol. Lett. **294:**45-51.

62. **Pilar, AVC, Reid-Yu, SA, Cooper, CA, Mulder, DT, Coombes, BK.** 2012. GogB is an anti-inflammatory effector that limits tissue damage during Salmonella infection through interaction with human FBXO22 and Skp1. PLoS Pathog. **8:**e1002773.

63. **Cao, Y, Wu, J, Liu, Q, Zhao, Y, Ying, X, Cha, L, Wang, L, Li, W.** 2010. sRNATarBase: a comprehensive database of bacterial sRNA targets verified by experiments. RNA. **16:**2051-2057. doi: 10.1261/rna.2193110 [doi].

64. **Solaimanpour, S, Sarmiento, F, Mrázek, J.** 2015. Tn-seq explorer: a tool for analysis of high-throughput sequencing data of transposon mutant libraries. PloS One. **10:**e0126070.

65. **Finn, S, Condell, O, McClure, P, Amézquita, A, Fanning, S.** 2013. Mechanisms of survival, responses and sources of Salmonella in low-moisture environments. Frontiers in Microbiology. **4:**331.

66. **Breeuwer, P.** 2014. Adaptation of Pathogenic Microorganisms to Dry Conditions, p. 37-48. *In* Anonymous The Microbiological Safety of Low Water Activity Foods and Spices. Springer.

67. **Chao, MC, Abel, S, Davis, BM, Waldor, MK.** 2016. The design and analysis of transposon insertion sequencing experiments. Nature Reviews Microbiology. **14:**119-128.

68. **Thorvaldsdottir, H, Robinson, JT, Mesirov, JP.** 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. **14:**178-192. doi: 10.1093/bib/bbs017 [doi].

CHAPTER 4:

Essential genome of *Campylobacter jejuni*

Rabindra K. Mandal[1], Tieshan Jiang[1], and Young Min Kwon[1, 2]*

[1]Department of Poultry Science, University of Arkansas, Fayetteville, AR, 72701

[2]Cell and Molecular Biology Program, University of Arkansas, Fayetteville, AR, 72701

*Corresponding author:

Department of Poultry Science

College of Agricultural, Food and Life Sciences

University of Arkansas

Fayetteville, AR 72701

Phone : (479) 575-4935

Email ykwon@uark.edu

ABSTRACT

Campylobacter species are a leading cause of bacterial foodborne illness worldwide. Despite of the global efforts to curb the infection, the *Campylobacter* infections have increased continuously in both developed and developing countries, which warrants development of effective strategies to control the infection. The essential genes of bacteria are usually the most prominent targets for the purpose. In this study, we used transposon sequencing (Tn-seq) of a genome-saturating library of Tn5 insertion mutants to redefine the essential genome of *C. jejuni.* We constructed a Tn5 mutant library of unprecedented level of complexity in *C. jejuni* NCTC 11168 with 95,929 unique insertions throughout the genome, and used the genomic DNAof the library for the reconstruction of Tn5 library in the same (*C. jejuni* NCTC 11168) and different strain background (*C. jejuni* 81-176) through natural transformation. We identified 166 essential protein-coding genes and 20 essential transfer RNA (tRNA) in *C. jejuni* NCTC 11168 which were intolerant to Tn5 insertions during *in vitro* growth. The reconstructed library *C. jejuni* 81-176 had 384 protein coding genes with no Tn5 insertions. Essential genes in both of the strains were highly enriched in cluster of orthologous group (COG) category like 'Translation, ribosomal structure and biogenesis (J)', 'Energy production and conversion (C)', and 'Coenzyme transport and metabolism (H)'. Comparative analysis among this and previous studies identified 50 core essential genes of *C. jejuni*, where most of them have been implicated in the development of drug target and vaccines against a wide range of bacteria. Thus, this essential gene list can be further investigated for the development of novel target to limit the spread of this notorious foodborne pathogen.

Keywords: *Campylobacter*, transposon sequencing (Tn-seq), essential genes,

INTRODUCTION

Campylobacter species are a leading cause of bacterial foodborne illness worldwide and one of the most common infectious agent of the last century. Despite the reduction in incidence of a number of major foodborne pathogens due to global efforts, the *Campylobacter* infections have increased in both developed and developing countries across the globe like USA, Europe, Australia, Africa, Asia and Middle East. Additionally, *Campylobacter* harbors antimicrobial genes with potential of horizontal transfer between pathogenic and commensal microorganism and emergence of multiple drugs resistance Hence, researchers speculate that *Campylobacter* will continue to remain threat to global public health for the years to come [1, 2]. Thus, it warrants multifaceted approaches to intervene and control *Campylobacter* infections including identification of indispensable essential genes, which are related to basic cellular functions or metabolic pathways/processes that have the potential of becoming therapeutic targets and/or vaccine development.

Essential genes are defined as those that are absolutely required for the viability of cellular life [3, 4]. Experimental techniques such as single-gene knockouts [5-7], transposon mutagenesis [8, 9], and antisense RNA and RNA interference [10, 11], have been used to identify essential genes. In addition, computational approach to track down essential genes involves comparative genomics, supervised machine learning, constraint-based methods, and integrative genomics approach based on orthology and phylogeny [12-16]. However, the most reliable method used to define the essential genome is transposon mutagenesis via transposon sequencing (Tn-seq). The basic principle of the approach involves creation of transposon insertion library in bacteria of interest, and identification of individual transposon-genome junction site on a global scale by Tn-seq method The process helps in identification of virtually all dispensable genes, which would

allow to identify the entire set of essential genes in the genome simultaneously in a single experiment by negative selection of transposon mutants [17]. These methods with little variations are named as InSeq [18], TraDIS [19], HITS [20], Tn-seq circle [21], Tn-seq[22], and RB-TnSeq [23]. Recently, Hutchison III et al. (2016) have used improved transposon mutagenesis methods for identification quasi-essential genes, which were then used as a basis to create minimal synthetic bacterial genome JCVI-syn3.0, smaller than the genomes of any autonomously replicating cell found in nature [24].

Stahl and Stintizi (2011) identified 195 essential genes of *Campylobacter jejuni* (*C. jejuni*) NCTC 11168 required for growth at 37ºC under a microaerophilic atmosphere on a rich Muller-Hinton medium with 7,201 individual mutants (Tn5) using microarray transposon-based tracing approach [25]. Furthermore, Metris et al. (2011) also identified 233 essential genes of *C. jejuni* NCTC 11168 strain *in vivo* based on a total of 9,550 transposon insertions in the genome using two different transposons (Mariner and Tn7) on Blood Agar Base no.2 (Oxoid) plates supplemented with 5% v/v defribinated horse blood at 42ºC under microaerophilic conditions[26]. More recently, Gao et al. (2014) identified 175 essential genes of *C. jejuni* 81-176 based on ~50,000 transposon insertion mutants of C. jejuni 87-176 strain on brucella agar plates at 37ºC in 10% CO_2 atmosphere [27]. However, these studies had little overlap between the essential genes on the lists probably because of the different culture conditions for recovery of the mutants, different strain backgrounds, varying levels of saturation of transposon insertions, and analysis approach. Since these approaches for essential gene discovery are based on the identification of the genomic regions that do not tolerate transposon insertions, the accuracy of essential gene discovery should be critically dependent on the saturation level of the transposon insertion library.

In this study, we created a highly complex Tn5 mutant library of *C. jejuni* NCTC 11168 (seed library) with more than 95,000 unique insertions in the genome. *C. jejuni* NCTC 11168 required 166 essential protein coding genes required for the growth on Muller-Hinton (MH) agar at 37ºC under microaerophilic condition. Additionally, we reconstructed Tn5 mutant library in the same (*C. jejuni* NCTC 11168) and different strain background (*C. jejuni* 81-176) by transferring the insertions in the seed library to the recipient cells via natural transformation to develop and validate a powerful approach for comparative functional genomics of *C. jejuni*. Furthermore, we combined all existing data from the previous and current studies to define a core set of essential genome of *Campylobacter.*

MATERIALS AND METHODS

Bacteria strains and growth conditions

C. jejuni NCTC11168 and *C. jejuni* 81-176 were grown on Muller-Hinton (MH) agar plates at 37ºC under microaerophilic conditions (O_2- 5%, CO_2- 10%, and N_2- Balance). Trimethoprim (TMP, 10 µg/ml,) and Kanamycin (Km, 50 µg/ml) was added to the MH agar when required. The bacteria pellet and extracted DNA were stored at -20ºC. *C. jejuni* frozen stocks were stored at -80ºC in 50% glycerol.

Construction of Tn5 Transposon Mutant Library

Construction of Tn5 seed library

Tn5 transposon mutant library of *C. jejuni* NCTC11168 was generated using EZ-Tn5™ <KAN-2>Tnp Transposome™ Kit (Cat. No. TSM99K2, Epicentre Biotechnologies, Madison, WI, USA) following manufacturer's protocol. Briefly, *in vitro* transposition reaction consisted of 2 µl 10x

EZ-Tn5 reaction buffer, 1 µl of transposome complex, 2 µg of chromosomal DNA, and 15 µl of distilled deionized H_2O and incubated for 4h at 37ºC. The transposed DNA was purified after adding of 60 µl of distilled deionized H_2O followed by phenol-chloroform extraction and then ethanol precipitation of DNA. DNA was recovered in 40 µl TE buffer (pH 8.0). Next, *in vitro* transposed DNA was repaired by adding 40 µl of transposed DNA, 6 µl of T4 DNA polymerase buffer (New England Biolabs, NEB), 4.8 µl of dNTPs mix (2.5 mM), 7.7 µl distilled H_2O, and 1.5 µl T4 DNA polymerase (1 U/ µl, NEB) and incubated at 11ºC for 20 min in thermal cycler. The reaction was inactivated by incubating at 75ºC for 15 min. The second repair reaction consisted of 60 µl reaction mixture (previous rxn), 12 µl T4 DNA ligase buffer (NEB), 1.5 µl T4 DNA ligase (NEB) and 46.5 µl dH_2O and was incubated for overnight at 16ºC. This was followed by DNA dialysis on tip of a nitrocellulose membrane floating on 10-20 ml distilled deionized water for 20 min. All of the reaction was used for one transformation of *C. jejuni* NCTC11168 following the natural transformation method described by Davis et al. 2008 (briefly explained in next section) [25]. Naturally transformed *C. jejuni* NCTC11168 were selected on MH agar plates with TMP and Km. The mutants were scrapped off the plate in 1x PBS, centrifuged, and the pellet was stored at -80ºC. We performed 14 transformations with each producing ~100,000 mutants. Equal volume of mutants from each transformation was combined together to create Tn5 seed mutant library (seed library) as shown in Figure 1.

Reconstruction of Tn5 library

Genomic DNA extracted from the complex Tn5 seed library (Tn5 seed library DNA) was used for reconstruction of Tn5 library in the same and different strain background of *C. jejuni*. Tn5 seed library was naturally transformed into *C. jejuni* NCTC11168 and *C. jejuni* 81-176 background following natural transformation protocol of *C. jejuni* as described by Davis et al.

2008 [28]. Briefly, *C. jejuni* strains from frozen stock was streaked on MH agar plate containing TMP and incubated for 16 h under microaerophilic conditions at 37ºC. Next day, a heavy inoculum from the plate was streaked on MH agar with TMP and incubated for 16 h. The entire bacteria from16 h growth plate was resuspended in 1 ml MH broth without antibiotics and OD_{600} was adjusted to 0.5 in MH broth. One ml fresh melted MH agar (without antibiotics) was pipetted in 5ml plastic test-tube and was allowed to solidify. An aliquot of 0.5 ml bacteria from adjusted OD_{600} was added in test-tube containing 1 ml of solidified MH agar and mixed gently and incubated for 3 h at 37ºC in microaerophilic conditions (bi-phasic medium). Then after, 500 µg of seed library DNA was added to biphasic medium and incubated for 4 h in the above conditions. The transformants were collected in microcentrifuge tube, centrifuged for 2 min and resuspended in MH broth. Finally, the transformants were plated directly or after serial dilutions on MH agar plates supplemented with TMP and Km and incubated for 2 days. The colonies were counted from dilution plates and also collected from direct plates in 1X PBS, and centrifuged. The supernatant was discarded and bacterial pellet was stored at -20ºC (Figure 1).

Transposon junction amplification and sequencing

Genomic DNA was extracted from the bacterial pellets of complex Tn5 libraries using QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA) following manufacturer's protocol. Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA) was used to check the concentration and purity of extracted DNA. Tn-seq DNA library for Illumina sequencing was prepared as previously developed protocol in our laboratory [29] with minor modifications as described in detail in File S1. The first step was to single primer extension step using a primer specific to one end of the transposon. In this study, for each library DNA, the linear extension step was performed either a regular primer, dual priming oligonucleotide (DPO) primer or both. DPO primers was designed

as described by [30], and used to increase the specificity of PCR amplification of the Tn5-chromosome junction sequences. The regular primer (5'-GATCCTCTAGAGTCGACCTGCAGGCATGCA-3') and DPO primer (5'-ACCGTGGCGGGGATCCTCTAGAGTIIIIITGCAGGCAT-3') were located 32 bp and 35 bp upstream of Tn5-genome junction, respectively as shown in Figure S1. Briefly, either regular or DPO primer was used for linear extension PCR using GoTaq G2 hot start colorless master mix (Promega Corporation, Madison, USA). Linear extension was followed by addition of C tail and then exponential PCR to amplify Tn5-genome junction sequences. Then after, exponential PCR product was heated at 65ºC for 15 min mixed with loading buffer and PCR product were ran on 1% agarose gel. DNA from 300-500 bp were gel-purified using Zymoclean™ Gel DNA Recovery Kit following manufacturer's protocol (Irvine, Ca, USA). Oligonucleotide used in this study are shown in Table S1. Equal quantity of DNA (10 ng) were mixed for each library and sent for Illumina HiSeq 4000 single end read with 90 cycles (DNA Technologies Core, UC Davis Genome Center, Davis, CA 95616).

Data analysis

Illumina sequencing reads were demultiplexed allowing prefect match of barcode and transposon mosaic end using custom Perl script. The 22 bp genomic junction sequence were extracted and used for downstream analysis in different manners for (1) the seed library and the library reconstructed in the same *C. jejuni* NCTC11168 strain background, and (2) the library reconstructed in different background (*C. jejuni* 81-176). For the libraries in *C. jejuni* NCTC11168 background, the junction sequences were aligned to the complete sequence of the same genome using Bowtie version 2.2.8 [31]. The aligned SAM mapping file were then fed to EL-ARIST for the analysis of essential genome following user manual [32]. Briefly, *C. jejuni*

NCTC11168 Tn5 library were mapped to 400 bp genomic windows and *C. jejuni* 81-176 Tn5 library to 500 bp genomic windows. Insertion sites were linked to their associated annotated genomic features. There was no obvious insertion bias according to the insertion sites in respect to the replication of origin as shown in Figure 2 (A, B, C, and D). Thus, raw data were used for downstream analysis. Then, sliding window analysis was used to define regions with lower read counts, which were used to train a hidden Markov model to predict each window to be essential or non-essential for growth. Because of the differences in read numbers and complexity of Tn5 library, we used the sliding windows of different sizes and different p-value threshold for calling a region significantly underrepresented in reads appropriate for each Tn5 library as described in Table S2.

Additionally, we used Tn-seq Explorer to assign the number of unique Tn5 insertions sites and read counts to each gene using Bowtie aligned SAM files [33]. Tn5 insertion read counts and unique insertion site were only considered in the central 80% of the gene (CDS) excluding insertion from the beginning and end 10% of the gene. While, for transfer RNA (tRNA), ribosomal (rRNA) and pseudogenes unique insertion and read counts were considered for the whole gene length.

