A QUASIMETAGENOMICS METHOD FOR CONCERTED DETECTION AND SUBTYPING OF *SALMONELLA ENTERICA* AND *E. COLI* O157:H7 FROM FRESH PRODUCE

by

ANNA M. TOWNSEND

(Under the Direction of Xiangyu Deng)

ABSTRACT

Quasimetagenomics refers to the sequencing of a modified food microbiome to facilitate combined detection and strain-level subtyping of targeted pathogens in a single workflow. Through quasimetagenomic sequencing, pathogens are detected and identified in a shortened time frame compared to traditional culture enrichment and whole genome sequencing-based analyses. While this method was previously used to detect and subtype *Salmonella enterica* from chicken, iceberg lettuce, and black pepper, it has not been applied to investigate several pathogens in one workflow. A quasimetagenomic method to concertedly detect and subtype *Salmonella enterica* and *Escherichia coli* O157:H7 from artificially contaminated fresh produce in a single workflow was developed. The majority of quasimetagenomic samples were serotyped after 12 hours of co-enrichment in a nonselective medium. SNP typing was achievable for some initial pathogen inoculum levels as low as 0.1 CFU/g, suggesting that this method can be used for concerted detection and identification of bacterial pathogens.

INDEX WORDS: quasimetagenomics, fresh produce, detection, subtyping, foodborne illness

A QUASIMETAGENOMICS METHOD FOR CONCERTED DETECTION AND SUBTYPING OF *SALMONELLA ENTERICA* AND *E. COLI* O157:H7 FROM FRESH PRODUCE

by

ANNA M. TOWNSEND

B.A., The University of Kentucky, 2017

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

MASTER OF SCIENCE

ATHENS, GEORGIA

2019

A QUASIMETAGENOMICS METHOD FOR CONCERTED DETECTION AND SUBTYPING OF *SALMONELLA ENTERICA* AND *E. COLI* O157:H7 FROM FRESH PRODUCE

by

ANNA M. TOWNSEND

Major Professor:	Xiangyu Deng
Committee:	Mark A. Harrison
	Hendrik C. den Bakker

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2019

ACKNOWLEDGMENTS

I would like to thank Dr. Xiangyu Deng for supporting my educational and professional career over the past two years. I would also like to acknowledge my committee, Dr. Mark Harrison and Dr. Henk den Bakker, for their time and guidance. I would like to especially acknowledge David Mann, who completed library preparation and sequencing on the Illumina platform for my samples and provided additional project guidance.

Thank you to Dr. Peter Cook, Dr. Fereidoun Forghani, Dr. Lee Katz, Shaoting Li, and Yingshu He for providing laboratory and bioinformatic support.

Lastly, thank you to my family and friends for sustaining my focus and motivation over the course of my degree.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

With the increased globalization of food systems and access to fresh produce, foodborne illness outbreak reports have also increased over the last twenty years (Olaimat and Holley, 2012). Leafy greens, such as lettuces, are susceptible carriers of foodborne bacteria due to an absence of complete inactivation methods or kill steps during their production (Gil et al., 2015). Leafy greens can contain a variety of microbiota that is able to attach and form biofilms along the plant surface (Gil et al., 2015). It is possible for non-pathogenic and pathogenic microbiota to be present on the surface of leafy greens. Pathogenic bacterial species, such as *Salmonella* species and *Escherichia coli* O157:H7, have been closely associated with lettuce (Ramos et al., 2013). Between 1973 and 2012, there were over 600 foodborne illness outbreaks associated with leafy green vegetables (Herman et al., 2015). Of these illnesses, at least 18% were caused by Shiga toxin-producing *E. coli* (STEC) and 11% from *Salmonella* species (Herman et al., 2015). These outbreaks have been responsible for more than 20,000 illnesses and 20 deaths (Herman et al., 2015). Notably, in 2018, there were two major foodborne illness outbreaks associated with romaine lettuce. The first outbreak began in March 2018, originated in the Yuma Valley region of Arizona, and caused 210 illnesses and five deaths from *E. coli* O157:H7 (Centers for Disease Control, 2018b). The second outbreak was traced to romaine grown in the Central Coastal regions of California and led to 62 illnesses from *E. coli* O157:H7 (Centers for Disease Control, 2018c).

In response to increased foodborne illness reports attributed to fresh produce, detection methods paired with subtyping by whole-genome sequencing have demonstrated more robust and

effective outbreak surveillance and response (Allard et al., 2016). Methods such as polymerase chain reaction (PCR) and whole genome sequencing have dramatically decreased the time taken to detect and subtype a bacterial pathogen (Law et al., 2014; Zhao et al., 2014). Detection and subtyping of foodborne pathogens are traditionally separate processes (Hyeon et al., 2018). However, quasimetagenomic sequencing combines these processes, allowing a rapid turnaround from contaminated food to pathogen fingerprints (Hyeon et al., 2018). The purpose of this research is to concertedly detect and subtype bacterial pathogens in fresh produce using a quasimetagenomic approach. These techniques have not been used previously for concerted detection and subtyping of multiple pathogens from food samples.

CHAPTER 2

LITERATURE REVIEW

Lettuce and leafy greens

The U.S. Food and Drug Administration (FDA) recognizes iceberg lettuce, romaine lettuce, leaf lettuce, butter lettuce, baby (immature) lettuce or greens, escarole, endive, spring mix, spinach, cabbage, kale, arugula, and chard as leafy greens (U.S. Food and Drug Administration, 2009). Leafy greens are becoming increasingly linked to global foodborne illness outbreaks. In the U.S. during the 1970s, there were fewer than 20 reported produce-related outbreaks; while in the 1990s, there were over 100 (Danyluk and Schaffner, 2011). The FDA reported that 34.1% of produce-related outbreaks between 1996 and 2008 were attributed to leafy greens (U.S. Food and Drug Administration, 2009). While, superficially, it appears outbreaks have increased in prevalence, greater collaboration between public health agencies and development of faster detection and robust subtyping methods have increased report of and response to foodborne diseases.

Approximately 30% of the lettuce consumed in the United States is romaine lettuce (Erickson and Ortega, 2018). Within the United States, romaine is primarily grown in California and Arizona (Erickson and Ortega, 2018). Many widely-consumed foods, such as Caesar salad and sandwiches, contain romaine lettuce. Romaine is oblong, with rib-like leaves growing in an upward direction from a central core or heart. In head and romaine lettuce varieties, leaves overlap each other forming pockets that are potentially able to harbor microorganisms. Leafy green production is typically divided into five stages: (1) production and harvest, (2) postharvest, (3) fresh-cut/value-added, (4) distribution, and (5) consumer, retail, and foodservice (U.S. Food and

Drug Administration, 2009). Most lettuces are grown outdoors in uncovered fields and harvested by hand (Kerns, 1999). Growth from seed to mature leaf lettuce takes between 65 and 80 days in summer or up to 130 days in fall and winter (Smith et al., 2011). Ideal temperatures for lettuce growth are between 63° and 83°F (Smith et al., 2011). Lettuce maturation in higher temperatures typically leads to bitter, tough greens due to bolting, which is when a plant goes to seed (Smith et al., 2011). To avoid quality defects, lettuce production within the United States takes place in different regions based on seasonality. Generally, production location transitions from central California to the Yuma region of Arizona in order to harvest year-round (Smith et al., 2011). Another method of preserving quality is field packaging, which is when the lettuce is harvested and immediately packaged (Kerns, 1999). Lettuce can be naked packed (no wrapping), film wrapped in cellophane or plastic (perforated or non-perforated), or bulked (multiple heads per bag) prior to shipment (Kerns, 1999). Wrapping lettuce helps prevent damage from pests and reduces contamination (Kerns, 1999).

Many foodborne illnesses linked to leafy greens have been traced to the production and harvest stage or at the final consumer, retail, and food service operation (Herman et al., 2015). Cross contamination with animal and human sources (such as feces), contaminated irrigation water, and unclean gloves or tools, can introduce pathogens to leafy greens in the field or during hand-harvest. For example, the mid-2018 *E. coli* O157:H7 outbreak associated with romaine lettuce was traced to irrigation water (Centers for Disease Control, 2018b). The FDA completed an environmental assessment in the Yuma Valley and found the outbreak strain in irrigation canal water (U.S. Food and Drug Administration (FDA), 2018). There is a concentrated large animal feeding operation near the irrigation canal. Large animals used in agricultural production, such as cattle, share a commensal relationship with *E. coli* O157:H7 and many other bacterial species.

Therefore, it is likely the canal water was contaminated with cattle feces harboring *E. coli* O157:H7 and the contaminated water was used on romaine lettuce crops. In addition to improper agricultural practices during production and harvest, incorrect handling, storage conditions, or hygienic practices can also lead to contamination of leafy greens during transportation and at the consumer level.

Bacteria of concern

The Centers for Disease Control and Prevention (CDC) estimate over 48 million cases of foodborne illness occur annually in the United States (Centers for Disease Control, 2016b). Of these 48 million cases, approximately 1.4 million are due to non-typhoidal *Salmonella* species and approximately 73,000 are due to *E. coli* O157:H7 (Mead et al., 1999). Both of these bacteria have been associated with leafy-green-related food foodborne illness.