On the contrary, to the libraries in *C. jejuni* 11168 background, we had to employ a different strategy for the downstream analysis using the 22 bp junction sequences for the library reconstructed in *C. jejuni* 81-176 background due to the previously known differences in the genomic regions [34]. For this library, genomic DNA from the seed library (*C. jejuni* NCTC 11168) was transferred to *C. jejuni* 81-176 through natural transformation. For the incoming genomic DNA fragments containing Tn5 insertion to integrate into the recipient genome, there should be sufficient homology between the two strains at DNA level. To determine the

homology levels in *C. jejuni* 81-176, DNA sequence flanking 1,000 bp upstream and downstream sequences of the coding sequences(CDS) was extracted along with the CDS for all genes in *C. jejuni* NCTC 11168 using custom Python script and BLASTed against *C. jejuni* 81-176 from command-line interface with BLASTn tabular output format 6. Single best BLAST hit was kept for each query sequence based on the highest bit score that gives indication of how good the alignment is with higher score having the better alignment. Then after, BLAST output table was filtered so that gene having high homology with flaking sequences can be retained. We arbitrarily used the following combination of condition to filter the BLAST tabular output file: alignment over > 55% of query length, percent identity ≥ 98%, mismatches < 50 nucleotide and gaps < 5 with were kept. Next, the genes of *C. jejuni* NCTC 11168 having high probability of homologous recombination in *C. jejuni* 81-176 were searched again for orthologous gene present in *C. jejuni* 81-1176. Only the orthologous gene with flanking sequence of *C. jejuni* 81-176 having higher homology in *C. jejuni* NCTC 11168 were considered for the analysis of essential genome as described above (Table S3).

RESULTS AND DISCUSSION

Evaluation and comparison of the libraries based on Illumina sequencing data

We generated a complex library of *C. jejuni* NCTC 11168 following natural transformation of *in vitro* mutagenized genomic DNA with commercially available Tn5 transposome complex (EZ-Tn5™ <KAN-2>Tnp Transposome™ Kit). Fourteen natural transformation were performed each producing ~100,000 mutants with 1.4 million mutants. Previously, various strategies have been attempted for efficient transposon mutagenesis of *C. jejuni*. For *in vivo* mutagenesis, the preformed Tn5 transposome complex was introduced into various *C. jejuni* strains via

electroporation [35]. This strategy yielded ~3,000 random mutants per electroporation for *C. jejuni* 81-176 strain, but the efficiency was extremely low for other strains tested, limiting the application of the approach. On the contrary, for *in vitro* mutagenesis, transposition reaction was conducted on the genomic DNA of *C. jejuni* using the purified transposon sequence plus purified transposase enzyme of either Tn5 [36] *mariner* [37]or Tn552 [38] Then the *in vitro* mutagenized DNA was used to transform *C. jejuni* cells through natural transformation, yielding 3,000-7,000 transposon mutants per reaction. In the current study, we achieved the efficiency of transposon mutagenesis far higher than previously reported (100,000 vs. 3,000-7,000 transposon mutants per transformation). We speculate that the high efficiency in our study was due to the use of preformed transposome complex for *in vitro* mutagenesis of genomic DNA in our study as compared to that transposase and transposon sequences were separately added into the reaction in all previous studies on *in vitro* transposon mutagenesis in *Campylobacter* [36-38].

Equal volume of the mutant pools was combined to make seed library (S-CJ11168). Genomic DNA of the seed library was used for the reconstruction of Tn5 mutant library. We collected 281,000 mutants from natural transformation of seed library in the same strain background (*C. jejuni* NCTC 11168: R-CJ11168-D) and 82,000 mutants in different strain background (*C. jejuni* 81-176: R-CJ81176-D) as shown in Figure 1 and Table 1.

Transposon junction sequence were amplified using two different primers (regular and DPO primer) in the linear extension step with downstream process remaining the same for all DNA library. Regular primer was used for linear extension of seed library *C. jejuni* NCTC 11168 (S-CJ11168) while DPO primer for seed library *C. jejuni* NCTC 11168 (S-CJ11168-D), reconstructed library in same strain background *C. jejuni* NCTC 11168 (R-CJ11168-D), and different strain background *C. jejuni* 81-176 (R-CJ81176-D). Dual priming oligonucleotide

(DPO) is believed to block mismatched priming, thereby accomplished of higher PCR specificity [30]. The site of Tn5 insertion in each Tn5 library was determined through next-generation sequencing on HiSeq platform. Demultiplexing Illumina sequencing reads without any mismatch in the barcode sequence and Tn5 mosaic end produced 9,040,241 reads for S-CJ11168; 6,920,934 reads for S-CJ11168-D; 6,052,446 reads for R-CJ11168-D and 1,638,463 reads for R-CJ81176-D. Similarly, R-CJ1168-D had highest reads per unique insertion and R-CJ81176-D had lowest reads per unique insertion with S-CJ11168 having highest and R-CJ81176-D having lowest median reads per unique insertion as shown in Table 1.

Next, sequencing reads with 22 bp transposon genomic junction sequence were mapped to respective genome using default parameter of Bowtie2.2.8, which reports best alignment. Interesting to note, *C. jejuni* NCTC 11168 has 30.6% GC content [39]. However, the GC content of 22bp genomic sequence across all the Tn5 library was little higher 40.25% (SE ± 2.22) that might reflect the preference of Tn5 transposon towards guanosine (G) and cytidine (C) rich sequences [40]. The overall alignment rate was 85.50% (SE ± 5.46). Regular primer had significantly lower alignment rate as compared to DPO primer library (75.36% vs 92.87 % respectively) clearly indicating the higher specificity in binding target DNA for the DPO primer. It was also observed that regular primer produced lower standard error (SE) with mean reads per unique insertion as compared to DPO primer (0.90 vs 11.54, respectively, for S-CJ11168 library). Thus, from this data, we can conclude that DPO primer produced better sequence library than regular primer (Table 1).

Furthermore, the seed library S-CJ11168 and S-CJ11168-D had the most unique insertions throughout the genome 95,920 and 79,178, respectively, followed by reconstructed library in same strain background (R-CJ11169-D: 52,607) and different strain background (R-CJ81176-D:

29,565) (Table 1). *C. jejuni* NCTC 11168 libraries had 47,090 shared unique insertions genome-wide. Importantly, only 2218 (2.1%) of the Tn5 insertion sites were unique to the reconstructed library of *C. jejuni* NCTC 11168 (R-CJ11168-D). Similarly, the seed library amplified using regular and DPO primer had 73,649 (71%) unique insertions in common (Figure S2). The fact that significantly lower number of unique insertions were detected in R-CJ11168-D in comparison to S-CJ11168-D (52,607 vs. 79,178) may be due to the insufficient number Tn5 mutants (281,000 mutant colonies) collected to form the mutant pool during the experiment. The considerably lower number of unique insertions in R-CJ81176-D in comparison to R-CJ11168-D (29,565 vs. 52,607) is probably due to the genomic differences in the two strains, and the insufficient number (82,000 mutants) of Tn5 mutants collected to form the reconstructed library R-CJ81176-D. The complexity of reconstructed libraries can be increased by increasing the number of Tn5 mutants with more natural transformation of the seed library.

Identification of essential gene in *Campylobacter jejuni* NCTC 11168

We used EL-ARTIST pipeline for the identification of essential genes of *C. jejuni* required for optimal growth on MH agar plates under microaerophilic condition at 37˚C. The *Campylobacter* Tn5 libraries had no noticeable replication bias in the reads distribution throughout the genome as shown in Figure 2 (A, B, C, and D) which is imparted by 'V' shaped read distribution with higher reads at origin of replication. This can be imparted due to higher doubling time (112 min) of *C. jejuni* NCTC 11168 grown in MH broth as compared to lower doubling time of some bacteria such as, *V. cholera* has doubling time of 16-20 min grown in rich media [41, 42]. Reads counts of Tn5 libraries were mapped to 400 bp genomic windows against *C. jejuni* NCTC 11168 genome. A high Spearman correlation was observed between the seed library prepared with regular (S-CJ11168) and DPO primer (S-CJ11168-D) based on read counts binned to 400 bp

window size ($R^2 = 0.95$, $p < 0.0001$) as shown in Figure 2 (E). Besides, a little lower Spearman correlation ($R2 = 0.92$, $p < 0.0001$) was observed amid the reconstructed library in same genetic background (R-CJ11168-D) with seed libraries prepared with DPO primer (S-CJ11168-D) as shown in Figure 2 (F and G).

Tn5 sequencing analysis using EL-ARTIST pipeline revealed ~250 genes essential in *C. jejuni* NCTC 11168 required for optimal growth on rich MH agar under microaerophilic condition at 37ºC. Seed library *C. jejuni* NCTC 11168 (S-CJ11168) required 280 genes (15 domain essential and 265 entirely essential genes) for optimal growth. Likewise, the same seed library prepared with DPO primer (S-CJ11168-D) had 278 essential genes (9 domain essential and 269 entirely essential genes) and reconstructed library in same genetic background (R-CJ11168-D) had 284 essential genes (18 domain essential and 266 entirely essential genes) in the same condition. All these three libraries shared 200 common essential genes (Figure S3).

However, when we looked Tn5 insertion at gene level, gene like *spoT* with 64 Tn5 unique insertions with 358 reads (in S-CJ11168 library) were called as essential for optimal growth. Furthermore, gene with less than 400 bp (window size used for EL-ARTIST) were not sensitive to be picked by the EL-ARTIST pipeline. In addition, windows smaller than 400 bp such as 100 bp can give more false positive result due to lack of insertions in many 100 bp windows. Bioinformatics pipeline analysis results are subjected to variation depending upon the algorithm and statistical power. Thus, genes having no insertion in all the three Tn5 libraries of *C. jejuni* NCTC 11168 were considered to be essential for growth in this study.

Genes that were not able tolerate Tn5 insertion in central 80% of the gene in all the three libraries of *C. jejuni* NCTC 11168 were considered essential. We identified 166 essential coding

sequence (CJ-11168) of *C. jejuni* NCTC 11168 with no Tn5 insertion (Table S2). 52.4% of the essential genes of *C. jejuni* NCTC 11168 were on negative strand while 47.5% were on positive strand. Genes that contain even one insertion can be called as non-essential but genes lacking insertions cannot be necessarily classified as essential due to sequence bias of Tn5 inseritons and the smaller genes with lower chance of transposon insertion [26]. However, in this study, there was significantly low correlation between the Tn5 insertion read counts and gene length (Spearman correlation = 0.1852, $p < 0.0001$) (Figure 3A). Also, Tn5 transposon are inserted randomly throughout the entire genome with some preference towards GC rich DNA sequences [40]. Nonetheless, we did not observed any correlation between the Tn5 read inserstion in the central 80% of genes (CDS) and GC content (%) of entire gene (CDS) as shown in Figure 3B (Spearman correlation = 0.0488, $p > 0.0531$).

Furthermore, to back up our analysis at gene level, we looked for the Tn5 insertion in pseudogenes of *C. jejuni* NCTC 11168. Out of total 38 pseudogenes in entire genome of *C. jejuni* NCTC 11168, all of the pseudogene had Tn5 insertion in at least one of the seed libraries of *C. jejuni* NCTC 11168. This also indicates the high complexity Tn5 transposon insertion mutagenesis of seed library. Nonetheless, seed library S-CJ11168 (sequenced with regular primer) and S-CJ11168-D (sequenced with DPO primer) missed Tn5 insertion in only one pseudogene Cj0740 and Cj0742, respectively. However, reconstructed library in same background R-CJ11168-D had zero Tn5 insertions in 11 pseudogenes. This could be due to the limited number of transposon mutants, which can be improved by increasing the number of Tn5 mutants during reconstruction of library.

Next, we assigned essential genes to Cluster of Orthologous Groups (COG) identifier using NCBI FTP site (ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_refseq/Bacteria/). COG

categories highly enriched among 166 essential genes were: J- Translation, ribosomal structure and biogenesis (20.48%); Not in COG (18.07%); M- Cell wall/membrane/envelope biogenesis (10.24%); H- Coenzyme transport and metabolism (8.43%); C- Energy production and conversion (7.83%); I- Lipid transport and metabolism (6.02%). COG moderately enriched were: U- Intracellular trafficking, secretion, and vesicular transport (4.22%); P– Inorganic ion transport and metabolism (3.01%); F- Nucleotide transport and metabolism (2.41%); O- Post-translational modification, protein turnover, and chaperones (2.41%); and R- General function prediction only (2.41%). Also, lower abundant COG categories with only one genes were CP, HR, JO, JT, Q (Secondary metabolites biosynthesis, transport, and catabolism) and TK as shown in Figure 4. Similar to our finding, most commonly enriched COG in essential genes of other bacteria like *Porphyromonas gingivalis*, *Herbaspirillum seropedicae*, *Vibrio cholera*, *Rhodopseudomonas palustris*, *Burkholderia cenocepacia*, and synthetic bacteria *Mycoplasma mycoides* JCVI-syn3.0 were related to Translation, ribosomal structure and biogenesis (J); Cell wall/membrane/envelope biogenesis (M); and Coenzyme transport and metabolism (H)' [8, 9, 24, 43-45].

Interestingly, we identified 20 essential transfer RNA (tRNA) out of total 43 tRNA gene in the entire genome of *C. jejuni* NCTC 1118 and none of the ribosomal RNA gene encoding 16S ribosomal RNA, 23S ribosomal RNA, and 5S ribosomal RNA, which is expected for these genes to be essential for growth on rich MH medium (Table S2). Essential tRNA gene were required for the transport of arginine, asparagine, aspartic acid, glutamic acid, glycine, leucine, lysine, methionine, selenocysteine, serine, tryptophan, tyrosine, and valine. Other essential ncRNA was *rnpB* which is a component RNA of ribonuclease P enzyme (RNAse P) together with RnpA protein (also essential gene identified in our study) acts in processing of 4.5S RNA and tRNA

precursor molecules in *E. coli* (https://ecocyc.org/gene?orgid=ECOLI&id=EG30069). In a recent study, Rosconi et al. (2016) found 22 tRNA and one 23S ribosomal RNA gene to be essential in *Herbaspirillum seropedicae*, an endophyte that colonizes crops like rice and maize, during *in vitro* growth in TY medium similar to our finding [44].

Reconstruction of insertion library in different strain background (*C. jejuni* 81-176)

We transformed the seed library genomic DNA through natural transformation into a different strain background (*C. jejuni* 81-176). The homologous recombination of the seed library genomic DNA into a different strain requires significant homology between the incoming DNA fragments and the target regions at the DNA level. *C. jejuni* strains have a high genomic diversity, and Poly et al. (2005) previously reported that *C. jejuni* 81-176 had 63 kb of new chromosomal DNA sequences and 87 novel genes as compared to *C. jejuni* NCTC 11168 based on microarray analysis [34]. To identify the genomic region in the donor strain with significant homology to the recipient genome, DNA sequence of 1 kb flanking upstream and downstream of the coding sequence (CDS) along with the CDS in *C. jejuni* NCTC 11168 were BLASTed against *C. jejuni* 81-176. This analysis produced 1,535 best BLAST hit based on highest bit score for each query sequence. Further filtering as described in Materials and Methods section resulted in 904 query genes with substantial homology of which 895 are orthologous gene commonly present in both *C. jejuni* 11168 and *C. jejuni* 81-1176 genomes (Table S3). This process was aiming at at eliminating the genes that are unique in *C. jejuni* 81-1176 genome from downstream analysis, so that we can prevent or minimize false identification of the genes with zero Tn5 insertions as essential genes, where the absence of insertion is in fact due to the lack of homology.

Out of 895 orthologous genes with significant homology including 1 kbp flanking sequences, 384 genes had zero Tn5 insertions, and thus considered as essential genes required for the *in vitro* growth of *C. jejuni* 81-176 on rich MH agar at 37°C under microaerophilic conditions. These 384 essential genes (CJ-81176) of *C. jejuni* 81-176 were broadly categorized into cluster of orthologous group (COG). The highly enriched COG were: J- Translation, ribosomal structure and biogenesis (16.67%); Not in COG (13.28%); C- Energy production and conversion (7.55%); H- Coenzyme transport and metabolism (6.25%); R- General function prediction only (5%); E- Amino acid transport and metabolism (5.73%); S- Function unknown (4.95%); M- Cell wall/membrane/envelope biogenesis (4.43%); and F- Nucleotide transport and metabolism (4.17%). Other moderately enriched COGs ranging from 3.5% to 2.5% in abundance were: Posttranslational modification, protein turnover, chaperones (O); Inorganic ion transport and metabolism (P); Replication, recombination and repair(L); Intracellular trafficking, secretion, and vesicular transport (U); Cell cycle control, cell division, chromosome partitioning (D); Carbohydrate transport and metabolism (G); Lipid transport and metabolism (I); and Transcription (K). While the other COGs with 1 or 2 genes had 12.50% in abundance as shown in Figure S4.

Both the strains *C. jejuni* NCTC 11168 and *C. jejuni* 81-176 had similar level of enrichment in COGs like: Nucleotide transport and metabolism (F); Transcription (K); Replication, recombination and repair (L); Posttranslational modification, protein turnover, chaperones (O); Cell cycle control, cell division, chromosome partitioning (D); Carbohydrate transport and metabolism (G); Inorganic ion transport and metabolism (P); Energy production and conversion (C); and Intracellular trafficking, secretion, and vesicular transport (U) with difference in relative abundance ranging from -1.76% to + 1.62%. While, COG category such as ‘Amino acid

transport and metabolism' (E) was relatively higher in *C. jejuni* 81-176 by 4.53% and 'Lipid transport and metabolism' (I) and 'Cell wall/membrane/envelope biogenesis' (M) was higher in *C. jejuni* NCTC 11168 by 3.68% and 5.81 %, respectively (Figure 5). These variations may be observed due to considerable variations in the genome contents of the two strains which in turn can affect their behavior towards environmental stimuli like availability of nutrients and temperature. In addition, the fact that the essential genes in *C. jejuni* 81-176 were identified only from the genomic regions common in both strains could have resulted in some bias in the overall enrichment representations.

Comparative Study

We compared the previously identified essential genes of *C. jejuni* using transposon mutagenesis with our finding. Surprisingly, there was not much correlation between essential genes reported by Metris et al. (2011) and Stahl and Stintzi (2011) and this study in *C. jejuni* NCTC 11168 as shown in Figure 6A. Likely explanations are: 1) limited number of transposon mutants (~ 10,000 mutants) in previous studies, 2) difference in growth conditions: 37˚C vs 42˚C, and MH agar vs Blood Agar, and 3) techniques used for transposon insertion site mapping: microarray used in previous studies vs next-generation sequencing in this study. These arguments are also substantiated by Gao et al. (2014).

Interestingly, substantial overlap was observed between the essential genes of *C. jejuni* 81-176 identified in this report with previously reported by Gao et al. (2014) as shown in Figure 6B, inspite of the considerable variation in experimental procedure. Twenty-three percent of essential genes from this study were common to Gao et al. (2014) and more than 50% of genes identified by Gao et al. (2014) were identified in our study. The primary source of variation with previous

study (Gao et al 2014) was the method of Tn5 library construction. In our study, genomic DNA of *C. jejuni* NCTC 11168 Tn5 library was naturally transformed into *C. jejuni* 81-176 to create Tn5 mutant library. Thus, Tn5 mutagenesis only occurred only at the genomic loci in *C. jejuni* 81-176 with significant amount of homology between the two strains.

Additionally, we were interested in comparing our results with the essential genes used for creation of JCVI-syn3.0, the first synthesized minimal bacterium with the smallest genome (473 genes) capable of self-replication in laboratory media. *C. jejuni* NCTC 11168 protein coding genes were searched for homologous proteins against JCVI-syn3.0 genome using BLASTP with similarity score cutoff of $1e^{-5}$. JCVI-syn3.0 and *C. jejuni* NCTC 11168 had 256 and 311 homologous protein hits, respectively. Among 311 homologous proteins of *C. jejuni*, 52 genes overlapped with the essential genes of *C. jejuni* NCTC 11168 identified in our study (Figure 6C). Unsurprisingly, little overlap was seen between the two bacteria probably due to of the significant differences in their genomic contents.

However, there was substantial overlap between essential genes of *C. jejuni* NCTC 11168 and *C. jejuni* 81-176 in this study (Figure 6D). Approximately, 65% percent essential genes of *C. jejuni* NCTC 11168 were common to *C. jejuni* 81-176. While fewer essential genes of *C. jejuni* 81-176 was shared (~28%) with *C. jejuni* NCTC 11168. This can be due to false assumption of genes with zero Tn5 insertion read count that led to increased essential gene list with no Tn5 insertion in *C. jejuni* 81-176. This might be possibly due to lack homologous recombination site in two different strain background of *C. jejuni* despite extensive filtering for homologous sequence to reduce the noise in data analysis. Other important factor that contributed to disagreement was the insufficient number of Tn5 mutants collected during reconstruction of library.

Notably, the essential genes of *C. jejuni* NCTC 11168 identified in this study had extensive homologs hit in Database of Essential Genes (DEG). DNA sequence of 166 essential genes of *C. jejuni* NCTC 11168 were BLASTed using BLASTX against 46 bacterial essential genes in the database with default parameters (Expect - 1E-05, Score - 100, and Matrix –BLOSUM62). Out of 166 essential protein-coding genes of *C. jejuni* NCTC 11168, 135 genes had homologs in the DEG and 2,879 DEG genes had homologs with our essential genes of *C. jeuni* NCTC 11168. Most of the essential genes with no hit in the DEG were hypothetical proteins (15 genes), 6 integral membrane proteins (Cj0369c, Cj0423, Cj0430, Cj0544, Cj0564, and Cj0851c), 3 periplasmic proteins (Cj0659c, Cj0854c, and Cj0114), and other genes were *pseH*, *rnpA*.

Next, we looked into the core essential genes of *C. jejuni* through comparative analysis of all essential genes identified in this research (*C. jejuni* NCTC 11168 and *C. jejuni* 81-176), previous studies by Metris et al. (2011), Stahl and Stintzi (2011), and Gao et al. (2014), and those in the synthetic bacterium, JCVI-syn3.0 as shown in Figure 7 and Table S4. Orthologous genes of essential genes in *C. jejuni* 81-176 in *C. jejuni* NCTC 11168 background and homologous proteins of JCVI-syn3.0 against *C. jejuni* NCTC 11168 were considered. There were 50 genes common to the six studies with each genes shared among at least four of the studies as shown in Figure 7. Most of the genes belonged to COG category 'Translation, ribosomal structure and biogenesis' (J, 34%), 'Carbohydrate metabolism and transport' (G, 10%), followed by genes not in COG (8%), 'Coenzyme transport and metabolism' (H, 6%) and 'Intracellular trafficking, secretion, and vesicular transport' (U, 6%). The ribosomal proteins are the most prominent drug targets in bacteria that have been used to control infections [46].

NrdF and *folD* were identified as essential in all of the above 6 studies while *ftsY*, *fba*, *engB*, and *rplD* in 5 of the studies. A live attenuated *Salmonella* Typhimurium *aroA* vector expressing

Mycoplasma hyopneumoniae ribonucleotide reductase R2 subunit (NrdF) gene can induce a cell-mediated immune response [47]. Moreover, *ftsY*, *fba*, *engB*, and *rplD* were identified as essential in all 5 studies. Deletion of *ftsY*, gene encoding a signal recognition particle protein in *Streptococcus pneumonia,* induced potent serotype-independent protection against otitis media, sinusitis, pneumonia and invasive pneumococcal disease [48]. Fba encodes class II fructose 1,6-bisphosphate aldolase enzyme important for bacterial, fungal and protozoan glycolysis and gluconeogenesis and is considered as a putative drug target against *Mycobacterium tuberculosis*, the causative agent for tuberculosis [49].

Other essential *Campylobacter* genes shared among at least four of above studies have been used as drug targets and vaccine constructions to mitigate several bacterial infections. BirA, a biotin protein ligase, is an emerging drug target against *E. coli* and other prokaryotes such as *Staphylococcus aureus*, and *Mycobacterium tuberculosis*. Inhibition of *lpxC* by a small-molecule antibiotic in *Acinetobacter baumannii* that has a role in lipid A biosynthesis protects mice from its infection by modulation of inflammation and enhancing opsonophagocytic killing [50]. MreB, a rod shape-determining protein, when blocked by MreB-specific antibiotics inhibits growth of *Chlamydia* [51]. IspH (4-hydroxy-3-methylbut-2-enyl diphosphate reductase) satisfied all criteria of being putative drug target against *Corynebacterium pseudotuberculosis*, pathogenic bacteria that causes caseous lymphadenitis (CLA), ulcerative lymphangitis, mastitis, and edematous in a broad spectrum of hosts [52]. KsgA (rRNA small subunit methyltransferase A) has been associated with clarithromycin resistance in *Mycobacterium tuberculosis* [53]. GalU (UTP-glucose-1-phosphate uridylyltransferase) mutation in *Francisella tularensis*, the causative agent of tularemia, was protective against homologous challenge in mice [54]. GapA (glyceraldehyde

3-phosphate dehydrogenase) was used for the construction of DNA vaccine against *Haemophilus parasuis*, the causative agent of swine polyserositis, polyarthritis, and meningitis [55].

CONCLUSION

We constructed an unprecedentedly complex Tn5 library of *C. jejuni* NCTC 11168 with more than 95,000 unique insertions in the genome. The genomic DNA of the seed library was effectively used for reconstruction of Tn5 library in the same strain background (*C. jejuni* NCTC 11168) and with limited value in different strain background (*C. jejuni* 81-176). Despite discrepancies among studies, comparative analysis of this report showed the core essential genes of *C. jejuni* shared between studies where most of the gene have been implicated as drug target or vaccine development against wide range of bacterial diseases. Usually, all the essential genes have potential to severely affect the survival fitness of a bacterium, which can be further exploited to develop novel strategies to curb this blatant food-borne pathogen.

ACKNOWLEDGEMENTS

We would like to thank Arkansas Biosciences Institute (ABI) for funding support and Arkansas High Performance Computing Center (AHPCC) - University of Arkansas for their computational support.

AUTHOR CONTRIBUTION

YMK- Conceived the study; RKM and TJ- performed experiment; RKM- analyzed data; RKM and YMK- wrote manuscript.

List of Tables

Table 1: Overview of C. jejuni Tn5 mutant libraries.

List of Supplementary Tables

Table S1: List of primers used for DNA library preparation of transposon sequencing.

Table S2: Essential gene of *C. jejuni* NCTC 11168.

Table S3: Essential gene of *C. jejuni* 81-176.

Table S4: Comparative analysis of essential genome of *C. jejuni* with previously identified essential genes.

List of Figures

Figure 1: Design of experiment. EZ-Tn5™ <KAN-2>Tnp Transposome™ Kit was used for in vitro transposition of genomic DNA of C. jejuni NCTC 11168. The transposed DNA was then naturally transformed to C. jejuni NCTC 11168 and mutants were collected MH agar plate with TMP and Km (Seed library: S-CJ11168). Seed library DNA was extracted and used for the reconstruction of Tn5 library in same (C. jejuni NCTC 11168: R-CJ11168-D) and different strain background (C. jejuni 81-176: R-CJ811176-D). Regular and DPO primer was used for linear extension to make DNA library for Illumina sequencing. [S: Seed library, R: Reconstruction, D: Dual priming oligonucleotide (DPO), CJ11168: C. jejuni NCTC 11168; and CJ81176: C. jejuni 81-176].

Figure 2: Overview of Illumina sequencing of C. jejuni Tn5 libraries. Read distribution of Tn5 mutant libraries of C. jejuni: A) S-CJ1118, B) S-CJ11168-D, C) R-CJ11168-D and D) R-CJ81176 (X-axis: Genomic coordinate of Campylobacter; Y-axis: log10 transformed read counts). Spearman correlation (R2) of Tn5 reads distribution based on 400 bp window size

between different libraries of C. jejuni NCTC 11168: E) S-CJ11168 vs S-CJ11168-D, F) S-CJ11168 vs R-CJ11168-D and G) R-Cj11168-D vs S-CJ11168-D. Tn5 libraries are named as in Figure 1.

Figure 3: Tn5 insertion bias analysis in Tn5 libraries of C. jejuni NCTC 11168. Spearman correlation (R2) between: A) gene length (amino acid length) and average Tn5 insertion reads B) GC content (%) of all genes and average Tn5 insertion reads. Average Tn5 reads is the average of Tn5 reads in the central 80% of gene in all the 3 libraries of C. jejuni NCTC 11168 (S-CJ11168, S-CJ11168-D, and R-CJ11168-D).

Figure 4: Cluster of orthologous group (COG) categories of essential genes of C. jejuni NCTC 11168 (CJ-11168). Eessential genes were not able to tolerate any Tn5 insertion in central 80% of genes among all the three libraries of C. jejuni NCTC 11168 (S-CJ11168, S-CJ11168-D, and R-CJ11168-D). Figure at the top of bar is the number of essential genes in that COG category. J- Translation, ribosomal structure and biogenesis; X– Not in COG; M- Cell wall/membrane/envelope biogenesis; H- Coenzyme transport and metabolism; C- Energy production and conversion; I- Lipid transport and metabolism; U- Intracellular trafficking, secretion, and vesicular transport; P– Inorganic ion transport and metabolism; F- Nucleotide transport and metabolism; O- Post-translational modification, protein turnover, and chaperones; R- General function prediction only; D– Cell cycle control, cell division, chromosome partitioning; G- Carbohydrate metabolism and transport; L- Replication, recombination and repair; S- Function unknown; E- Amino acid transport and metabolism; K- Transcription; N- Cell motility; V- Defense mechanisms; Q- Secondary metabolites biosynthesis, transport, and catabolism; and T- Signal transduction mechanisms.

Figure 5. Comparison of major COG categories enriched in essential genome of C. jejuni NCTC 11168 (CJ11168) and C. jejuni 81-176 (CJ81176). COG categories annotations are same as in Figure 4.

Figure 6: Venn diagram indicating numbers of shared essential genes of C. jejuni between this and previous studies. A) Essential genes of C. jejuni NCTC 11168 common between Metris et al. (2011), Stahl and Stintzi (2011) and this study (CJ-11168). B) Essential genes of C. jejuni 81-176 common between Gao et al. (2014) and this study (CJ-811176). Number of genes inside small bracket did not had high homologous sequence in C. jejuni NCTC 11168 background according to our criteria. C) Common genes shared between this study (CJ-11168) and homologous proteins of JCVI-syn3.0 against C. jejuni NCTC 11168. D) Common essential genes between the two Campylobacter strains identified in this study (CJ-11168 - C. jejuni NCTC 11168 and CJ-81176 - C. jejuni 81-176).

Figure 7: Core essential genes of C. jejuni NCTC 11168. Genes are colored to indicate the COG category. Numbers after legend indicate percentage of COG category enriched in core essential gene list. (Soft edge rectangle- Various studies; Metris[26], Stahl-Stintzi [25], Gao-O [27]; JCVI-syn3-H [24]; CJ-811176-O: C. jejuni 81-176; CJ-11168: C. jejuni NCTC 11168; O-orthologous; H- Homologs; Oval shape – essential gene). Genes circled in thick red and thin red are common in six and five studies respectively. All the genes in this network were identified as essential in at least four of the studies.