Salmonella enterica

Salmonella bacteria are facultatively anaerobic, gram-negative bacteria that are rod-shaped and motile (Andino and Hanning, 2015). The infectious dose of *Salmonella* can be as low as one cell (U.S. Food and Drug Administration, 2012). *Salmonella enterica* is able to cause illness if it bypasses the gastric barrier, enabling irritation and colonization of the intestines. This leads to gastroenteritis characterized by diarrhea, nausea, or vomiting (Andino and Hanning, 2015). If the infection is severe, *S. enterica* can also cause bacteremia, which is defined as bacteria in the bloodstream (Andino and Hanning, 2015). Foodborne disease from *Salmonella* is traditionally associated with eggs, poultry, unpasteurized milk and juice, and raw fruits and vegetables (U.S. Food and Drug Administration, 2018a). Since 2004, there have been 17 confirmed *Salmonella*-caused illness outbreaks associated with leafy greens (Marler, 2018). In 2004, there were 97 cases of *Salmonella* Newport infections resulting from contaminated iceberg lettuce (Marler, 2018). In

2006, 2009, and 2011, there were four separate lettuce-associated foodborne illness outbreaks of *Salmonella* Typhimurium (Marler, 2018). *Salmonella* Saintpaul has been associated with three produce-related foodborne disease outbreaks from contaminated tomatoes and jalapeno and serrano peppers (2008), alfalfa sprouts (2009), and cucumbers (2013) (Centers for Disease Control, 2009; Klontz et al., 2010; Yao et al., 2017).

Escherichia coli O157

E. coli O157 is a facultatively anaerobic, gram-negative bacterial serovar that is rod-shaped, motile, and a type of enterohemorrhagic *E. coli* (EHEC) (Lim et al., 2010). This bacterium causes hemorrhagic colitis, or bloody diarrhea, via Shiga-like toxins that destroy the gastrointestinal epithelium. If left untreated, the bacteria can cause hemolytic uremic syndrome (HUS), which is a life-threatening kidney disease. The reported infectious dose of *E. coli* O157:H7 is approximately 10 to 100 cells (U.S. Food and Drug Administration, 2012). Foodborne disease from *E. coli* is typically traced to fecal contamination in water or food, improperly cooked beef, unpasteurized milk and juice, and raw vegetables (U.S. Food and Drug Administration, 2018a). *Escherichia coli* bacteria have a long, prolific history with leafy greens, especially romaine lettuce. At least thirty-nine *E. coli* O157 illnesses were caused by or associated with leafy greens in the last 23 years (Marler, 2018). Of these illnesses, at least eight have been traced to romaine lettuce (Centers for Disease Control, 2018b, c; Marler, 2018).

Aside from an absence of a total lethality step in fresh produce production, one possible explanation for frequent disease from *Salmonella* and *E. coli* O157:H7 on leafy greens is bacterial adherence. *Salmonella enterica* serovars can readily attach to both intact and cut lettuces, with bacteria attaching more strongly to romaine lettuce compared to cabbage and iceberg lettuce (Patel and Sharma, 2010). After attachment to intact romaine lettuce leaves, *Salmonella enterica* has

been shown to maintain a consistent population size when stored for nine days at 4°C (Kroupitski et al., 2009). *Escherichia coli* O157:H7, from artificially contaminated irrigation water, has been shown to readily adhere to lettuce seedling roots and inside stomatal pores (Wachtel et al., 2002). *Escherichia coli* O157:H7 biofilms can decrease the efficacy of both irradiation and disinfectant solutions on romaine and spinach (Niemira and Cooke, 2010). In addition, damaged plant tissue can harbor and promote the growth of *E. coli* O157:H7 populations (Brandl, 2008). Both *Salmonella* and *E. coli* O157:H7 have curli, which enable plant adhesion, cell aggregation, and biofilm formation (Barnhart and Chapman, 2006; Seo and Matthews, 2012). *Salmonella enterica* growth was observed in salad leaf juices, which release after plant damage, with direct *Salmonella* attachment to leaves increasing over 350% when juice was present (Koukkidis et al., 2017). Therefore, these pathogens have biological adaptations that enable them to readily adhere to plant material, which may lead to increased disease associated with fresh produce, especially leafy greens. This also indicates that poor lettuce quality may be an indicator of potential bacterial residence. A more aggressive lethality step is needed to prevent the attachment of bacteria and potential biofilm formation on fresh produce without impacting quality.

Pathogen concentration and detection

The turnaround time from receiving contaminated food samples to determining subtypes could be up to 17 days for a *Salmonella*-associated outbreak and 12 days for an *E. coli* O157:H7 outbreak (Centers for Disease Control, 2014, 2019). After consistent reports of similar illnesses, an outbreak investigation may take months or years until the pathogen source is identified. Then, after the contaminated food matrix or item is identified, a recall may be needed if the product is still available to consumers. Therefore, more rapid, efficient methods for laboratory investigation of *Salmonella* and *E. coli* species are needed to shorten the time between receiving contaminated

food samples to subtyping the pathogen, and also prevent additional foodborne illnesses. Several technologies used to detect foodborne pathogens include culture enrichment, PCR, and enzyme-linked immunosorbent assays (ELISAs) (Law et al., 2014). These detection methods are used to determine if a pathogen is present in a food item. Previous research has demonstrated an original food microbiome can be modified for the concentration of *Salmonella enterica* genomic DNA through culture enrichment, immunomagnetic separation (IMS), and multiple displacement amplification (MDA) before sequencing (Hyeon et al., 2018). This process, termed as quasimetagenomics, refers to the sequencing of a modified food microbiome to facilitate detection and subtyping of targeted pathogens. Quasimetagenomics sequencing is partially reliant on metagenomics, which is the study of genetic material taken directly from the environment (Lin, 2006). Metagenomics is a burgeoning field within food safety and quality. This approach can characterize spoilage organisms without relying on selective media and provide a better understanding of microbiological ecosystems in a food matrix (Cocolin et al., 2018). In 2015, a metagenomic approach was used to analyze extracted DNA from artificially contaminated spinach with Shiga toxin-producing *E. coli* (STEC) after enrichment using a modified FDA Bacteriological Analytical Manual (BAM) protocol (Leonard et al., 2015). In addition to identifying spoilage and pathogenic bacteria, metagenomics has also been used to identify viruses on fresh produce (Aw et al., 2016). Aw et al. (2016) identified rotaviruses and picobirnaviruses on romaine and iceberg lettuces in field and retail samples. One of the pitfalls of metagenomics is the proportion of human, food, or microbial DNA compared to the target pathogen (Cocolin et al., 2018). Typically, pathogen DNA is in much smaller quantities compared to non-target DNA. Quasimetagenomics sequencing addresses this issue with integration of culture enrichment and IMS beads to concentrate the target pathogen prior to amplifying DNA.

Immunomagnetic Separation (IMS)

IMS-MDA serves as an optimization method to traditional culture enrichment (Hyeon et al., 2018). Culture enrichment, depending on the organism, can take up to 48 h and involve several types of microbiological media. For example, the FDA BAM protocol for *Salmonella* detection involves 48 hours of enrichment in three different media. IMS-MDA has successfully allowed real-time PCR or genome sequencing-based detection and/or subtyping of *S. enterica* with less than 12 h of enrichment (Hyeon and Deng, 2017; Hyeon et al., 2018). IMS uses magnetic beads, coated in antibodies, which are able to remove the organism of interest from a food matrix, environmental swab, or enrichment culture (Seth-Smith et al., 2013). The antibodies target and bind to certain bacterial pathogens, which are drawn out of the supernatant using a magnetic current between the magnetized beads and a magnetic rack. By utilizing species- or serotype-specific immunomagnetic beads for *Salmonella enterica* and *E. coli* O157:H7, this study demonstrates an expanded quasimetagenomic method to concertedly detect multiple bacterial pathogens in a single workflow.

Multiple Displacement Amplification

Similar to PCR, MDA also involves polymerases for strand displacement and replication (de Bourcy et al., 2014). While PCR amplifies a small, selective portion of the genome, MDA is able to replicate large portions of the genome (Chen et al., 2014). Polymerase phi29 is used to replicate DNA strands with the aid of random hexamers, which are short oligodeoxy-ribonucleotides that anneal to random sections of complementary DNA (Kumar, 2007). Shown in Figure 1, there are three main steps that are repeated twice in the MDA process: polymerization, debranching, and primer annealing. Polymerization refers to the act of synthesizing DNA via a DNA polymerase (phi29), debranching is the process of breaking down DNA, and primer

annealing is the attachment of a primer (random hexamers) to the shards of DNA after debranching (Lasken, 2007). While PCR uses thermal cycling to replicate DNA strands, MDA is performed isothermally (de Bourcy et al., 2014).

Real-Time PCR (qPCR)

Real-time PCR was used in previous studies (Bustin et al., 2009; Law et al., 2014) to evaluate the quality of IMS-MDA products by using fluorescence to detect and quantify nucleic acid samples in real-time. Relative abundances of the amplicons were assessed by C_T values, which are the number of cycles at which a fluorescent dye signal passes a predetermined value, or threshold (Schmittgen and Livak, 2008). The C_T values suggest if there is a sufficient amount of target pathogen DNA present in the sample for genomic sequencing. Low C_T values indicate fewer cycles are needed for the signal to reach a threshold value, meaning the sample already contains a relatively high concentration of target DNA (Schmittgen and Livak, 2008). Previous use of quasimetagenomic sequencing to detect and serotype *S.* Enteritidis illustrated that a C_T value less than 25 led to the recovery of more than 50% of the bacterial genome (Hyeon et al., 2018).