List of Supplementary Figures

Figure S1: Primer design for linear extension of Tn5 junction sequence during DNA library preparation for Illumina Sequencing. DPO-Tn5-Kn2 primer (green) is dual priming oligonucleotide primer and Ez-Tn5 primer3 (blue) is regular primer. ME: Mosaic end sequence.

Figure S2: Number of unique Tn5 insertion sites in Tn5 mutant libraries of *C. jejuni* NCTC 11168. Number inside the bracket is percentage. Tn5 libraries are named accordingly as in the main body of manuscript.

Figure S3: Essential gene of *C. jejuni* NCTC 11168 identified using EL-Artist pipeline. Figures outside brackets are number of identified essential genes and numbers inside bracket are overall percentage (%). Libraries are named accordingly as in the main body of manuscript.

Figure S4: Assignment of cluster of orthologous group (COG) to essential genes of *C. jejuni* 81-176 (CJ-811176). Essential gene were only analyzed for genes of *C. jejuni* 81-176 that had sufficient homologous sequence in upstream and downstream including coding sequence in the *C. jejuni* NCTC 11168 background (Since Tn5 library in *C. jejuni* 811176 were constructed using genomic DNA of Tn5 library of *C. jejuni* NCTC 11168). Figure at the top of bar indicated the number of essential genes.

Figure S5: Schematic diagram of DNA library preparation of transposon library for Illumina sequencing using DPO primer. Tn-specific primer 1 (DPO-Tn5-Kan2, Table S1) is used for linear extension and Tn-specific primer 2 (library specific barcoded primer according) is used for exponential PCR in conjunction with C-tail specific primer (HTM-Primer, Table S1).

Figure S6: Agarose gel electrophoresis of Tn5 mutant library after exponential PCR. DNA from 300-500 bp is extracted from agarose gel (1%), pooled in equal quantity (10 ng/ sample) and sent for sequencing. [1,2: Tn5 library of *C. jejuni*; 3: Control (genomic DNA of C. jejuni NCTC 11168); and M: Hi-Low DNA Marker).

List of Supplementary File

File S1: DNA library preparation protocol for transposon sequencing (Tn-seq) using dual priming oligonucleotide (DPO). (Figure S5 and Figure S6).

Table 1: Overview of *C. jejuni* Tn5 mutant libraries.

Library	# Tn5 Mutants	Total Reads	Mapped Reads (%)	# Unique Insertions	Mean (± SE)
S-CJ11168	1,400,000	9,040,241	6,812,731 (75.36)	95,929	71.02 ± 0.90
S-CJ11168-D	1,400,000	6,920,934	6,448,244 (93.1)	79,178	81.44 ± 11.54
R-CJ11168-D	281,000	6,052,446	5,685,107 (93.93)	52,607	108.07 ± 13.89
R-CJ81176-D (a)	82,000	1,638,463	1,303,248 (79.54)	29,565	44.08 ± 5.89
*a->CJ11168			1,493,694 (91.16)	32,623	45.79 ± 5.46

S- Seed library; R- Reconstruction; D- DPO; CJ11168- *C. jejuni* NCTC 11168; and CJ81116- *C. jejuni* 81-176. Mean is the average reads per unique insertion in the Tn5 library with standard error (SE). *Tn5 library (R-CJ811176, a) when mapped against donor strain *C. jejuni* NCTC 11168 genome (a->CJ11168), a higher alignment rate was achieved (91.6 %).

Figure 1: Design of experiment**.** EZ-Tn5™ <KAN-2>Tnp Transposome™ Kit was used for *in vitro* transposition of genomic DNA of *C. jejuni* NCTC 11168. The transposed DNA was then naturally transformed to *C. jejuni* NCTC 11168 and mutants were collected MH agar plate with TMP and Km (Seed library: S-CJ11168). Seed library DNA was extracted and used for the reconstruction of Tn5 library in same (*C. jejuni* NCTC 11168: R-CJ11168-D) and different strain background (*C. jejuni* 81-176: R-CJ811176-D). Regular and DPO primer was used for linear extension to make DNA library for Illumina sequencing. [S: Seed library, R: Reconstruction, D: Dual priming oligonucleotide (DPO), CJ11168: *C. jejuni* NCTC 11168; and CJ81176: *C. jejuni* 81-176].

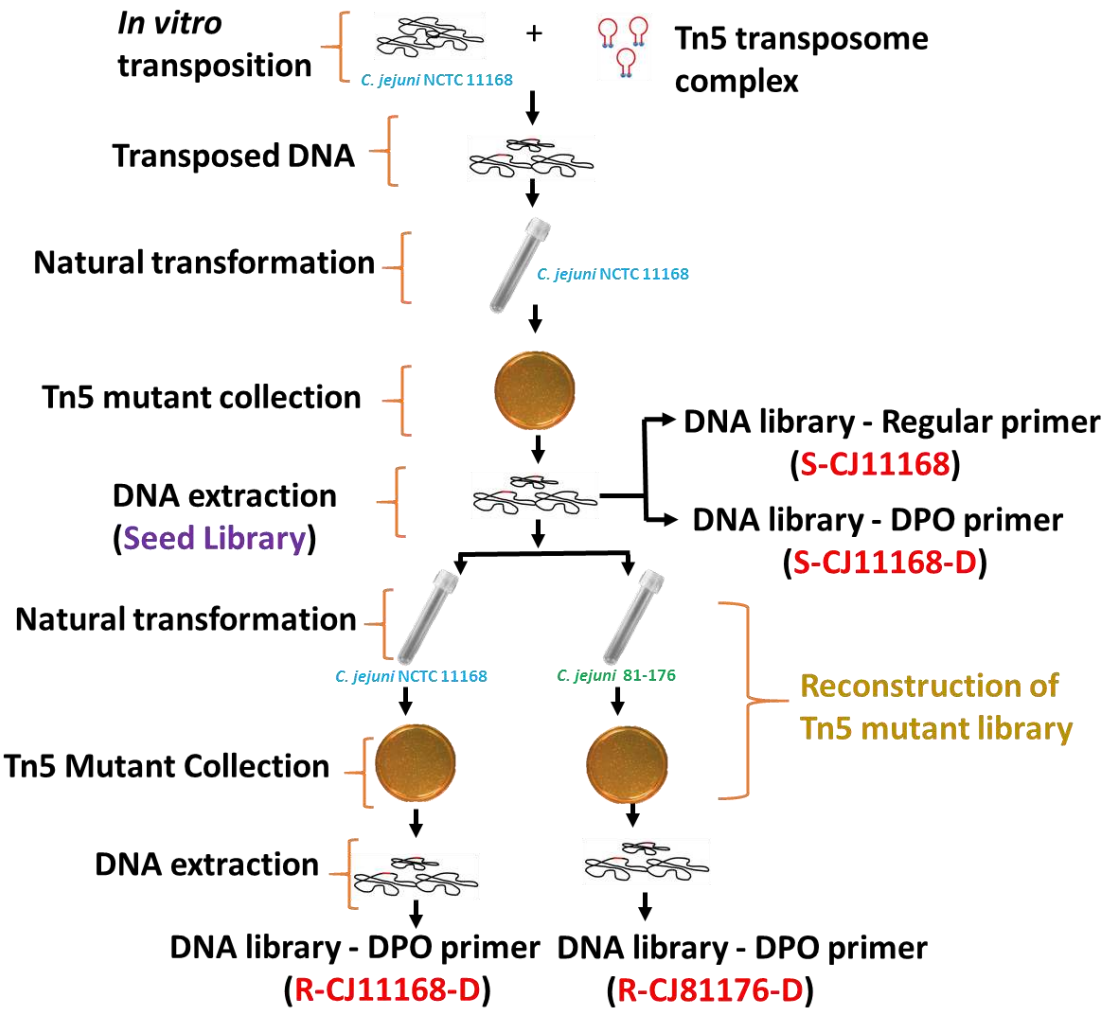


Figure 2: Overview of Illumina sequencing of C. jejuni Tn5 libraries. Read distribution of Tn5 mutant libraries of *C. jejuni*: A) S-CJ1118, B) S-CJ11168-D, C) R-CJ11168-D and D) R-CJ81176 (X-axis: Genomic coordinate of *Campylobacter*; Y-axis: $\log_{10}$ transformed read counts). Spearman correlation (R^2) of Tn5 reads distribution based on 400 bp window size between different libraries of *C. jejuni* NCTC 11168: E) S-CJ11168 vs S-CJ11168-D, F) S-CJ11168 vs R-CJ11168-D and G) R-Cj11168-D vs S-CJ11168-D. Tn5 libraries are named as in Figure 1.

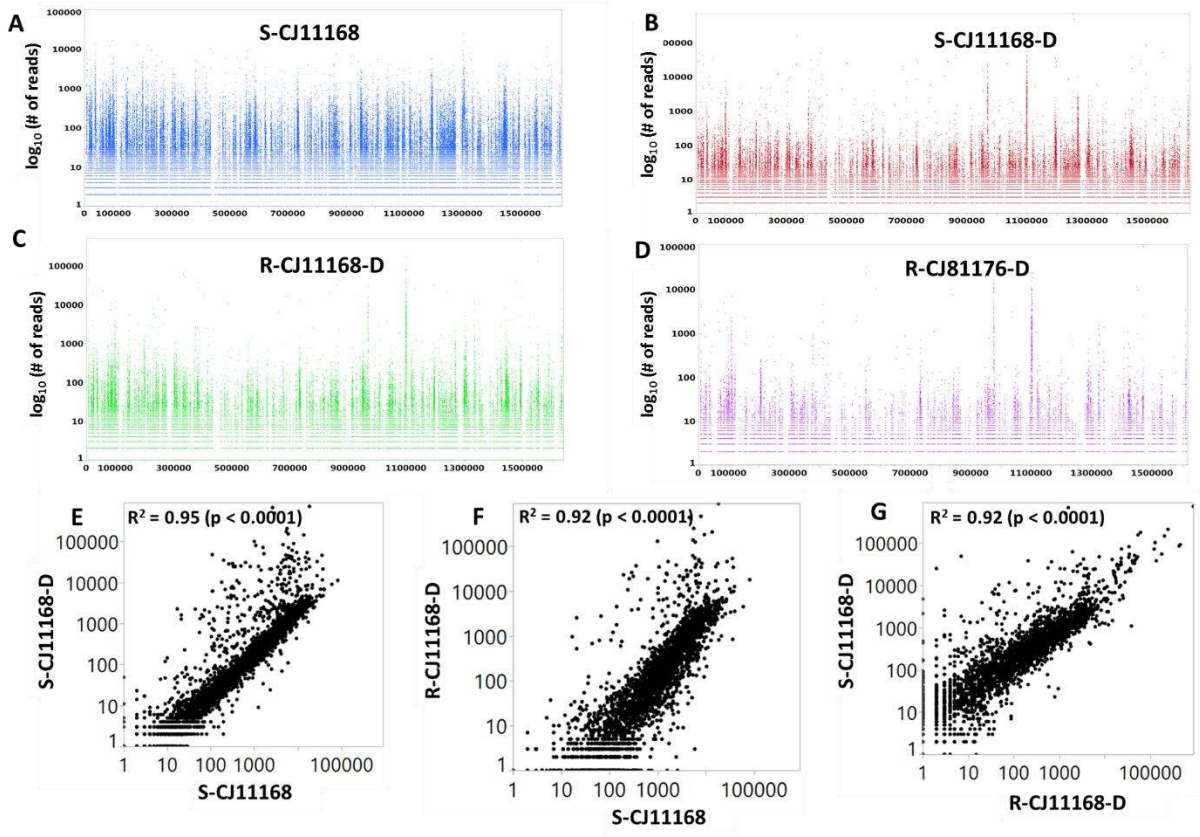


Figure 3: Tn5 insertion bias analysis in Tn5 libraries of *C. jejuni* NCTC 11168. Spearman correlation (R^2) between: A) gene length (amino acid length) and average Tn5 insertion reads B) GC content (%) of all genes and average Tn5 insertion reads. Average Tn5 reads is the average of Tn5 reads in the central 80% of gene in all the 3 libraries of *C. jejuni* NCTC 11168 (S-CJ11168, S-CJ11168-D, and R-CJ11168-D).

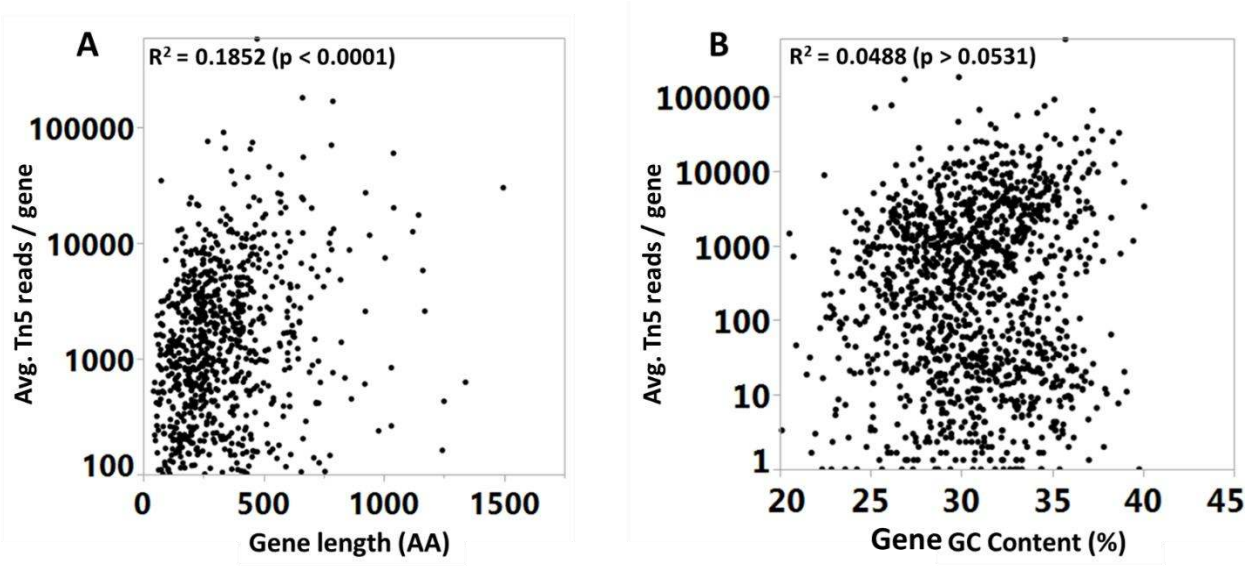


Figure 4: Cluster of orthologous group (COG) categories of essential genes of *C. jejuni* NCTC 11168 (CJ-11168)**.** Eessential genes were not able to tolerate any Tn5 insertion in central 80% of genes among all the three libraries of *C. jejuni* NCTC 11168 (S-CJ11168, S-CJ11168-D, and R-CJ11168-D). Figure at the top of bar is the number of essential genes in that COG category. J- Translation, ribosomal structure and biogenesis; M- Cell wall/membrane/envelope biogenesis; H- Coenzyme transport and metabolism; C- Energy production and conversion; I- Lipid transport and metabolism; U- Intracellular trafficking, secretion, and vesicular transport; P– Inorganic ion transport and metabolism; F- Nucleotide transport and metabolism; O- Post-translational modification, protein turnover, and chaperones; R- General function prediction only; D– Cell cycle control, cell division, chromosome partitioning; G- Carbohydrate metabolism and transport; L- Replication, recombination and repair; S- Function unknown; E- Amino acid transport and metabolism; K- Transcription; N- Cell motility; V- Defense mechanisms; Q- Secondary metabolites biosynthesis, transport, and catabolism; and T- Signal transduction mechanisms.

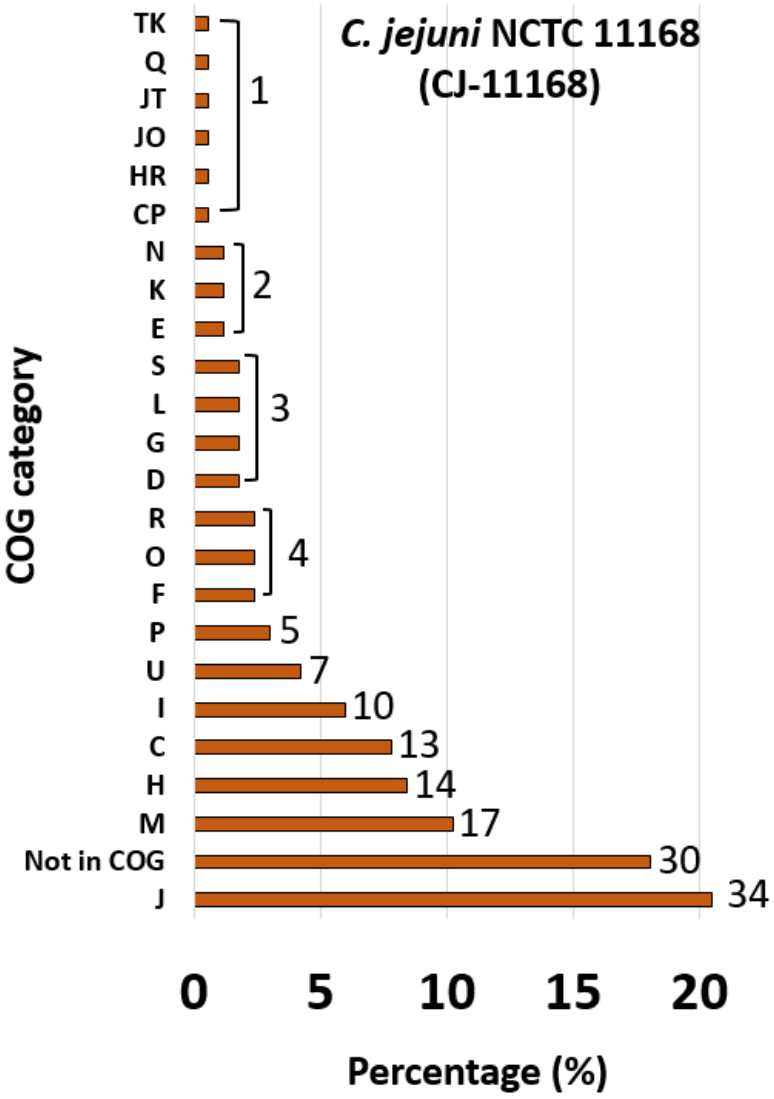


Figure 5. Comparison of major COG categories enriched in essential genome of *C. jejuni* NCTC 11168 (CJ11168) and *C. jejuni* 81-176 (CJ81176). COG categories annotations are same as in Figure 4.

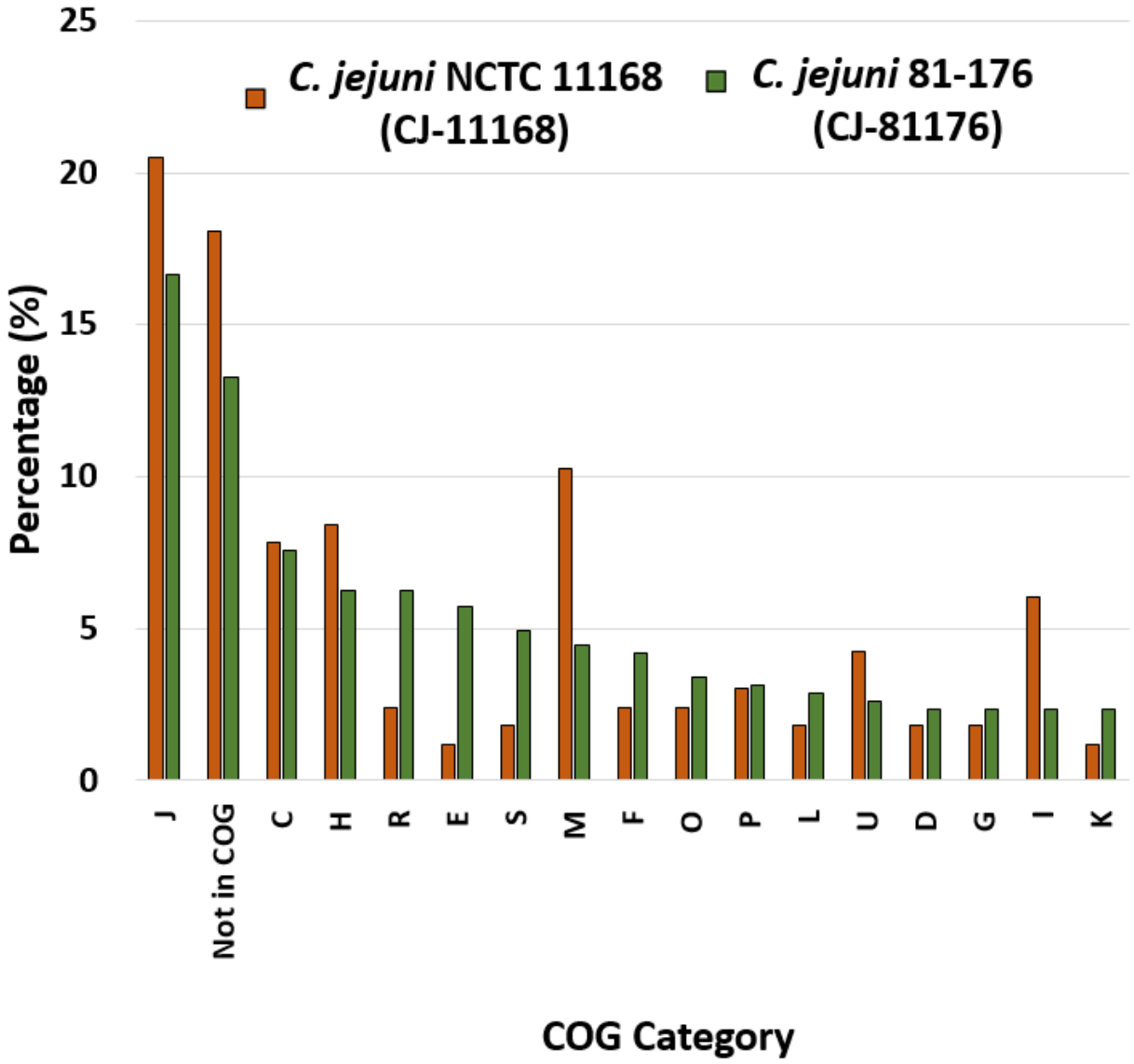


Figure 6: Venn diagram indicating numbers of shared essential genes of *C. jejuni* between this and previous studies. A) Essential genes of *C. jejuni* NCTC 11168 common between Metris et al. (2011), Stahl and Stintzi (2011) and this study (CJ-11168). B) Essential genes of *C. jejuni* 81-176 common between Gao et al. (2014) and this study (CJ-811176). Number of genes inside small bracket did not had high homologous sequence in *C. jejuni* NCTC 11168 background according to our criteria. C) Common genes shared between this study (CJ-11168) and homologous proteins of JCVI-syn3.0 against *C. jejuni* NCTC 11168. D) Common essential genes between the two *Campylobacter* strains identified in this study (CJ-11168 - *C. jejuni* NCTC 11168 and CJ-81176 - *C. jejuni* 81-176).

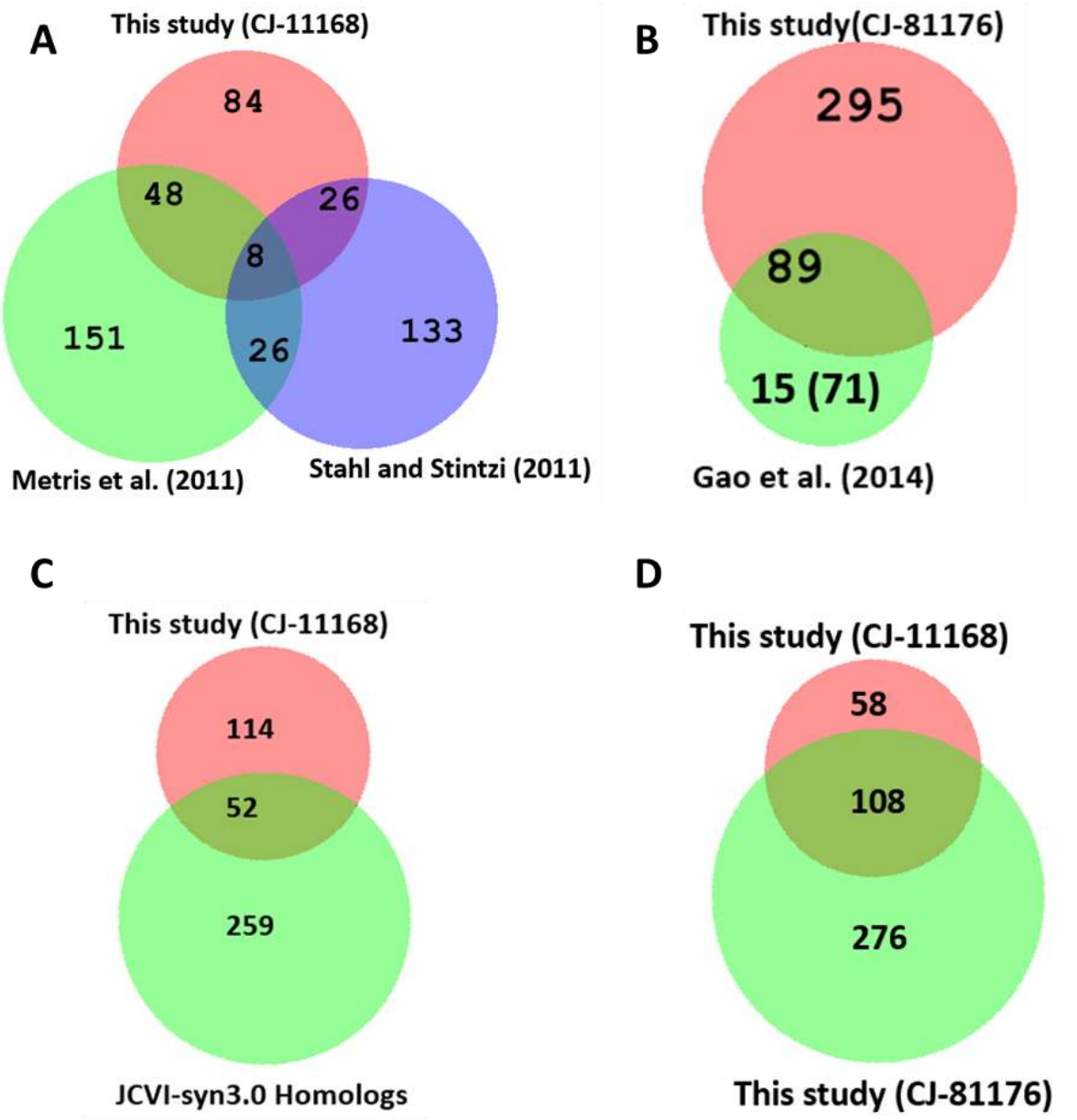


Figure 7: Core essential genes of *C. jejuni* NCTC 11168. Genes are colored to indicate the COG category. Numbers after legend indicate percentage of COG category enriched in core essential gene list. (Soft edge rectangle- Various studies; Metris[26], Stahl-Stintzi [25], Gao-O [27]; JCVI-syn3-H [24]; CJ-811176-O: *C. jejuni* 81-176; CJ-11168: *C. jejuni* NCTC 11168; O- orthologous; H- Homologs; Oval shape – essential gene). Genes circled in thick red and thin red are common in six and five studies respectively. All the genes in this network were identified as essential in at least four of the studies.

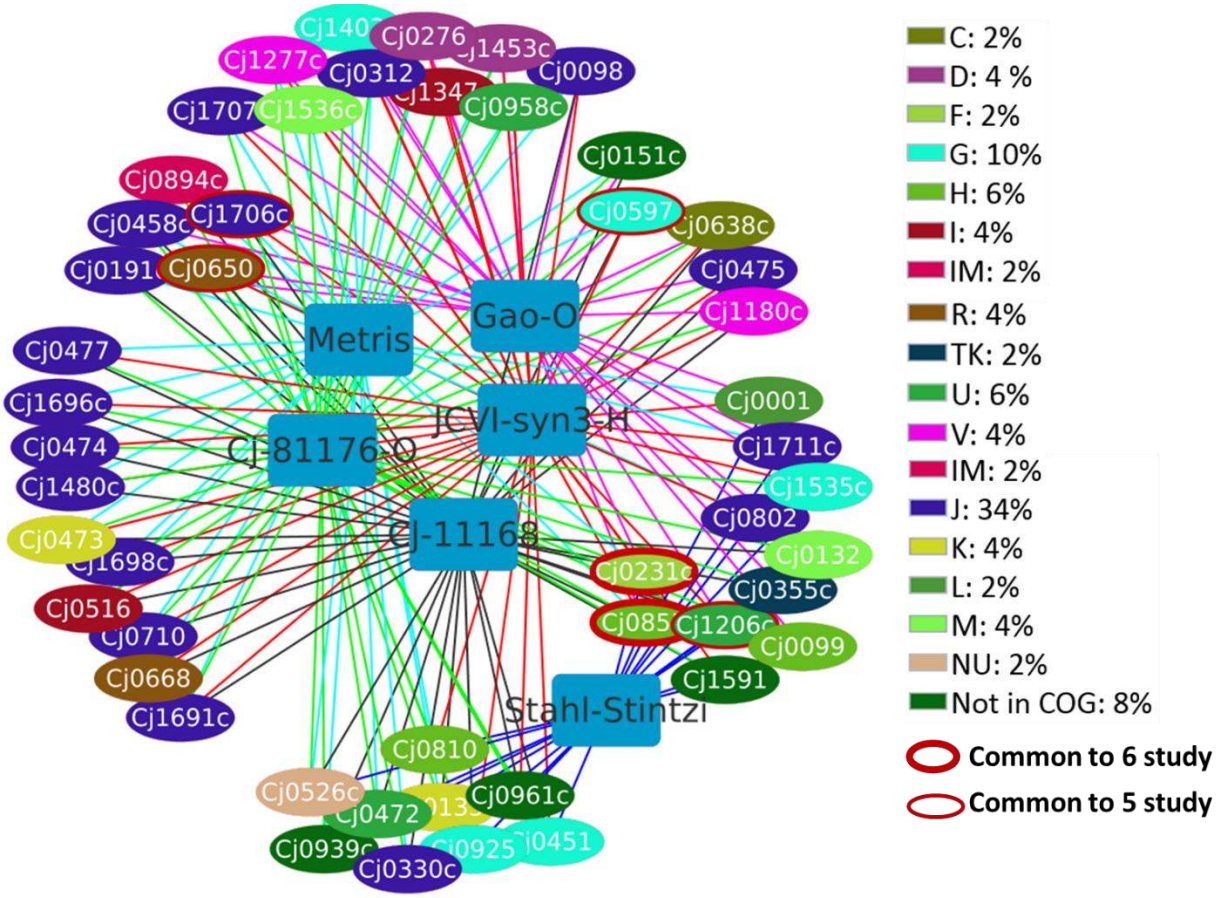


Figure S1: Primer design for linear extension of Tn5 junction sequence during DNA library preparation for Illumina Sequencing. DPO-Tn5-Kn2 primer (green) is dual priming oligonucleotide primer and Ez-Tn5 primer3 (blue) is regular primer. ME: Mosaic end sequence.