Pathogen subtyping

Subtyping classifies an organism at the subspecies level (Wiedmann, 2002). Development of molecular subtyping methods to differentiate foodborne pathogens has increased to replace slower, labor-intensive biochemical assays (Law et al., 2014). Pulsed-field gel electrophoresis (PFGE) was considered to be the gold standard for molecular subtyping of foodborne pathogens (Graves and Swaminathan, 2001). While PFGE is capable of subtyping some bacteria, it cannot differentiate between closely related strains, like *Salmonella* strains of certain serotypes (Deng et al., 2014). Whole genome sequencing is a one-stop shop for a variety of applications, which can address a range of research questions. Whole genome sequencing (WGS) has become an

increasingly routine method to subtype foodborne pathogens (Allard et al., 2016). Subtyping is important for foodborne illness surveillance because it enables differentiation between strains of the same species. For example, *Salmonella enterica* contains over 2,500 serotypes, with many subtypes per serotype and varying degrees of virulence and hosts (Braden, 2006). *Escherichia coli* has over 200 serotypes, which also have a broad range of virulence mechanisms and reservoirs (Fratamico et al., 2016). There are two primary methods to subtype bacteria using WGS data: 1) serotyping through identification of serotype-determinant genes and 2) single nucleotide polymorphism (SNP)-typing through phylogenetics. Single nucleotide polymorphism typing is used to differentiate among individual cells within the same serotype (Coll et al., 2014). This allows fingerprinting of the pathogen, which enables investigators to pinpoint the source of the pathogen and relevant clinical cases (Centers for Disease Control, 2016a). Additionally, as bacteria undergo natural selection, serotypes may become uncommon or extinct, or are created. This is a concern especially among bacteria that are adopting antimicrobial resistance genes through horizontal gene transfer. Many foodborne bacteria are adopting antibiotic resistance genes, which complicates treatment of foodborne disease because antibiotics are less or not effective (Centers for Disease Control, 2018a). Genomic data generated through WGS facilitates antimicrobial resistance monitoring and helps predict emerging pathogens (U.S. Food and Drug Administration, 2018b).

Next-generation sequencing (NGS) is a more powerful tool compared to older sequencing methods, such as Sanger sequencing (Shendure and Ji, 2008). In contrast to older sequencing methods, NGS is less expensive making it a reasonable tool for large-scale sequencing projects (Shendure and Ji, 2008). Illumina sequencing is a type of NGS that uses random fragmentation and single-molecule arrays to amplify and determine the sequence of single-stranded DNA

(Morozova and Marra, 2008). The Illumina platform is able to sequence DNA through fluorescent labeling on oligonucleotide chains (Morozova and Marra, 2008). Fluorescent labels are unique for each type of nucleotide, allowing individual nucleotides in a DNA sequence to be assigned according to the pattern of fluorescence (Morozova and Marra, 2008). Illumina sequencing produces many short reads that can be aligned to a DNA template or reference genome (Morozova and Marra, 2008). Within 24 hours, the Illumina MiSeq platform is able to generate up to 5.1 Gb of sequencing data consisting of paired-end reads up to 150 bases in length (Illumina, 2019).

CHAPTER 3

MATERIALS AND METHODS

Primer and probe design

The primers and probe for *Salmonella enterica* were designed by Malorny et al. (2004) and are shown in Table 1. The primers and probe correspond to a section of the *ttr* gene, which is a highly conserved region of the *Salmonella* genome and is responsible for tetrathionate respiration (GenBank accession no. AF 282268; (Malorny et al., 2004).

The primers and probes for *E. coli* O157:H7 are from FDA BAM protocol "Testing Methodologies for *E. coli* O157:H7 and *Salmonella* species in Spent Sprout Irrigation Water (or Sprouts)" and are shown in Table 2 (U.S. Food and Drug Administration, 2015). The primers and probe targets Shiga toxin genes (*stx*1, *stx*2, or *uid*A).

Bacteria strains

The *Salmonella* species used in this project is *Salmonella enterica* subspecies *enterica* serovar Saintpaul E2008001236 isolated from jalapenos associated with a 2008 outbreak. The *E. coli* O157:H7 strain used in this project is *E. coli* O157:H7 H1730, which is a human isolate from a lettuce outbreak. Both bacterial strains are from the University of Georgia Center for Food Safety culture collection, with the *Salmonella* and *E. coli* strains contributed by Dr. Francisco Diez and Dr. Larry Beuchat, respectively.

Romaine lettuce procurement and storage

Heads of romaine lettuce were purchased from a local grocery store in Griffin, GA. Romaine was stored at 4°C and was used within 48 h of purchase.

Culture transfer and enumeration

Glycerol freezer stocks of the bacteria were transferred into 10 mL of tryptic soy broth (TSB; Becton, Dickson and Company, Sparks, MD). The IMS-MDA workflow for each individual organism used a new overnight culture of the bacteria of interest in 10 mL of TSB. A single loopful (~10 μL) of the culture was aseptically transferred to a new 15-mL tube with 10 mL of TSB. The culture was placed at 37°C for approximately 24 h. After incubation, serial dilutions were performed from 1 mL of the overnight culture (ca. 8 log CFU) using six 15-mL tubes with 9 mL of 1x PBS (Fisher Scientific, Fair Lawn, NJ) in each tube. One hundred microliters of the three weakest dilutions (10^4 – 10^6) were aseptically plated in duplicate on tryptic soy agar (TSA; Acumedia, Lansing, MI). The plates were placed at 37°C for approximately 24 h. After incubation, colony counts were used to estimate the inoculum size.

Lettuce inoculation, enrichment, and sample collection

Figure 2 gives an overview of a single biological replicate. Romaine lettuce (25 g each) was placed in filtered sample bags on one side of the filter. Each lettuce sample was inoculated with 250 μL from the appropriate previously performed serial dilutions in 1x PBS. Sample bags were stored at 4°C for approximately 24 h to mimic cold storage during transport or at a grocery store. After 24 h, 225 mL of TSB added to each sample bag with inoculated lettuce. The bags were hand massaged for 30 sec and then the placed at 42°C. After 12 h of enrichment, the bag contents were stomached for 30 sec at 230 rpm. Approximately 45 mL of TSB was carefully poured from the filtered side of the bag into a 50-mL conical tube. Figure 1 gives a generalized overview of the quasimetagenomic workflow after enrichment.

Centrifugation

After collecting enrichment samples, the 50 mL conical tubes were weighed and paired based on similar weights to balance the centrifuge. Tubes were spun in an accuSpin 400 centrifuge (Fisher Scientific) for 10 min at 100 x g to separate large particulates from the supernatant. The supernatant was transferred to a new 50-mL conical tube and spun at 6,000 x g. After spinning, the supernatant was carefully poured from the tube keeping the pellet intact. The pellets were washed with 5 mL TSB and resuspended using a vortex mixer (Fisher Scientific). The tubes were spun once more for 10 min at 6,000 x g, then the supernatant was removed, and the pellets were washed and resuspended with 5 mL of TSB.

Immunomagnetic separation

One milliliter was removed from each resuspended solution and placed in a 1.5-mL tube. Twenty microliters of organism-specific Dynabeads (Invitrogen) (anti-*Salmonella* for *Salmonella* or anti-*E. coli* O157 for *E. coli* O157) were placed in each 1.5-mL tube containing 1 mL of sample. The 1.5-mL tubes were placed in a HulaMixer (Invitrogen) for 30 min with orbital 25, reciprocal 21°, and vibro 0°. The following steps took place inside of a biosafety cabinet. After mixing, the 1.5-mL tubes were placed in a DynaMag magnetic rack (Applied Biosystems) for three min. While keeping the tubes in the rack, the supernatant was carefully removed from each tube avoiding the beads collected on the side of the tube touching the magnet. Excess liquid in the cap of the tube was also discarded. One milliliter of 1x PBS (Fisher Scientific) with 0.05% Tween-20 (Fisher Scientific) was added to each tube. The plastic tube holder was carefully removed from the magnetic rack and, placing hands over the tops of the 1.5-mL tubes to prevent the tubes from falling out of the rack, the rack was carefully inverted 10 to 15 times to wash the beads in the 1x PBS and Tween-20 detergent. The tube holder was placed back on the magnetic rack, leaving the

tubes to rest again for three min. After three min, the tubes were aspirated, the supernatant removed, and the beads were washed. This process was repeated two more times without the washing step after the fourth 3-min bead collection step. After removing the supernatant after the fourth 3-min bead collection, the tubes were spun for approximately two sec in a mini-centrifuge to collect beads at the bottom of the tubes (Fisher Scientific). The tubes were placed back into the magnetic rack for three min, and then any remaining supernatant was removed from the tubes using a small pipette volume (100-200 μL).