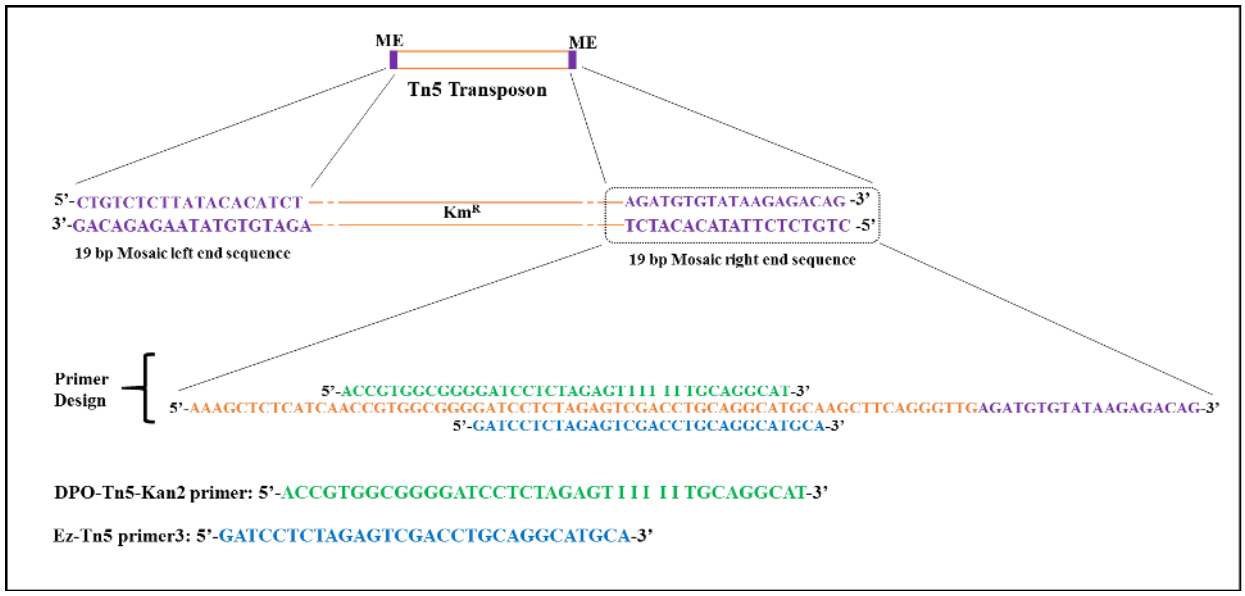


Figure S2: Number of unique Tn5 insertion sites in Tn5 mutant libraries of *C. jejuni* NCTC 11168. Number inside the bracket is percentage. Tn5 libraries are named accordingly as in the main body of manuscript.

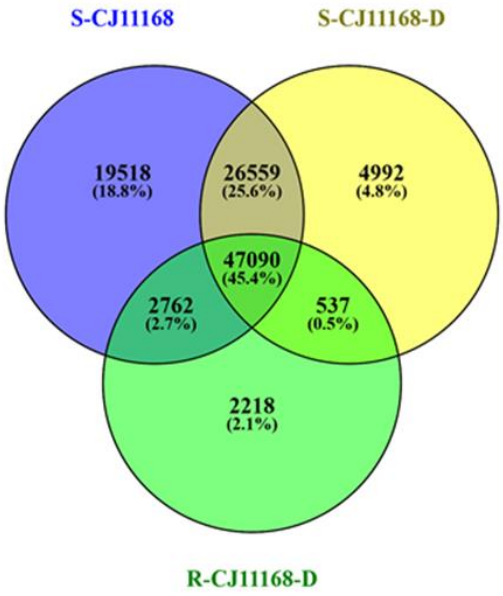


Figure S3: Essential gene of *C. jejuni* NCTC 11168 identified using EL-Artist pipeline. Figures outside brackets are number of identified essential genes and numbers inside bracket are overall percentage (%). Libraries are named accordingly as in the main body of manuscript.

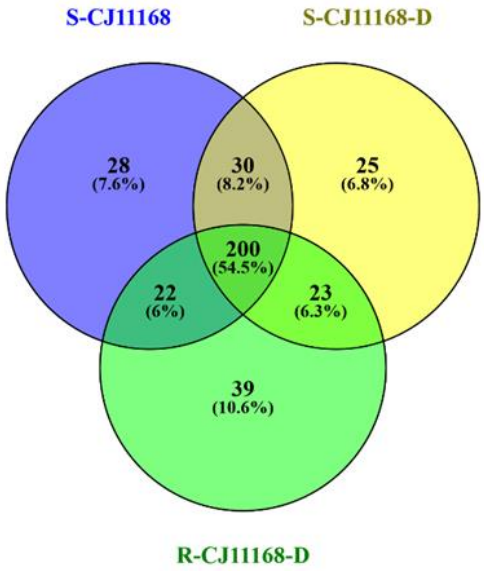


Figure S4: Assignment of cluster of orthologous group (COG) to essential genes of *C. jejuni* 81-176 (CJ-811176). Essential gene were only analyzed for genes of *C. jejuni* 81-176 that had sufficient homologous sequence in upstream and downstream including coding sequence in the *C. jejuni* NCTC 11168 background (Since Tn5 library in *C. jejuni* 811176 were constructed using genomic DNA of Tn5 library of *C. jejuni* NCTC 11168). Figure at the top of bar indicated the number of essential genes.

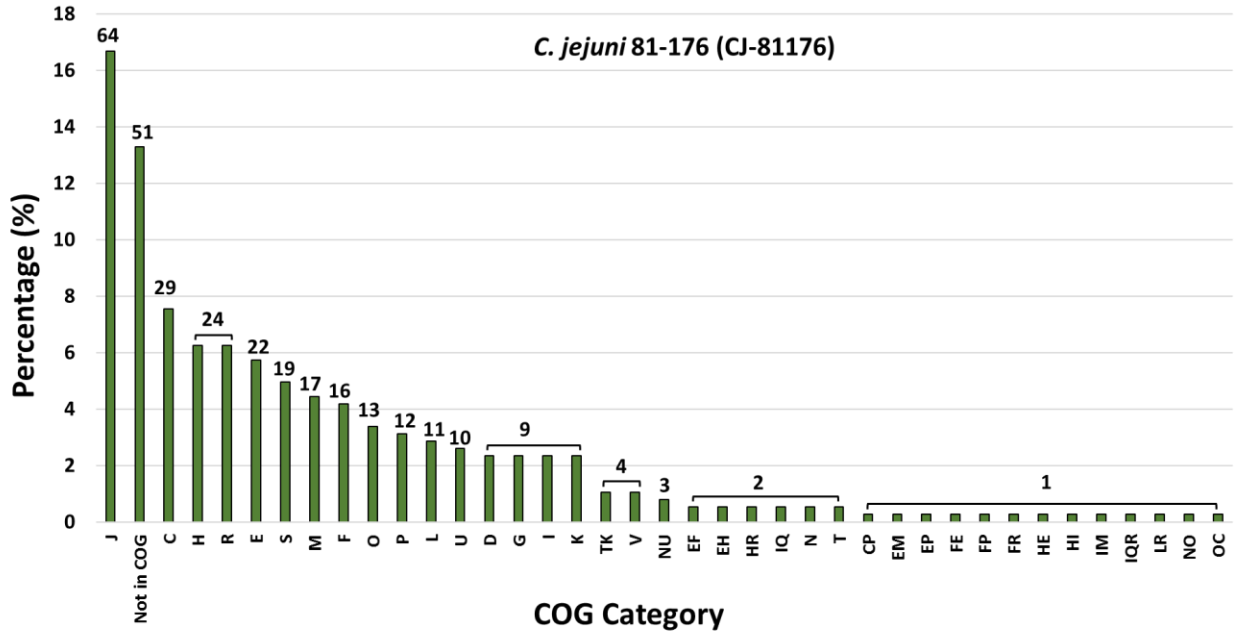


Table S1: List of primers used for DNA library preparation of transposon sequencing.

Primer Name	Sequence (5'-3')	Sample
Ez-Tn5 primer3	5’-GATCCTCTAGAGTCGACCTGCAGGC ATGCA-3’	
DPO-Tn5-Kan2	5’-ACCGTGGCGGGGATCCTCTAGAGTCGA CCTGCAGGCAT-3’	
IR2-IS-B4	5’-AATGATACGGCGACCACCGAGATCTAC ACTCTTTCCCTACACGACGCTCTTCCGA TCTNNNNAGaatgataTCAGGGTTGAGATG TGTATAAGGGACAG-3’	S-CJ11168
IR2-IS-B19	5’-AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATC TNNNNAGatcgacTCAGGGTTGAGATGTGT ATAAGAGACAG-3’	S-CJ11168-D
IR2-IS-B7	5’-AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATC TNNNNAGcagatcTCAGGGTTGAGATGTGT ATAAGAGACAG-3’	R-CJ11168-D
IR2-IS-B8	5’-AATGATACGGCGACCACCGAGATCTACA CTCTTTCCCTACACGACGCTCTTCCGATCT NNNNAGacttgaTCAGGGTTGAGATGTGTAT AAGAGACAG-3’	R-CJ81176-D
HTM-Primer	5’-CAAGCAGAAGACGGCATACGAGCTCTTC CGATCTGGGGGGGGGGGGGGG-3’	

Ez-Tn5 primer3 (regular primer) and DPO-Tn5-Kan2 (dual priming oligonucleotide) were used for linear extension during DNA library preparation. NNNN: four random nucleotide used for efficient clustering. Nucleotide in small letter are barcode designed to allow sorting of Illumina sequence reads according to sample. Barcoded primer and HTM-primer were used for exponential PCR.

File S1:

DNA library preparation protocol for transposon sequencing (Tn-seq) using dual priming oligonucleotide (DPO)

This method of DNA library preparation is improved version of previously developed methodology in our laboratory [1, 2]. A single dual priming oligonucleotide (DPO) primer is used for linear extension of transposon junction [3]. The purified PCR product is subjected to addition of C tail, which is controlled effectively by the mixture of deoxycytidine triphosphate (dCTP) and dideoxy CTP (ddCTP) [4]. Then, the purified C-tailed product is amplified by transposon specific primer and poly G primer.

MATERIALS

DNA extracted from Tn5 mutant library

Wild type DNA (Control DNA)

QIAamp DNA Mini Kit (Qiagen, Valencia, CA, USA)

Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA)

Oligonucleotide **(Supplementary Table 1)**

GoTaq® G2 Hot Start Colorless Master Mix (Promega, WI, USA)

Nuclease-free water

DNA Clean and Concentrator Kit (Zymo Research, Irvine, CA, USA)

Terminal Transferase (TdT, New England Biolabs, Ipswich, MA, USA)

TdT Reaction Buffer (10X)

CoCl2 (2.5 mM) (TdT, New England Biolabs, Ipswich, MA, USA)

dCTP (100 mM) (Promega, Madison, WI, USA)

ddCTP (10 mM) (Promega Madison, WI, USA)

1 % Agarose gel

Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA)

Thermocycler (PCR machine)

STEPS:

Step 1: Linear extension PCR

5. Reaction mixture

Nuclease free H_2O	22 µl
GoTaq G2 Hot Start Colorless Mx	25 µl
DPO-Tn5-Kan2 (10 µm)	1 µl
Genomic DNA of Tn library (~100ng)	2 µl
Total	50 µl

6. PCR cycle

	95°C	2 min
	[95°C	30 sec]
50 cycles	[**63°C**	45 sec]
	[72°C	10 sec]
	4°C	hold

Note: Regular primer was used during the linear extension of the seed library.

7. Purify the linear extension PCR products using DNA clean and concentrator kit. Elute DNA in 11µl EB buffer and store at -20°C.

Step 2: C-tailing reaction

6. Preparation of dNTP working stock

   Dilute 100 mM dCTP to 10 mM dCTP with ddH_2O (nuclease-free)
   Dilute 10 mM ddCTPto 1 mM ddCTP with ddH_2O (nuclease-free)

7. Reaction mixture

DNA (linear extension products)		10.0 µl
TdT Buffer (10X)	2.0 µl	
2.5 mM $CoCl_2$	2.0 µl	
10 mM dCTP	2.4 µl	
1 mM ddCTP	1.0 µl	
ddH_2O	2.1 µl	
Terminal transferase	0.5 µl	
Total	20.0 µl	

8. Incubate the reaction tube at 37°C for 1 hr.

9. Incubate the reaction tube at 75°C for 20 min for heat inactivation of TdT.

10. Purify the C-tailed products using DNA clean and concentrator kit. Elute DNA in 10 µl EB buffer and store at -20°C.

Step 3: PCR to amplify Tn-flanking sequences

2. Reaction mixture

ddH_2O	22.5 µl
GoTaq G2 Hot Start colorless Mx	25 µl
IR2 BC primer with an unique barcode (10 µM)	1 µl
HTM primer (20 µM)	0.5 µl
C-tailed DNA	1 µl
Total	50 µl

8. PCR cycle

	95°C	2 min
	[95°C	30 sec]
36 cycles	[**58°C**	45 sec]
	[72°C	20 sec]
	72°C	10 min
	4°C	hold

Step 4: Gel-purification of PCR products

4. Mix the sample with loading buffer and heat at 65ºC for 15 min.
5. Run 10 µl/sample on 1% agarose gel.
6. Cut 300-500bp bands and gel-purify DNA fragments.

References (File S1):

1. Dawoud TM, Jiang T, Mandal RK, Ricke SC, Kwon YM: Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers. *Mol Biotechnol* 2014, 56(11)**;**1004-1010.

2. Kwon YM, Ricke SC, Mandal RK: Transposon sequencing: methods and expanding applications. *Appl Microbiol Biotechnol* 2016, 100(1)**;**31-43.

3. Chun JY, Kim KJ, Hwang IT, Kim YJ, Lee DH, Lee IK, Kim JK: Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Res* 2007, 35(6)**;**e40.

4. Lazinski DW, Camilli A: Homopolymer tail-mediated ligation PCR: a streamlined and highly efficient method for DNA cloning and library construction. *BioTechniques* 2013, 54(1)**;**25-34.

Figure S5: Schematic diagram of DNA library preparation of transposon library for Illumina sequencing using DPO primer. Tn-specific primer 1 (DPO-Tn5-Kan2, Supplementary Table1) is used for linear extension and Tn-specific primer 2 (library specific barcoded primer according) is used for exponential PCR in conjunction with C-tail specific primer (HTM-Primer, Supplementary Table 1).