Multiple displacement amplification

Multiple displacement amplification materials (sample buffer, reaction buffer, and enzyme) from the illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare Life Sciences) were previously aliquoted into separate 0.5-mL tubes to limit contamination of the original reagent stock. The sample buffer, reaction buffer, and enzyme were retrieved from storage at -20°C to thaw on ice. Use of MDA reagents took place inside of a biosafety cabinet. Nine microliters of the sample buffer were placed into the 1.5-mL tube containing the cleaned bead-bacteria complexes. The sample buffer was gently pipetted up and down and the pipette tip swirled to mix the sample buffer with the beads. A 9 μL sample was drawn from the tube and placed in a new 0.2-mL tube. This tube was immediately placed on ice to avoid a premature reaction. After completing this process for each sample, the 0.2-mL tubes were placed in a thermocycler to denature the DNA. The thermocycler was pre-programmed with the following parameters: 3 min at 95°C to denature the DNA and a final hold at 4°C. While the thermocycler was running, the master mix of enzyme and reaction buffer was prepared. Nine microliters of reaction buffer per sample were mixed with 1 μL of enzyme per sample in a 0.5-mL tube. The enzyme was added to the reaction buffer, and the mixture was gently mixed by pipetting up and down. After mixing, the master mix was placed

on ice. After the thermocycler program completed, the 0.2-mL tubes were placed on ice. Ten microliters of the master mix were placed into each 0.2-mL tube and gently mixed by pipetting up and down. After adding the master mix, the tube was placed on ice. The 0.2-mL tubes were placed in the thermocycler once more and ran on a program for DNA amplification. The following thermocycler parameters were used: 30°C for 2 h for amplification, then 65°C for 10 min to inactivate the enzyme, and a final hold at 4°C.

Quantitative PCR

Reagent preparation for qPCR took place in a PCR cabinet. The qPCR master mix contained 2X TaqMan Universal master mix (Applied Biosystems), forward primers, reverse primers, probe, and nuclease-free water. For *Salmonella,* 10 μL of universal master mix, 2 μL of each primer (900 nM), 2 μL of the probe (250 nM), and 2 μL of molecular grade water were used per sample. For *E. coli* O157:H7, 10 μL of universal master mix, 1.8 μL of each primer (900 nM), 0.5 μL of the probe (250 nM), and 3.9 μL of molecular grade water were used per sample. Two microliters of the MDA product and 18 μL of the master mix were used per sample per optical 0.1-mL qPCR tube (Applied Biosystems). Quantitative PCR was completed on a StepOne Real-Time PCR system (Applied Biosystems). Parameters for qPCR include two holds, one at 50°C for 2 min and another at 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 60 sec.

Genome sequencing

Prior to sequencing, DNA concentrations of MDA products were determined using a Qubit Broad Range (BR) dsDNA assay kit (Invitrogen). Concentrations of each sample were diluted to approximately 0.2 ng/μL before sequencing. The MiSeq Nextera DNA Library Prep kit, MiSeq

Reagent kit (V2), and MiSeq Index kit (all from Illumina) were used for Illumina paired-end sequencing.

Bioinformatic analysis

Raw reads were quality checked using FastQC (v0.10.1) (Andrews, 2010). Raw reads were trimmed using Trimmomatic (v0.36), which removed three nucleotides from the leading and trailing ends of reads, used a four nucleotide sliding window to remove additional nucleotides from the 3' end when the Phred score was less than 20 and discarded reads fewer than 20 base pairs after trimming (Bolger et al., 2014). Kraken2 (v2.0.7-beta) with the standard database was used to taxonomically classify reads. Then, reads with taxonomy IDs matching *Salmonella enterica* (ID: 28901) or *Escherichia coli* (ID: 562) were extracted and used for *de novo* assembly (Wood and Salzberg, 2014). *De novo* assembly was completed using SPAdes (v3.9.0) with the "--careful" option (Bankevich et al., 2012).

Extracted taxonomic-specific reads were serotyped using SeqSero (Zhang et al., 2015) for *Salmonella enterica*, and SeroTypeFinder 2.0 (Joensen et al., 2015), for *Escherichia coli.* With SeqSero, raw sequencing reads were directly used. With SeroTypeFinder, raw reads were *de novo* assembled into draft genomes using SPAdes before serotype prediction. The N_{50} value was determined using QUAST (v4.4) (Gurevich et al., 2013). Output (total million bases of raw reads per sample), coverage (percentage of reference genome mapped by reads), and depth ratio (ratio between the total size of *Salmonella* or *E. coli* sequences per 100 million bases of sequencing data and the size of the corresponding reference genome) were also determined.

$$\text{Depth} = \frac{\text{number of base pairs in quasimetagenomic reads}}{\text{number of base pairs in reference genome}}$$

$$\text{Output} = \text{size (Mb) of metagenomic reads after trimming}$$

$$\text{Depth ratio} = \frac{\text{depth} * 10^7}{\text{output}}$$

A core genome single nucleotide polymorphism (SNP) phylogeny for each pathogen was constructed using quasimetagenomics samples and isolates selected from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA). Tables 3 and 4 contain metadata for NCBI isolates for *Salmonella* Saintpaul and *E. coli* O157:H7, respectively. Isolates were screened, using either SeqSero for *Salmonella* or SeroTypeFinder 2.0 for *E. coli,* to ensure their designated serotype was listed correctly on NCBI. In the *Salmonella* Saintpaul phylogeny, two isolates represented one outbreak associated with jalapeno and serrano peppers in 2008. Two other well-known Saintpaul outbreaks, cucumbers (2013) and alfalfa sprouts (2009), were not included because publicly-available and applicable SRA data could not be found. Some isolates from the *E. coli* phylogeny were from two outbreaks: a 2006 spinach outbreak and a 2008 lettuce outbreak. Each SNP phylogeny was assembled with the Center for Food Safety and Applied Nutrition (CFSAN) SNP Pipeline (v2.0.2) with default settings.

CHAPTER 4

RESULTS

qPCR data

Given in Tables 5 and 6, C_T values for all samples, except for negative controls, were less than 27. Samples are designated by the strain used for inoculation (EC for *E. coli* O157:H7 and SE for *S. enterica* serovar Saintpaul), as well as inoculum level (0.1, 1, or 10 CFU/g). Negative controls include a no inoculum enrichment for each biological replicate, as well as an MDA negative control to ensure reagents were not previously contaminated. All negative controls had undetermined C_T values. Overall, C_T values associated with *Salmonella* detection are lower than that of EHEC detection.

Sequencing metrics

Table 7 exhibits various bioinformatic metrics for 28 samples from three biological replicates with varying inoculum levels. These metrics include sequencing output, reference genome coverage, N_{50} of target genome assembly, serotype prediction, percent EC or *S. enterica* reads of all reads, and percent abundance of EC or *S. enterica* reads of all microbial reads. Sequencing output per sample ranged from 109 Mb to 660 Mb. Reference genome coverage, which is the percent of the reference genome that was mapped by raw reads, ranged from 4.77% to 98.41%. The lowest coverage was 4.77% of the *Salmonella* Saintpaul reference genome from a sample with 0.1 CFU/g inocula of both pathogens. This sample also exhibited the lowest depth ratio (0.0240). Values for N_{50}, which is the minimum contig size that covers 50% of the genome, ranged from 545 to 19,919. The percent reads of EC or *S. enterica*, which is the percentage of

reads that were identified as EC or *S. enterica* out of all reads, were between 0.21% and 61.73%. The percent abundance, which is the percent of reads identified as EC or *S. enterica* out of all bacterial reads, ranged from 0.42% to 88.41%.

Serotyping

Serotype information is also given in Table 7. Of all 28 samples, 19 (~68%) were correctly serotyped using only serotype determinant genes. Samples containing a single pathogen inoculum at various levels were all accurately serotyped with the exception of EC only at 0.1 CFU/g. For co-enrichment samples, eight of eleven using anti-*E. coli* O157 beads were serotyped. Of those samples using anti-*Salmonella* IMS beads, six of eleven were serotyped. Some co-enrichment samples were able to be serotyped at very low levels (0.1 CFU/g). For example, among co-enrichment samples with EC having the lowest inoculum level (0.1 CFU/g) and using anti-*E. coli* O157:H7 IMS beads, four out of five were correctly serotyped. However, among co-enrichment samples with *S. enterica* having the lowest inoculum level (0.1 CFU/g) and using anti-*Salmonella* IMS beads, only one of five was successfully serotyped.

"Sample pairs" refers to two enrichment samples drawn from the same enrichment bag in a single biological replicate but are subjected to different targeted IMS beads. Overall, five sample pairs were serotyped. For samples inoculated with ~1 CFU/g of each pathogen, all sample pairs were serotyped except for the anti-EC IMS bead sample in replicate three. For samples inoculated with ~0.1 CFU/g of each pathogen, neither of the two sample pairs were serotyped, except the anti-EC IMS bead sample in replicate three. Sample pairs with high EC inoculum (~1 CFU/g) and low *S. enterica* inoculum (~0.1 CFU/g) were both serotyped in replicate 1, not serotyped in replicate 2, and only serotyped for the anti-EC IMS bead sample in replicate 3. Sample pairs with low EC

inoculum (~0.1 CFU/g) and high *S. enterica* inoculum (~1 CFU/g) were both serotyped in replicates 2 and 3. However, the anti-*Salmonella* IMS bead sample in replicate 1 was not serotyped.