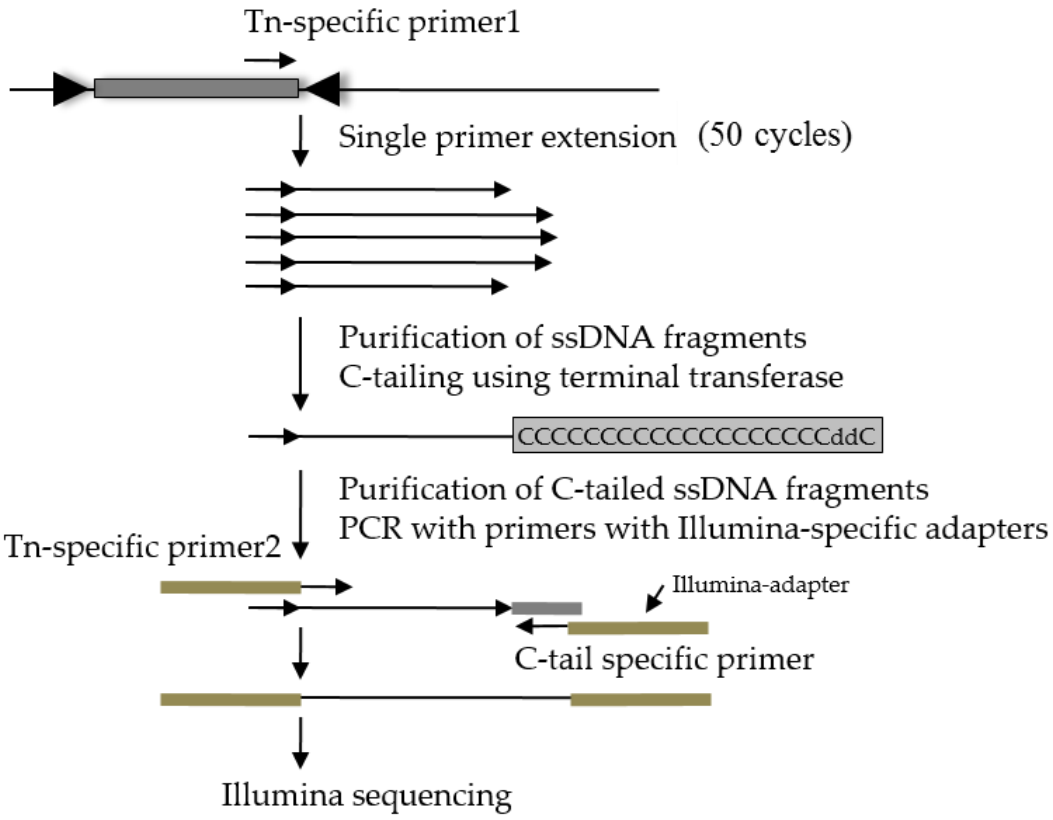
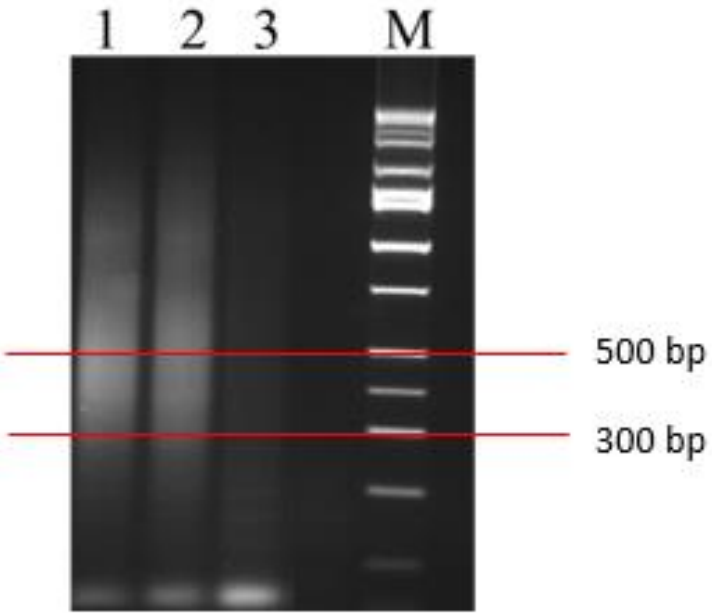


Figure S6: Agarose gel electrophoresis of Tn5 mutant library after exponential PCR. DNA from 300-500 bp is extracted from agarose gel (1%), pooled in equal quantity (10 ng/ sample) and sent for sequencing. [1,2: Tn5 library of *C. jejuni*; 3: Control (genomic DNA of C. jejuni NCTC 11168); and M: Hi-Low DNA Marker).



References

1. Kaakoush NO, Castano-Rodriguez N, Mitchell HM, Man SM: Global Epidemiology of Campylobacter Infection. *Clin Microbiol Rev* 2015, 28(3)**;**687-720.

2. Epps SV, Harvey RB, Hume ME, Phillips TD, Anderson RC, Nisbet DJ: Foodborne Campylobacter: infections, metabolism, pathogenesis and reservoirs. *International journal of environmental research and public health* 2013, 10(12)**;**6292-6304.

3. Juhas M, Eberl L, Church GM: Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol* 2012, 30(11)**;**601-607.

4. Xu P, Ge X, Chen L, Wang X, Dou Y, Xu JZ, Patel JR, Stone V, Evans K, Kitten T: Genome-wide essential gene identification in Streptococcus sanguinis. *Scientific reports* 2011, 1**;**125.

5. de Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A, Cruaud C, Samair S, Lechaplais C, Gyapay G, Richez C, Durot M, Kreimeyer A, Le Fevre F, Schachter V, Pezo V, Doring V, Scarpelli C, Medigue C, Cohen GN, Marliere P, Salanoubat M, Weissenbach J: A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. *Mol Syst Biol* 2008, 4**;**174.

6. Porwollik S, Santiviago CA, Cheng P, Long F, Desai P, Fredlund J, Srikumar S, Silva CA, Chu W, Chen X: Defined single-gene and multi-gene deletion mutant collections in Salmonella enterica sv Typhimurium. *PloS one* 2014, 9(7)**;**e99820.

7. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2006, 2**;**2006.0008.

8. Pechter KB, Gallagher L, Pyles H, Manoil CS, Harwood CS: Essential Genome of the Metabolically Versatile Alphaproteobacterium Rhodopseudomonas palustris. *J Bacteriol* 2015, 198(5)**;**867-876.

9. Wong Y, El Ghany MA, Naeem R, Lee K, Tan Y, Pain A, Nathan S: Candidate essential genes in Burkholderia cenocepacia J2315 identified by genome-wide TraDIS. *Frontiers in Microbiology* 2016, 7**;**.

10. Ji Y, Zhang B, Van SF, Horn, Warren P, Woodnutt G, Burnham MK, Rosenberg M: Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* 2001, 293(5538)**;**2266-2269.

11. Juhas M, Eberl L, Glass JI: Essence of life: essential genes of minimal genomes. *Trends Cell Biol* 2011, 21(10)**;**562-568.

12. Mushegian AR, Koonin EV: A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc Natl Acad Sci U S A* 1996, 93(19)**;**10268-10273.

13. Deng J: An Integrated Machine-Learning Model to Predict Prokaryotic Essential Genes. *Gene Essentiality: Methods and Protocols* 2015, **;**137-151.

14. Perumal D, Samal A, Sakharkar KR, Sakharkar MK: Targeting multiple targets in Pseudomonas aeruginosa PAO1 using flux balance analysis of a reconstructed genome-scale metabolic network. *J Drug Target* 2011, 19(1)**;**1-13.

15. Wei W, Ning L, Ye Y, Guo F: Geptop: a gene essentiality prediction tool for sequenced bacterial genomes based on orthology and phylogeny. *PloS one* 2013, 8(8)**;**e72343.

16. Long JL: Gene Essentiality: Methods and Protocols: Humana Press; 2015.

17. Kwon YM, Ricke SC, Mandal RK: Transposon sequencing: methods and expanding applications. *Appl Microbiol Biotechnol* 2016, 100(1)**;**31-43.

18. Goodman AL, McNulty NP, Zhao Y, Leip D, Mitra RD, Lozupone CA, Knight R, Gordon JI: Identifying genetic determinants needed to establish a human gut symbiont in its habitat. *Cell host & microbe* 2009, 6(3)**;**279-289.

19. Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J, Parkhill J, Turner AK: Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Res* 2009, 19(12)**;**2308-2316.

20. Gawronski JD, Wong SM, Giannoukos G, Ward DV, Akerley BJ: Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for Haemophilus genes required in the lung. *Proc Natl Acad Sci U S A* 2009, 106(38)**;**16422-16427.

21. Gallagher LA, Shendure J, Manoil C: Genome-scale identification of resistance functions in Pseudomonas aeruginosa using Tn-seq. *MBio* 2011, 2(1)**;**e00315-10.

22. van Opijnen T, Bodi KL, Camilli A: Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature methods* 2009, 6(10)**;**767-772.

23. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, Blow MJ, Bristow J, Butland G, Arkin AP, Deutschbauer A: Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *MBio* 2015, 6(3)**;**e00306-15.

24. Hutchison CA, Chuang R, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L: Design and synthesis of a minimal bacterial genome. *Science* 2016, 351(6280)**;**aad6253.

25. Stahl M, Stintzi A: Identification of essential genes in C. jejuni genome highlights hyper-variable plasticity regions. *Functional & integrative genomics* 2011, 11(2)**;**241-257.

26. Metris A, Reuter M, Gaskin DJ, Baranyi J, van Vliet AH: In vivo and in silico determination of essential genes of Campylobacter jejuni. *BMC Genomics* 2011, 12(1)**;**1.

27. Gao B, Lara-Tejero M, Lefebre M, Goodman AL, Galan JE: Novel components of the flagellar system in epsilonproteobacteria. *MBio* 2014, 5(3)**;**e01349-14.

28. Davis L, Young K, DiRita V: Genetic manipulation of Campylobacter jejuni. *Current protocols in microbiology* 2008, **;**8A. 2.1-8A. 2.17.

29. Dawoud TM, Jiang T, Mandal RK, Ricke SC, Kwon YM: Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers. *Mol Biotechnol* 2014, 56(11)**;**1004-1010.

30. Chun JY, Kim KJ, Hwang IT, Kim YJ, Lee DH, Lee IK, Kim JK: Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene. *Nucleic Acids Res* 2007, 35(6)**;**e40.

31. Langmead B, Salzberg SL: Fast gapped-read alignment with Bowtie 2. *Nature methods* 2012, 9(4)**;**357-359.

32. Pritchard JR, Chao MC, Abel S, Davis BM, Baranowski C, Zhang YJ, Rubin EJ, Waldor MK: ARTIST: high-resolution genome-wide assessment of fitness using transposon-insertion sequencing. *PLoS Genet* 2014, 10(11)**;**e1004782.

33. Solaimanpour S, Sarmiento F, Mrázek J: Tn-seq explorer: a tool for analysis of high-throughput sequencing data of transposon mutant libraries. *PloS one* 2015, 10(5)**;**e0126070.

34. Poly F, Threadgill D, Stintzi A: Genomic diversity in Campylobacter jejuni: identification of C. jejuni 81-176-specific genes. *J Clin Microbiol* 2005, 43(5)**;**2330-2338.

35. Lin J, Wang Y, Van Hoang K: Systematic identification of genetic loci required for polymyxin resistance in Campylobacter jejuni using an efficient in vivo transposon mutagenesis system. *Foodborne pathogens and disease* 2009, 6(2)**;**173-185.

36. Reid AN, Pandey R, Palyada K, Whitworth L, Doukhanine E, Stintzi A: Identification of Campylobacter jejuni genes contributing to acid adaptation by transcriptional profiling and genome-wide mutagenesis. *Appl Environ Microbiol* 2008, 74(5)**;**1598-1612.

37. Hendrixson DR, Akerley BJ, DiRita VJ: Transposon mutagenesis of Campylobacter jejuni identifies a bipartite energy taxis system required for motility. *Mol Microbiol* 2001, 40(1)**;**214-224.

38. Colegio OR, Griffin TJ,4th, Grindley ND, Galan JE: In vitro transposition system for efficient generation of random mutants of Campylobacter jejuni. *J Bacteriol* 2001, 183(7)**;**2384-2388.

39. Parkhill J, Wren B, Mungall K, Ketley J, Churcher C, Basham D, Chillingworth T, Davies R, Feltwell T, Holroyd S: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. *Nature* 2000, 403(6770)**;**665-668.

40. Green B, Bouchier C, Fairhead C, Craig NL, Cormack BP: Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA* 2012, 3(1)**;**1.

41. Han F, Pu S, Wang F, Meng J, Ge B: Fitness cost of macrolide resistance in Campylobacter jejuni. *Int J Antimicrob Agents* 2009, 34(5)**;**462-466.

42. Dryselius R, Izutsu K, Honda T, Iida T: Differential replication dynamics for large and small Vibrio chromosomes affect gene dosage, expression and location. *BMC Genomics* 2008, 9(1)**;**1.

43. Klein BA, Tenorio EL, Lazinski DW, Camilli A, Duncan MJ, Hu LT: Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. *BMC Genomics* 2012, 13**;**578-2164-13-578.

44. Rosconi F, de Vries SP, Baig A, Fabiano E, Grant AJ: Essential genes for in vitro growth of the endophyte Herbaspirillum seropedicae SmR1 revealed by transposon insertion sequencing. *Appl Environ Microbiol* 2016, **;**.

45. Chao MC, Pritchard JR, Zhang YJ, Rubin EJ, Livny J, Davis BM, Waldor MK: High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Res* 2013, 41(19)**;**9033-9048.

46. Gladki A, Kaczanowski S, Szczesny P, Zielenkiewicz P: The evolutionary rate of antibacterial drug targets. *BMC Bioinformatics* 2013, 14(1)**;**1.

47. Chen AY, Fry SR, Forbes-Faulkner J, Daggard GE, Mukkur T: Comparative immunogenicity of M. hyopneumoniae NrdF encoded in different expression systems delivered orally via attenuated S. typhimurium aroA in mice. *Vet Microbiol* 2006, 114(3)**;**252-259.

48. Rosch JW, Iverson AR, Humann J, Mann B, Gao G, Vogel P, Mina M, Murrah KA, Perez AC, Edward Swords W, Tuomanen EI, McCullers JA: A live-attenuated pneumococcal vaccine elicits CD4+ T-cell dependent class switching and provides serotype independent protection against acute otitis media. *EMBO Mol Med* 2014, 6(1)**;**141-154.

49. Capodagli GC, Sedhom WG, Jackson M, Ahrendt KA, Pegan SD: A noncompetitive inhibitor for Mycobacterium tuberculosis's class IIa fructose 1, 6-bisphosphate aldolase. *Biochemistry (N Y)* 2013, 53(1)**;**202-213.

50. Lin L, Tan B, Pantapalangkoor P, Ho T, Baquir B, Tomaras A, Montgomery JI, Reilly U, Barbacci EG, Hujer K, Bonomo RA, Fernandez L, Hancock RE, Adams MD, French SW, Buslon VS, Spellberg B: Inhibition of LpxC protects mice from resistant Acinetobacter baumannii by modulating inflammation and enhancing phagocytosis. *MBio* 2012, 3(5)**;**10.1128/mBio.00312-12. Print 2012.

51. Ouellette SP, Karimova G, Subtil A, Ladant D: Chlamydia co-opts the rod shape-determining proteins MreB and Pbp2 for cell division. *Mol Microbiol* 2012, 85(1)**;**164-178.

52. Hassan SS, Tiwari S, Guimarães LC, Jamal SB, Folador E, Sharma NB, de Castro Soares S, Almeida S, Ali A, Islam A: Proteome scale comparative modeling for conserved drug and vaccine targets identification in Corynebacterium pseudotuberculosis. *BMC Genomics* 2014, 15(7)**;**1.

53. Phunpruch S, Warit S, Suksamran R, Billamas P, Jaitrong S, Palittapongarnpim P, Prammananan T: A role for 16S rRNA dimethyltransferase (ksgA) in intrinsic clarithromycin resistance in Mycobacterium tuberculosis. *Int J Antimicrob Agents* 2013, 41(6)**;**548-551.

54. Jayakar HR, Parvathareddy J, Fitzpatrick EA, Bina XR, Bina JE, Re F, Emery FD, Miller MA: A galU mutant of Francisella tularensis is attenuated for virulence in a murine pulmonary model of tularemia. *BMC Microbiol* 2011, 11**;**179-2180-11-179.

55. Fu S, Zhang M, Ou J, Liu H, Tan C, Liu J, Chen H, Bei W: Construction and immune effect of Haemophilus parasuis DNA vaccine encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in mice. *Vaccine* 2012, 30(48)**;**6839-6844.

CONCLUSION

In this dissertation, we have applied high throughput transposon sequencing (Tn-seq) for systems-level analysis of foodborne pathogens that can screen thousands of transposon mutants in a single experiment accompanied by massively parallel next-generation sequencing (NGS). Foodborne pathogens *Salmonella* and *Campylobacter* represents a leading cause of foodborne pathogen illness worldwide. Increased understanding of the genetic determinants of foodborne pathogens can help formulate the effective strategies to mitigate these nasty bugs from the food web without altering the consumer preference. In the first part, we reviewed current methods and applications of Tn-seq (Chapter 1), In the second part, *Salmonella* Typhimurium 14028S (*S.* Typhimurium) complex Tn5 mutant library was subjected to growth and survival fitness assay (Chapter 2 and 3). In the third part, we investigated the essential genome of two stains of *Campylobacter jejuni* (*C. jejuni*) NCTC 11168 and *C. jejuni* 81-176 (Chapter 4).

We have improved the previously developed protocol in our laboratory of DNA library preparation for Illumina sequencing using regular and dual priming oligonucleotide (DPO) primer. Additionally, comparative bioinformatics studies were heavily employed in our studies to pinpoint and narrow down the most indispensable genetic factors required for the fitness of *Salmonella* and *Campylobacter*. Furthermore, we constructed an unprecedented level of highly complex Tn5 mutant library in *C. jejuni* NCTC 11168 and uniquely used the library DNA for the reconstruction of Tn5 library in the same (*C. jejuni* NCTC 11168) and different strain background (*C. jejuni* 81-176).

In the Chapter 2, a highly saturating Tn5 library of *S.* Typhimurium 14028s (≥ 186, 000 unique insertions) was subjected to selection during growth in the presence of short chain fatty acid (100

mM Propionate), osmotic stress (3% NaCl) or oxidative stress (1mM H2O2) or survival in extreme acidic pH (30 min in pH 3) or starvation (12 days in PBS). A comparative study with *in vivo* (previously study) and *in vitro* genes (our study), we identified 25 set of other than virulence genes (non-pathogenicity island genes) that can potentially invoke host immune response with has great prospects to be exploited as vaccine development and/or drug target to curb the *Salmonella* infection.

In chapter 3, we performed global screening of *S*. Typhimurium 14028S Tn5 mutant library in abiotic stress such as desiccation stress that *Salmonella* faces outside the body of host. *Salmonella* desiccation resistance genes were mostly enriched for energy production and conversion; cell wall/membrane/envelope biogenesis, inorganic ion transport and metabolism; regulation of biological process; DNA metabolic process; ABC transporters; and two component system with more than 20% genes being putative or hypothetical.

In chapter, 4 we constructed an unprecedented level of highly complex Tn5 mutant library in *C. jejuni* NCTC 11168 with 95,929 unique insertions and reconstructed Tn5 to same and different strain background. Essential genes in both of the strains were highly enriched in cluster of orthologous group (COG) category like 'Translation, ribosomal structure and biogenesis (J)', 'Energy production and conversion (C)', 'Coenzyme transport and metabolism (H)' and not in COG category. Comparative analysis among our and previous studies identified 50 core essential genes of *C. jejuni*, where most genes have been implicated in the development of drug target and vaccines against a wide range of bacteria.

In short, comprehensive analysis of *in vitro* fitness genes of *Salmonella* and *Campylobacter* provided a solid background research for the development of more effective strategies to alleviate the foodborne pathogens.

Appendix I

CURRICULUM VITAE

Rabindra Kumar Mandal, B.V.Sc. & A.H.
Ph.D. Candidate

INTERESTS

Job Interests: Research; teaching; working in collaboration; academic settings

Areas of interest: molecular microbiology; microbial ecology; immunology; bacterial pathogenesis; transposon sequencing; next-generation sequencing; antimicrobial resistance; host-pathogen interaction; probiotics; and bioinformatics

---

EDUCATION

Ph.D. Candidate, Poultry Science, University of Arkansas, Fayetteville, Arkansas, 72701
May 2012 – Present (Expected graduation- *Fall 2016*)
GPA-3.76
Thesis Title: Genetic Determinants of *Salmonella* and *Campylobacter* required for *in vitro* fitness
Major Professor: Dr. Young Min Kwon

B.V.Sc. & A.H., Veterinary Science, Institute of Agriculture and Animal Science, Chitwan, Nepal *August 2005 - August 2010*
GPA-3.99
Thesis Title: Common Elephant Diseases Recorded in the Captive Elephants of Chitwan National Park, Nepal.
Major Advisor: Dr. Kamal Prasad Gairhe

RESEARCH AND PROFESSIONAL EXPERIENCE

Graduate Research Assistant, Department of Poultry Science
P.I. Young Min Kwon, University of Arkansas, Fayetteville, AR *May 2012 – Present*

- ❖ Planned and conducted Tn-seq (transposon sequencing) experiments for functional genomics analysis of *Salmonella* and *Campylobacter jejuni*
- ❖ Worked on chicken microbiome analysis based on Illumina sequencing of 16S rRNA gene
- ❖ Established bioinformatics pipeline for analysis of Illumina sequencing data for microbiome analysis using Pearl scripts, QIIME, LEfSe, PICRUSt, and Cytoscape
- ❖ Wrote peer reviewed scientific papers
- ❖ Taught 4 fellow graduate students to use bioinformatics pipelines like QIIME, ARTIST

- ❖ Made PowerPoint presentations and scientific posters
- ❖ Revised and wrote lab protocols for *Salmonella* mutant construction
- ❖ Collaborated with graduate students and post-docs on research projects
- ❖ Constructed candidate *Salmonella* vaccines
- ❖ Collected samples (blood) from chicken
- ❖ Conducted extensive lab work such as bacterial mutant construction, phenotype testing, PCR, bacterial culture
- ❖ Collected and maintained bacterial strains
- ❖ Made database of 16S rRNA to better identify bacteria at species level compatible to QIIME
- ❖ Took training on RNA sequencing (RNA-seq)

Adjunct Assistant Lecturer, Himalayan College of Agricultural Science and Technology (HICAST), Kathmandu, Nepal *February 2011 – March 2012*

- ❖ Taught a course on Veterinary Public Health (Milk Hygiene)
- ❖ Helped to conduct lab experiments
- ❖ Ran a 50 min theory class and 2-hour lab class
- ❖ Helped with grading exams and quizzes

Techno-Marketing Officer, Adhunik Poultry Breeding Farm Pvt. Ltd. Kathmandu, Nepal *November 2011 – February 2012*

- ❖ Provided veterinary service to poultry farmers
- ❖ Helped in diagnosis, control and treatment of disease in poultry farms
- ❖ Marketed the farm products like feed and chicks
- ❖ Taught farmers on better poultry practices

Internship, Chitwan National Park, Chitwan, Nepal *February 2010 – July 2010*

- ❖ Actively participated in the medical treatment, routine management of captive elephants
- ❖ Involved in rescue of orphan wild animals
- ❖ Participated in rhinoceros patrolling, camera trapping of tigers, and darting of tigers
- ❖ Conducted postmortem of one-horned rhinoceros and gharial
- ❖ Participated in care and management of vultures and gharial at the breeding center

TECHNICAL EXPERTISE

- Design and conduct research on functional genomics of bacteria using transposon sequencing
- Design, collect, conduct, analyze and interpret results of chicken microbiomes
- Next generation sequencing
- RNA sequencing (RNA-seq)
- Bacteria culture and counting
- *Salmonella* and *Campylobacter* genome manipulation
- *Lactobacillus* motility, aggregation, and pathogen inhibition (agar overlay assay)

- Mutant construction for *Salmonella*-vectored vaccine
- Mutant construction for *Bacillus* vectored vaccine
- DNA Isolation
- PCR
- ELISA
- Flow Cytometry
- Mass sepctrometry
- Growth phenotyping (TECAN)
- Agarose gel electrophoresis
- Gel imaging
- Polyacrylamide gel electrophoresis
- Cluster computing
- Bash scripting
- Perl programming
- Python (Beginner level)
- Bioinformatics analysis (QIIME, ARTIST, String, BLAST, Panther among others)
- SAS
- JMP
- R

GRANTS & FELLOWSHIP

Conservation and Sustainable Use of Wetlands in Nepal (CSUWN) Research Grant, Kathmandu, Nepal *2011*
Funded: $350 towards research costs

Himalayan Nature Research Grant, Kathmandu, Nepal *2011*
Funded: $200 towards research costs

Community Livestock Development Program (CLDP) Research Grant, Institute of Agriculture and Animal Science, Chitwan, Nepal *2010*
Funded: $150 towards research costs

Government Scholarship for Undergraduate Study, Institute of Agriculture and Animal Science, Rampur, Chitwan, Nepal *2005 – 2010*
Full scholarship to pursue B.V.Sc & A.H.

AWARDS & HONORS

Midwest Big Data Summer School, **Ames**, **IA** *June 20-24 2016*
Recipient of Travel Grant

Bumper college Honor Student Board Poster competition, University of Arkansas, Fayetteville, Arkansas, 72701
First Place Graduate Poster *April 2016*

Outstanding Graduate Student Award, Department of Poultry Science, University of Arkansas, Fayetteville, Arkansas, 72701
Nominated *November 2015*
Graduate Student Poster Presentation Competition, Fayetteville, Arkansas, 72701 Arkansas Association of Food Protection (AAFP)
First place in section "Fundamental understandings of Foodborne Pathogens". *September 2015*

PUBLICATIONS

In preparation:

(1) **Mandal, Rabindra K**, Tieshan Jiang and Young Min Kwon. "Genetic mechanisms of *Salmonella* enterica serovar Typhimurium for overcoming host stressors".

(2) **Mandal, Rabindra K**, Tieshan Jiang, Robert F. Wideman Jr., Troy Lohrmann, and Young Min Kwon. "Molecular survey of the microbiomes in broilers to understand probiotics- induced reduction of BCO lameness".

(3) Yichao Yang, Amanda Wolfenden, **Rabindra K. Mandal**, Olivia Faulkner, Billy Hargis, Yong Min Kwon, Lisa Bielke. Evaluation of Recombinant *Salmonella* Vaccines to Provide Cross-Serovar Protection and Prevent *Campylobacter* Infection.

(4) **Mandal, Rabindra K**., Tieshan Jiang, Robert F. Wideman Jr., Troy, Lohrmann, and Young Min Kwon. "Microbiome analysis to understand translocation of gut microbiota into blood in chickens.

(5) **Mandal, Rabindra K** and Young Min Kwon. "Global screening of *Salmonella* genes for desiccation survival".

(6) **Mandal, Rabindra K**, Tieshan Jiang and Young Min Kwon. "Essential genome of *Campylobacter jejuni*".

Published:

(5) **Mandal, Rabindra K.**, Tieshan Jiang, Adnan A. Al-Rubaye, Douglas D. Rhoads, Robert F. Wideman, Jiangchao Zhao, Igal Pevzner, and Young Min Kwon. "An investigation into blood microbiota and its potential association with Bacterial Chondronecrosis with Osteomyelitis (BCO) in Broilers." ***Scientific Reports*** 6 (2016): 25882.

(6) Kwon, Young Min, Steven C. Ricke, and **Rabindra K. Mandal**. "Transposon sequencing: methods and expanding applications." ***Applied microbiology and biotechnology*** 100, no. 1 (2016): 31-43.

(7) Jiang, Tieshan, **Rabindra K. Mandal**, Robert F. Wideman Jr, Anita Khatiwara, Igal Pevzner, and Young Min Kwon. "Molecular survey of bacterial communities associated with bacterial chondronecrosis with osteomyelitis (BCO) in broilers." ***PloS one*** 10, no. 4 (2015): e0124403.

(8) Dawoud, Turki M., Tieshan Jiang, **Rabindra K. Mandal**, Steven C. Ricke, and Young Min Kwon. "Improving the efficiency of transposon mutagenesis in Salmonella enteritidis by overcoming host-restriction barriers." ***Molecular biotechnology*** 56, no. 11 (2014): 1004-1010.

ABSTRACTS AND CONFERENCE PROCEEDINGS

(1) ***Rabindra K. Mandal**, Audrianna Rogers, Nicholas B. Anthony and Young Min Kwon. Exploring the effect of host genotype and environment on gut microbiomes in broiler chickens.
-Oral presentation at the 2016 Poultry Science Association (PSA), New Orleans, Louisiana July 11–14, 2016.

(2) Bishnu Adhikari, ***Rabindra K. Mandal**, and Young Min Kwon. Characterization of lactic acid bacteria population associated with different regions in gastrointestinal tract of chicken.
-Poster presentation at the 2016 PSA, New Orleans, Louisiana July 11–14, 2016.

(3) Anita Mandal, **Rabindra K. Mandal**, Young Min Kwon. In vitro characterization of *Lactobacillus* isolates from chicken gut for probiotics potentials.
-Accepted for poster presentation at the 2016 Poultry Science Association (PSA), New Orleans, Louisiana July 11–14, 2016.

(4) A.D. Wolfenden, Y. Yang, O.B. Faulkner, **R.K. Mandal**, B.M. Hargis, L.R. Berghman, Y.M. Kwon, L.R. Bielke. Evaluation of Recombinant *Salmonella* Typhimurium and Enteritidis Vaccines to Protect Against *Salmonella* Heidelberg Infection.
-Oral presentation at the 2016 Poultry Science Association (PSA), New Orleans, Louisiana July 11–14, 2016.

(5) ***Rabindra K. Mandal**, Tieshan Jiang, Robert F. Wideman Jr., Troy Lohrmann, and Young Min Kwon. Tracking down the source of microbiome in sterile organs of chickens.
-Poster presentation at the 2016 Arkansas Academy of Science (AAS), University of Arkansas, Fayetteville, April 01-02 2016.

(6) ***Rabindra K. Mandal**, and Young Min Kwon. Genetic mechanisms of Salmonella enterica serovar Typhimurium for overcoming host stressors.
-Oral Presentation at the 2015 Conference of Research Workers in Animal Disease (CRWAD), Chicago, December 6-8, 2015.

(7) ***Rabindra K. Mandal**, Tieshan Jiang, Adnan A. Alrubaye, Douglas D. Rhoads, Robert F. Wideman Jr., Jiangchao Zhao, Igal Pevzner and Young Min Kwon. Bacteremia and

Lameness: An investigation into blood microbiota associated with Bacterial Chondronecrosis with Osteomyelitis (BCO) in Broilers.

-Oral presentation at the 2015 Symposium on Gut Health in Production of Food Animals, Kansas, November 9-11, 2015.

(8) ***Rabindra K. Mandal**, Tieshan Jiang, and Young Min Kwon. Genome-wide genetic requirements of Salmonella Typhimurium for optimal growth in the presence of propionate.
-Poster presentation at the 2015 American Society of Microbiology, New Orleans, LA, USA, May 30 - June 2, 2015.
-Poster presentation at the 2015 Arkansas Association for Food Protection, Fayetteville, Arkansas, September 8-10, 2015.

(9) Jiang, Tieshan, **Rabindra K. Mandal**, Robert Wideman Jr, T. Lohrman and Young Min Kwon. Molecular survey of the microbiomes in broilers to understand probiotics-induced reduction of BCO lameness.
- Oral presentation at the 2015 Symposium on Gut Health in production of Food Animals. Kansas City, MO, November 10-12, 2014.

* Presenting author