Microbial abundance

Figure 3 displays a stacked bar graph illustrating the percent microbial abundance of bacterial genera within each *Salmonella* Saintpaul-targeted quasimetagenomic sample. Each sample contains *Salmonella,* with SE1 having the greatest abundance (~70%) of *Salmonella*. Samples with the lowest *Salmonella* abundance (< ~2%) include EC1 SE0.1 and EC0.1 SE0.1 from both replicates 2 and 3. Figure 4 also illustrates a stacked bar graph providing the percent microbial abundance of several bacteria genera within each quasimetagenomic sample targeting *E. coli* O157:H7. Samples with the greatest percent abundance (~95%) of *E. coli* include EC1 SE1 from replicates 1 and 2, as well as EC0.1 SE1 (rep1). Samples with the lowest percent abundance are EC1 SE0.1 (rep2) and EC0.1 SE0.1 (rep2). Overall, *E. coli*-targeted quasimetagenomic samples display a greater relative abundance of *E. coli* compared to that of *Salmonella* in *Salmonella*-targeted quasimetagenomic samples.

All samples contained DNA from other species of bacteria. In Figure 3, quasimetagenomic sample SE1 displays the greatest abundance (~80%) of *Salmonella*. Samples EC1 SE1 (rep2), EC0.1 SE1 (rep2), and SE10 show similar abundances (~30-40%) of *Salmonella*. EC 0.1 SE1 (rep1) has the greatest abundance (~90%) of *Streptococcus.* Samples EC1 SE0.1 (rep1), EC1 SE1 (rep2), EC1 SE1 (rep3), EC0.1 SE1 (rep2), and EC0.1 SE0.1 (rep2) have similar abundances (~60%) of *Streptococcus*. Other frequently present genera include *Paeniclostridium, Bacillus, Staphylococcus,* and *Acinetobacter*.

The most abundant bacteria genera in Figure 4 are *Escherichia, Streptococcus, Paeniclostridium, Bacillus, and Klebsiella.* Samples EC1 SE1 (rep1), EC1 SE1 (rep2), and EC0.1

SE1 (rep #) show the largest percentage abundance (~90%) of *Escherichia*. Consistent with data in Figure 3, *Streptococcus* is also seen in samples in Figure 4, with EC1 SE0.1 (rep2) and EC1 only having approximately 70% abundance of *Streptococcus. Paeniclostridium, Bacillus, and Klebsiella* are also present in some samples at abundances no greater than 40%.

SNP typing

The phylogenetic tree, given in Figure 5, contains only *Salmonella* Saintpaul isolates from Salmonella-associated food sources, such as kratom and ground turkey. Four samples were not included in Figure 5 because their sequencing data did not allow the detection of any high-quality SNP loci, which are given in Table 8. The average number of SNP loci across all *Salmonella*-targeted samples is approximately 411. The EC SNP phylogeny (Figure 6) only contains *E. coli* O157:H7 isolates from food sources, which are primarily leafy greens and beef. All EC-targeted quasimetagenomic samples given in Figure 6 contained SNP loci, which are shown in Table 9. The average number of SNP loci across *E. coli*-targeted samples is approximately 1,294. Both tree figures used multi-FASTA alignment files without gaps from the CFSAN SNP pipeline.

CHAPTER 5

DISCUSSION

The lack of an efficient or complete "kill step" during fresh produce production can allow pathogens to survive on fresh produce. Therefore, it is important to have rapid pathogen detection and subtyping methods to assist identification of pathogen contaminants and contaminated food, which can lead to quicker response to foodborne disease outbreaks. Quasimetagenomics sequencing has been previously described as a method to detect and subtype a single pathogen in one workflow (Hyeon et al., 2018). However, this method has not been used to detect and subtype multiple foodborne pathogens. Therefore, this study assessed the ability of the quasimetagenomics sequencing method to be used to concertedly subtype two pathogens within a common food matrix.

Results from Hyeon et al. (2018) indicate C_T values below 25 for *Salmonella enterica* yield large enough target DNA concentration for accurate pathogen identification. All C_T values given in Tables 5 and 6 are approximately 27 or below, which indicates that the majority of samples were likely to provide enough target DNA for pathogen identification. Table 5 shows lower C_T values (approximately 18) for samples with 1 CFU/g inoculums of both pathogens and greater *Salmonella* inoculum compared to that of EHEC. Overall, C_T values associated with *Salmonella* detection are lower than that of EHEC detection. This may be due to differences in cell capacity for each IMS bead type or variations in qPCR efficiency between each target organism. Results in Table 6 illustrates very similar C_T values for EHEC detection regardless of inoculum size. This result may be due to the maximum capacity of cells allowed on anti-EHEC IMS beads or MDA efficiency.

Co-enrichment samples refer to enrichment bags containing inoculums of both *Salmonella* Saintpaul and *E. coli* O157:H7. For co-enrichment samples, those subjected to anti-*E. coli* O157:H7 IMS beads were more frequently serotyped (8/11, or 72.7%) compared to those using anti-*Salmonella* IMS beads (6/11, or 54.5%). This is further supported by the difference in average sequencing coverage across all *Salmonella*- or *E. coli*-targeted samples. For *Salmonella*-targeted samples, the average sequencing coverage was approximately 65%, while for *E. coli*-targeted samples, the average sequencing coverage was approximately 85%. This may have occurred for several reasons. First, *E. coli* O157:H7 may have a competitive advantage over *Salmonella* Saintpaul within the first 12 h of enrichment in TSB. Several studies have investigated *E. coli* and *Salmonella* survival in a competitive environment. *Escherichia coli* has demonstrated a greater survival capacity than *Salmonella* Typhimurium in tropical estuarine water (Chandran and Hatha, 2005). Indole, which can be produced by *E. coli* during the stationary phase, has shown to suppress genes involved in *Salmonella* motility and invasion (Nikaido et al., 2012). In addition, anti-*E. coli* O157:H7 IMS beads may have a greater cell capacity compared to anti-*Salmonella* IMS beads. This may increase the overall number of cells carried on the bead, which would also increase the amount of genetic material associated with those cells.

Approximately 70% of quasimetagenomic samples were able to be serotyped. Variation in successful serotyping could be due to the duration of culture enrichment, the food matrix and corresponding microbiome, or variation of bacterial growth (Hyeon et al., 2018). Samples unable to be serotyped all had depth ratios below ~1, indicating a relatively low total size of target organism sequences (per 100 million bases) compared against the size of the corresponding reference genome. Also, these samples typically covered less than ~68% of the genome, which led to low or no sequencing coverage at one or more serotype determinant loci and affected serotype

determination. On average, samples without an assigned serotype had a percent of *S. enterica* or EC reads of all total reads less than 4%, as well as a percent of *S. enterica* or EC reads of all bacteria reads (percent abundance) less than 12%. The only exception was a sample targeting *Salmonella*, EC1 SE1 (rep3), which was successfully serotyped with 2.38% *Salmonella enterica* reads out of all reads and 5.82% abundance of *Salmonella enterica* reads out of only bacterial reads. This may have occurred due to the genetic data contained within these reads, which may happen to contain enough information to allow serotyping. This data could include more SNPs on average or areas of the genome characteristic to the Saintpaul serovar.

For the stacked bar charts in Figures 3, which show *Salmonella*-targeted quasimetagenomic samples, three of the samples from biological replicate 2 have a relatively large abundance of *Streptococcus*. This may suggest the head of romaine used in replicate 2 possibly had a relatively large presence of *Streptococcus*. Many of the non-target genera shown in Figures 3 and 4 are prevalent in soil, so they were most likely introduced from the field prior to harvest. This also indicates non-target species carryover on IMS beads. Recently published evidence suggests smaller IMS bead size can increase recovery of cells (Chen and Park, 2018). Therefore, the use of a smaller IMS bead size may increase the recovery of target cells, which would theoretically increase sequencing data to assist in pathogen identification. Background flora on lettuce may vary in seasonality, location, and types of handling and packaging. This study used naked packed romaine from refrigerated shelving at a local grocery store, which is subjected to handling from both consumers and employees, water misting systems, and contact with other produce. It is possible that romaine contained within plastic bags or boxes may have less exposure to sources of bacterial contamination and a less diverse background microbiota. Using a wrapped packaged

romaine may help decrease background flora, which may increase the capture efficiency of IMS beads.

In the *Salmonella* Saintpaul SNP phylogeny (Figure 5), clustering is observed between the quasimetagenomic samples, the reference strain used for inoculation, and isolates from the 2008 *Salmonella* Saintpaul jalapeno outbreak. This indicates that strain-level subtyping is achievable for a variety of inoculum levels, including co-enrichment with *S. enterica* inoculum as low as 0.1 CFU/g. The *E. coli* O157:H7 SNP phylogeny (Figure 6) also illustrates clustering between the reference and quasimetagenomic samples, which demonstrates that SNP-level serotyping is achievable for EC inoculums as low as 0.1 CFU/g. Overall, 24/28 samples (85.7%) were able to be subtyped through SNP-typing.

In conclusion, these results provide evidence that concerted detection of *Salmonella enterica* and *E. coli* O157:H7 is possible using a quasimetagenomics method. Compared against its original iteration in Hyeon et at. (2018), this method is also able to subtype samples within 12 h of enrichment but does not use selective enrichment media. It also has high sensitivity, enabling sufficient growth and concentration of target pathogens for SNP-based subtyping from an initial inoculum of 0.1 CFU/g. Future applications of this method may include outbreaks with multiple virulent serovars or species within the same food matrix. With an increase in reported fresh produce-related illness in the last twenty years, it is anticipated this method will assist in faster traceback and pathogen identification during foodborne disease outbreaks to prevent additional illnesses. This method can also be used to monitor pathogen presence in various areas of the food supply chain.

REFERENCES

Allard, M.W., Strain, E., Melka, D., Bunning, K., Musser, S.M., Brown, E.W., Timme, R., 2016. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J. Clin. Microbiol. 54, 1975-1983. DOI: 10.1128/jcm.00081-16.

Andino, A., Hanning, I., 2015. *Salmonella enterica*: survival, colonization, and virulence differences among serovars. Sci. World J. 2015, 520179. DOI: 10.1155/2015/520179.

Andrews, S., 2010. FastQC: a quality control tool for high throughput sequence data.

Aw, T.G., Wengert, S., Rose, J.B., 2016. Metagenomic analysis of viruses associated with field-grown and retail lettuce identifies human and animal viruses. Int. J. Food Microbiol. 223, 50-56. DOI: 10.1016/j.ijfoodmicro.2016.02.008.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455-477.

Barnhart, M.M., Chapman, M.R., 2006. Curli biogenesis and function. Annu. Rev. Microbiol. 60, 131-147. DOI: 10.1146/annurev.micro.60.080805.142106.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30, 2114-2120.

Braden, C.R., 2006. *Salmonella enterica* serotype Enteritidis and eggs: a national epidemic in the United States. Clin. Infect. Dis. 43, 512-517. DOI: 10.1086/505973.

Brandl, M.T., 2008. Plant lesions promote the rapid multiplication of *Escherichia coli* O157:H7 on postharvest lettuce. Appl. Environ. Microbiol. 74, 5285-5289. DOI: 10.1128/aem.01073-08.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611-622.

Centers for Disease Control, 2009. Multistate outbreak of *Salmonella* Saintpaul infections linked to raw alfalfa sprouts (final update). https://www.cdc.gov/salmonella/2009/raw-alfalfa-sprouts-5-8-2009.html. Accessed date: 20 May 2019.

Centers for Disease Control, 2014. Timeline for reporting cases of *E. coli* O157 infection. https://www.cdc.gov/ecoli/reporting-timeline.html. Accessed date: 9 March 2018.

Centers for Disease Control, 2016a. About PulseNet: frequently asked questions. https://www.cdc.gov/pulsenet/about/faq.html. Accessed date: 9 June 2019.

Centers for Disease Control, 2016b. Burden of foodborne illness: overview. https://www.cdc.gov/foodborneburden/estimates-overview.html. Accessed date: 9 March 2018.

Centers for Disease Control, 2018a. Antibiotic resistance and food safety. https://www.cdc.gov/foodsafety/challenges/antibiotic-resistance.html. Accessed date: 10 June 2019.

Centers for Disease Control, 2018b. Multistate outbreak of *E. coli* O157:H7 infections linked to romaine lettuce (final update). https://www.cdc.gov/ecoli/2018/o157h7-04-18/index.html. Accessed date: 6 December 2018.

Centers for Disease Control, 2018c. Outbreak of *E. coli* infections linked to romaine lettuce. https://www.cdc.gov/ecoli/2018/o157h7-11-18/index.html. Accessed date: 6 December 2018.

Centers for Disease Control, 2019. Timeline for reporting cases of *Salmonella* infection. https://www.cdc.gov/salmonella/reporting-timeline.html. Accessed date: 9 June 2019.

Chandran, A., Hatha, A.A.M., 2005. Relative survival of *Escherichia coli* and *Salmonella typhimurium* in a tropical estuary. Water Res. 39, 1397-1403. DOI: 10.1016/j.watres.2005.01.010.

Chen, J., Park, B., 2018. Effect of immunomagnetic bead size on recovery of foodborne pathogenic bacteria. Int. J. Food Microbiol. 267, 1-8.

Chen, M.F., Song, P.F., Zou, D., Hu, X.S., Zhao, S.C., Gao, S.J., Ling, F., 2014. Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. Plos One. 9, 12. DOI: 10.1371/journal.pone.0114520.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.F., Jagadeesan, B., Rantsiou, K., Phister, T., 2018. Next generation microbiological risk assessment meta-omics: the next need for integration. Int. J. Food Microbiol. 287, 10-17. DOI: 10.1016/j.ijfoodmicro.2017.11.008.

Coll, F., McNerney, R., Guerra-Assuncao, J.A., Glynn, J.R., Perdigao, J., Viveiros, M., Portugal, I., Pain, A., Martin, N., Clark, T.G., 2014. A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. Nat. Commun. 5, 5. DOI: 10.1038/ncomms5812.

Danyluk, M.D., Schaffner, D.W., 2011. Quantitative assessment of the microbial risk of leafy greens from farm to consumption: preliminary framework, data, and risk estimates. J. Food Prot. 74, 700-708. DOI: 10.4315/0362-028x.Jfp-10-373.

de Bourcy, C.F.A., De Vlaminck, I., Kanbar, J.N., Wang, J.B., Gawad, C., Quake, S.R., 2014. A quantitative comparison of single-cell whole genome amplification methods. Plos One. 9, 9. DOI: 10.1371/journal.pone.0105585.

Deng, X.Y., Desai, P.T., den Bakker, H.C., Mikoleit, M., Tolar, B., Trees, E., Hendriksen, R.S., Frye, J.G., Porwollik, S., Weimer, B.C., Wiedmann, M., Weinstock, G.M., Fields, P.I., McClelland, M., 2014. Genomic epidemiology of *Salmonella enterica* serotype Enteritidis based on population structure of prevalent lineages. Emerging Infectious Diseases. 20, 1481-1489. DOI: 10.3201/eid2009.131095.

Erickson, M., Ortega, Y., 2018. Romaine lettuce. http://fsi.colostate.edu/romaine-lettuce-2/. Accessed date: 10 December 2018.

Fratamico, P.M., DebRoy, C., Liu, Y.H., Needleman, D.S., Baranzoni, G.M., Feng, P., 2016. Advances in molecular serotyping and subtyping of *Escherichia coli*. Front. Microbiol. 7, 8. DOI: 10.3389/fmicb.2016.00644.

Gil, M.I., Selma, M.V., Suslow, T., Jacxsens, L., Uyttendaele, M., Allende, A., 2015. Pre-and postharvest preventive measures and intervention strategies to control microbial food safety hazards of fresh leafy vegetables. Crit. Rev. Food Sci. Nutr. 55, 453-468.

Graves, L.M., Swaminathan, B., 2001. PulseNet standardized protocol for subtyping *Listeria monocytogenes* by macrorestriction and pulsed-field gel electrophoresis. Int. J. Food Microbiol. 65, 55-62. DOI: 10.1016/s0168-1605(00)00501-8.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29, 1072-1075.

Herman, K.M., Hall, A.J., Gould, L.H., 2015. Outbreaks attributed to fresh leafy vegetables, United States, 1973-2012. Epidemiol. Infect. 143, 3011-3021. DOI: 10.1017/S0950268815000047.

Hyeon, J.Y., Deng, X.Y., 2017. Rapid detection of *Salmonella* in raw chicken breast using real-time PCR combined with immunomagnetic separation and whole genome amplification. Food Microbiol. 63, 111-116. DOI: 10.1016/j.fm.2016.11.007.

Hyeon, J.Y., Li, S.T., Mann, D.A., Zhang, S.K., Li, Z., Chen, Y., Deng, X.Y., 2018. Quasimetagenomics-based and real-time-sequencing-aided detection and subtyping of *Salmonella enterica* from food samples. Appl. Environ. Microbiol. 84, 15. DOI: 10.1128/aem.02340-17.

Illumina, 2019. Specifications for the MiSeq system. https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html. Accessed date: 9 June 2019.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., Scheutz, F., 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J. Clin. Microbiol. 53, 2410-2426.

Kerns, D.L., Matheron, M.E., Palumbo, J.C., Sanchez, C.A., Still, D.W., Tickes, B.R., Umeda, K., 1999. Guidelines for head lettuce production in Arizona. https://cals.arizona.edu/crop/vegetables/cropmgt/az1099.html#variety. Accessed date: 23 May 2019.

Klontz, K.C., Klontz, J.C., Mody, R.K., Hoekstra, R.M., 2010. Analysis of tomato and jalapeño and serrano pepper imports into the United States from Mexico before and during a national outbreak of *Salmonella* serotype Saintpaul infections in 2008. J. Food Prot. 73, 1967-1974. DOI: 10.4315/0362-028x-73.11.1967.

Koukkidis, G., Haigh, R., Allcock, N., Jordan, S., Freestone, P., 2017. Salad leaf juices enhance *Salmonella* growth, colonization of fresh produce, and virulence. Appl. Environ. Microbiol. 83, e02416-02416.

Kroupitski, Y., Pinto, R., Brandl, M.T., Belausov, E., Sela, S., 2009. Interactions of *Salmonella enterica* with lettuce leaves. J. Appl. Microbiol. 106, 1876-1885. DOI: 10.1111/j.1365-2672.2009.04152.x.

Kumar, G., Rech, R., Kapolka, K., Lavrenov, K., Garnova, E., Lavasini, S., Deadman, R., Hamilton, S., 2007. Genomic DNA preparation using illustra GenomiPhi V2 and HY DNA amplification kits., Nat. Methods.

Lasken, R.S., 2007. Single-cell genomic sequencing using multiple displacement amplification. Curr. Opin. Microbiol. 10, 510-516.

Law, J.W.-F., Ab Mutalib, N.-S., Chan, K.-G., Lee, L.-H., 2014. Rapid methods for the detection of foodborne bacterial pathogens: principles, applications, advantages and limitations. Front. Microbiol. 5, 770. DOI: 10.3389/fmicb.2014.00770.

Leonard, S.R., Mammel, M.K., Lacher, D.W., Elkins, C.A., 2015. Application of metagenomic sequencing to food safety: detection of Shiga toxin-producing *Escherichia coli* on fresh bagged spinach. Appl. Environ. Microbiol. 81, 8183-8191. DOI: 10.1128/aem.02601-15.

Lim, J.Y., Yoon, J.W., Hovde, C.J., 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. J. Microbiol. Biotechn. 20, 5-14.

Lin, D., 2006. Metagenomics: a new look into the world of microbes. http://www.hcs.harvard.edu/~hsr/wp-content/themes/hsr/pdf/spring2006/lin.pdf. Accessed date: 16 April 2018.

Malorny, B., Paccassoni, E., Fach, P., Bunge, C., Martin, A., Helmuth, R., 2004. Diagnostic real-time PCR for detection of *Salmonella* in food. Appl. Environ. Microbiol. 70, 7046-7052. DOI: 10.1128/AEM.70.12.7046-7052.2004.

Marler, B., 2018. A long, long history of outbreaks linked to leafy greens. https://www.foodpoisonjournal.com/foodborne-illness-outbreaks/a-long-long-history-of-outbreaks-linked-to-leafy-greens/. Accessed date: 30 January 2019.

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

Morozova, O., Marra, M.A., 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics. 92, 255-264. DOI: 10.1016/j.ygeno.2008.07.001.

Niemira, B.A., Cooke, P.H., 2010. *Escherichia coli* O157:H7 biofilm formation on romaine lettuce and spinach leaf surfaces reduces efficacy of irradiation and sodium hypochlorite washes. J. Food Sci. 75, M270-M277. DOI: 10.1111/j.1750-3841.2010.01650.x.

Nikaido, E., Giraud, E., Baucheron, S., Yamasaki, S., Wiedemann, A., Okamoto, K., Takagi, T., Yamaguchi, A., Cloeckaert, A., Nishino, K., 2012. Effects of indole on drug resistance and virulence of *Salmonella enterica* serovar Typhimurium revealed by genome-wide analyses. Gut Pathog. 4, 13. DOI: 10.1186/1757-4749-4-5.

Olaimat, A.N., Holley, R.A., 2012. Factors influencing the microbial safety of fresh produce: a review. Food Microbiol. 32, 1-19. DOI: 10.1016/j.fm.2012.04.016.

Patel, J., Sharma, M., 2010. Differences in attachment of *Salmonella enterica* serovars to cabbage and lettuce leaves. Int. J. Food Microbiol. 139, 41-47. DOI: 10.1016/j.ijfoodmicro.2010.02.005.

Ramos, B., Miller, F.A., Brandao, T.R.S., Teixeira, P., Silva, C.L.M., 2013. Fresh fruits and vegetables - an overview on applied methodologies to improve its quality and safety. Innov. Food Sci. Emerg. Technol. 20, 1-15. DOI: 10.1016/j.ifset.2013.07.002.

Schmittgen, T.D., Livak, K.J., 2008. Analyzing real-time PCR data by the comparative C-T method. Nat. Protoc. 3, 1101-1108. DOI: 10.1038/nprot.2008.73.

Seo, S., Matthews, K.R., 2012. Influence of the plant defense response to *Escherichia coli* O157: H7 cell surface structures on survival of that enteric pathogen on plant surfaces. Appl. Environ. Microbiol. 78, 5882-5889.

Seth-Smith, H.M.B., Harris, S.R., Scott, P., Parmar, S., Marsh, P., Unemo, M., Clarke, I.N., Parkhill, J., Thomson, N.R., 2013. Generating whole bacterial genome sequences of low-abundance species from complex samples with IMS-MDA. Nat. Protoc. 8, 2404-2412. DOI: 10.1038/nprot.2013.147.

Shendure, J., Ji, H.L., 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26, 1135-1145. DOI: 10.1038/nbt1486.

Smith, R., Cahn, M., Daugovish, O., Koike, S., Natwick, E., Smith, H., Subbarao, K., Takele, E., Turini, T., 2011. Leaf lettuce production in California. University of California Agriculture and Natural Resources.

U.S. Food and Drug Administration, 2009. Draft guidance for industry: guide to minimize microbial food safety hazards of leafy greens. https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/ProducePlantProducts/ucm174200.htm. Accessed date: 7 December 2018.

U.S. Food and Drug Administration, 2012. Bad bug book, foodborne pathogenic microorganisms and natural toxins, Second ed.

U.S. Food and Drug Administration, 2015. Testing methodologies for *E. coli* O157:H7 and *Salmonella* species in spent sprout irrigation water (or sprouts). https://www.fda.gov/downloads/Food/FoodScienceResearch/LaboratoryMethods/UCM467055.pdf. Accessed date: 18 July 2018.

U.S. Food and Drug Administration, 2018a. Foodborne illnesses: what you need to know. https://www.fda.gov/food/resourcesforyou/consumers/ucm103263.htm. Accessed date: 31 January 2019.

U.S. Food and Drug Administration, 2018b. Whole genome sequencing (WGS) program. https://www.fda.gov/food/science-research-food/whole-genome-sequencing-wgs-program#Monitoring. Accessed date: 10 June 2019.

U.S. Food and Drug Administration (FDA), 2018. Environmental assessment of factors potentially contributing to the contamination of romaine lettuce implicated in a multi-state outbreak of *E. coli* O157:H7. https://www.fda.gov/Food/RecallsOutbreaksEmergencies/Outbreaks/ucm624546.htm. Accessed date: March 29.

Wachtel, M.R., Whitehand, L.C., Mandrell, R.E., 2002. Association of *Escherichia coli* O157:H7 with preharvest leaf lettuce upon exposure to contaminated irrigation water. J. Food Prot. 65, 18-25. DOI: 10.4315/0362-028x-65.1.18.

Wiedmann, M., 2002. Subtyping of bacterial foodborne pathogens. Nutr. Rev. 60, 201-208.

Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46.

Yao, K., Muruvanda, T., Allard, M.W., Hoffmann, M., 2017. Complete genome sequences of three *Salmonella enterica* subsp. *enterica* serovar Saintpaul isolates associated with a 2013 multistate outbreak in the United States. Microbiol. Resour. Ann. 5, 2. DOI: 10.1128/genomeA.00456-17.

Zhang, S., Yin, Y., Jones, M.B., Zhang, Z., Kaiser, B.L.D., Dinsmore, B.A., Fitzgerald, C., Fields, P.I., Deng, X., 2015. *Salmonella* serotype determination utilizing high-throughput genome sequencing data. J. Clin. Microbiol. 53, 1685-1692.

Zhao, X., Lin, C.W., Wang, J., Oh, D.H., 2014. Advances in rapid detection methods for foodborne pathogens. J. Microbiol. Biotechn. 24, 297-312. DOI: 10.4014/jmb.1310.10013.

Table 1
qPCR primers and probes for *Salmonella* detection (Malorny et al., 2004).

	Target gene	GenBank number	Sequence
Primers	*ttr6*	AF282268	CTCACCAGGAGATTACAACATGG
	ttr4		AGCTCAGACCAAAAGTGACCATC
Probe	*ttr5*		FAM-CACCGACGGCGAGACCGACTTT-Dark Quencher

Table 2
qPCR primers and probes for *E. coli* O157:H7 detection (U.S. Food and Drug Administration, 2015).

	Target gene	Orientation	GenBank number	Sequence (5' to 3')
Primers	*Stx1*	Forward	M19473	GTGGCATTAATACTGAATTGTCATCA
	Stx1	Reverse	M19473	GCGTAATCCCACGGACTCTTC
	Stx2	Forward	X07865	GATGTTTATGGCGGTTTTATTTGC
	Stx2	Reverse	X07865	TGGAAAACTCAATTTTACCTTTAGCA
	UidA	Forward	AF305917	CAGTCTGGATCGCGAAAACTG
	UidA	Reverse	AF305917	ACCAGACGTTGCCCACATAATT
Probes	*Stx1*	N/A	M19473	TXRD-TGATGAGTTTCCTTCTATGTGTCCGGCAGAT-BHQ2
	Stx2	N/A	X07865	6FAM-TCTGTTAATGCAATGGCGGCGGATT-BHQ1
	UidA	N/A	AF305917	TET-ATTGAGCAGCGTTGG-MGB/NFQ

Table 3

List of NCBI isolates for *Salmonella enterica* subsp. *enterica* serovar Saintpaul used in the SNP phylogeny in Figure 5.

Experiment Accession	Experiment Title	Organism Name	Instrument	Submitter	Study Accession	Study Title	Sample Accession	Total Size (Mb)	Total Spots	Total Bases	Library Name
SRX346838	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004125 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004125	Illumina MiSeq	CFSAN	SRP029641	Foodborne Pathogen Survey	SRS477772	89.14	399152	157574498	Salmonella enterica str. CFSAN004125_01 shotgun
SRX4022964	Whole genome Illumina MiSeq sequence of Salmonella enterica subsp. enterica serovar Saintpaul	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	CFSAN	SRP018785	GenomeTrakr Project: US Food and Drug Administration	SRS3243450	159.26	639897	279358898	Nextera XT library SEQ000074000
SRX4149038	Whole genome Illumina MiSeq sequence of Salmonella enterica subsp. enterica serovar Saintpaul	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	CFSAN	SRP032981	GenomeTrakr Project: Minnesota Department of Health	SRS3362623	124.35	517653	246275156	Nextera XT library SEQ000075382
SRX347764	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004139 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004139	Illumina MiSeq	CFSAN	SRP029835	Foodborne Pathogen Survey	SRS478554	190.79	1072031	318678728	Salmonella enterica str. CFSAN004139_01 shotgun
SRX347765	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004140 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004140	Illumina MiSeq	CFSAN	SRP029836	Foodborne Pathogen Survey	SRS478555	329.31	1812388	537669955	Salmonella enterica str. CFSAN004140_01 shotgun
SRX346849	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004116 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004116	Illumina MiSeq	CFSAN	SRP029651	Foodborne Pathogen Survey	SRS477783	37.88	142686	65557532	Salmonella enterica str. CFSAN004116_01 shotgun
SRX347762	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004136 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004136	Illumina MiSeq	CFSAN	SRP029833	Foodborne Pathogen Survey	SRS478552	211.29	1323795	370117900	Salmonella enterica str. CFSAN004136_01 shotgun
SRX347790	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004159 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004159	Illumina MiSeq	CFSAN	SRP029855	Foodborne Pathogen Survey	SRS478580	679.05	2368226	1096862153	Salmonella enterica str. CFSAN004159_01 shotgun
SRX347796	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004166 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004166	Illumina MiSeq	CFSAN	SRP029861	Foodborne Pathogen Survey	SRS478586	261.39	858748	411840498	Salmonella enterica str. CFSAN004166_01 shotgun
SRX347798	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004188 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004188	Illumina MiSeq	CFSAN	SRP029863	Foodborne Pathogen Survey	SRS478588	251.38	1020737	485129194	Salmonella enterica str. CFSAN004188_01 shotgun
SRX346840	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004128 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004128	Illumina MiSeq	CFSAN	SRP029643	Foodborne Pathogen Survey	SRS477774	39.87	158212	67605919	Salmonella enterica str. CFSAN004128_01 shotgun
SRX346841	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004129 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004129	Illumina MiSeq	CFSAN	SRP029644	Foodborne Pathogen Survey	SRS477775	95.61	337154	161210441	Salmonella enterica str. CFSAN004129_01 shotgun
SRX1643543	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004120	Illumina MiSeq	CFSAN	SRP029647	Foodborne Pathogen Survey	SRS477779	186.75	638566	305196578	Salmonella enterica subsp. enterica Nextera XT shotgun
SRX346846	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica	Salmonella enterica subsp. enterica	Illumina MiSeq	CFSAN	SRP029648	Foodborne Pathogen Survey	SRS477780	121.46	444128	211264499	Salmonella enterica str.

	serovar Saintpaul str. CFSAN004121 by Illumina MiSeq	serovar Saintpaul str. CFSAN004121									CFSAN004121_01 shotgun
SRX2100609	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004123	Illumina MiSeq	CFSAN	SRP029649	Foodborne Pathogen Survey	SRS477781	124.41	444292	206409106	Salmonella enterica subsp. enterica Nextera shotgun
SRX346847	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004123 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004123	Illumina MiSeq	CFSAN	SRP029649	Foodborne Pathogen Survey	SRS477781	71.04	257363	120403460	Salmonella enterica str. CFSAN004123_01 shotgun
SRX347777	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004145 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004145	Illumina MiSeq	CFSAN	SRP029842	Foodborne Pathogen Survey	SRS478567	83.05	273428	129715546	Salmonella enterica str. CFSAN004145_01 shotgun
SRX347780	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004149 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004149	Illumina MiSeq	CFSAN	SRP029845	Foodborne Pathogen Survey	SRS478570	359.06	1213824	591212320	Salmonella enterica str. CFSAN004149_01 shotgun
SRX347793	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004163 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004163	Illumina MiSeq	CFSAN	SRP029858	Foodborne Pathogen Survey	SRS478583	541.15	1875335	887503185	Salmonella enterica str. CFSAN004163_01 shotgun
SRX347794	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004164 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004164	Illumina MiSeq	CFSAN	SRP029859	Foodborne Pathogen Survey	SRS478584	784.07	2863751	1269884859	Salmonella enterica str. CFSAN004164_01 shotgun
SRX347788	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004157 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004157	Illumina MiSeq	CFSAN	SRP029853	Foodborne Pathogen Survey	SRS478578	316.86	1025252	491531533	Salmonella enterica str. CFSAN004157_01 shotgun
SRX347789	Whole genome shotgun sequencing of Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004158 by Illumina MiSeq	Salmonella enterica subsp. enterica serovar Saintpaul str. CFSAN004158	Illumina MiSeq	CFSAN	SRP029854	Foodborne Pathogen Survey	SRS478579	553.8	1892663	857039632	Salmonella enterica str. CFSAN004158_01 shotgun
SRX1308794	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N48697	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102042	271.34	778236	428146694	CVM N48697
SRX1308795	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N48698	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102041	182.64	505487	279753280	CVM N48698
SRX1308670	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N45926	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102178	165.24	566567	263271961	CVM N45926
SRX1308759	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N47719	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102077	223.55	739674	400461101	CVM N47719
SRX1308892	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N51291	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1101939	202.4	555480	318638206	CVM N51291
SRX1308575	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N43456	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102274	243.35	946595	440544229	CVM N43456
SRX1308649	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N45394	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102199	228.45	755395	347028393	CVM N45394
SRX1308753	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N47713	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102087	188.41	545578	301189978	CVM N47713
SRX1308775	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N48678	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102061	280.52	1029434	486629287	CVM N48678

SRX1308735	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N46846	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102113	152.01	487801	252865600	CVM N46846
SRX1308580	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N43461	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102269	172.49	657711	304185671	CVM N43461
SRX1308813	WGS of Salmonella enterica: 2013 NARMS Ground Turkey: Sample CVM N50423	Salmonella enterica subsp. enterica serovar Saintpaul	Illumina MiSeq	FDA	SRP063697	CVM NARMS Salmonella	SRS1102024	238.02	672070	373323611	CVM N50423

Table 4

List of NCBI isolates for *E. coli* O157:H7 used in the SNP phylogeny in Figure 6.

Experiment Accession	Experiment Title	Organism Name	Instrument	Submitter	Study Accession	Study Title	Sample Accession	Total Size (Mb)	Total Spots	Total Bases	Library Name
SRX1939122	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP074329	GenomeTrakr project: Alaska State Public Health Laboratory-Anchorage	SRS1540222	305.88	995908	482853554	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX2692806	Whole genome shotgun sequencing of Escherichia coli by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP071258	GenomeTrakr Project: California Department of Health - FDLB Micro	SRS2089431	210.9	836696	389166760	Escherichia coli Nextera XT shotgun
SRX3468426		Escherichia coli O157:H7 str. K5607	Illumina MiSeq	edlb-cdc	SRP046387	PulseNet Escherichia coli and Shigella genome sequencing	SRS2756117	514.07	1835187	921263874	NexteraXT
SRX1618666	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP071258	GenomeTrakr Project: California Department of Health - FDLB Micro	SRS1327815	305.73	1108825	506473602	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX1618667	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP071258	GenomeTrakr Project: California Department of Health - FDLB Micro	SRS1327814	92.93	278493	129977631	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX5078675	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP065993	GenomeTrakr Project: New Mexico State University, Food Safety Laboratory	SRS4091848	535.91	2310741	935420977	Nextera XT library SEQ000083419
SRX3030563	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP083072	GenomeTrakr Project: State Hygienic Laboratory at the University of Iowa	SRS2379452	328.21	1015969	496226514	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX2887571	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP058582	GenomeTrakr Project: US Food and Drug Administration	SRS2255590	94.02	368389	157415312	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX2887699	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP071258	GenomeTrakr Project: California Department of Health - FDLB Micro	SRS2255669	299.26	1246502	541675359	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX2887660	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP071258	GenomeTrakr Project: California Department of Health - FDLB Micro	SRS2255632	403.78	1917496	746664042	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX5170917	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP065993	GenomeTrakr Project: New Mexico State University, Food Safety Laboratory	SRS4178551	463.47	1788633	687901627	Nextera XT library SEQ000084924
SRX3851954	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP041719	GenomeTrakr Project: New York State Dept. of Health, Wadsworth Center	SRS3096174	432.99	1534422	731288096	Nextera XT library SEQ000071386
SRX3851959	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP041719	GenomeTrakr Project: New York State Dept. of Health, Wadsworth Center	SRS3096179	460.48	1705683	769911546	Nextera XT library SEQ000071384
SRX1960699	FDA-CFSAN: microbial foodborne pathogen research	Escherichia coli	Illumina MiSeq	FDA/CFSAN	SRP078859	GenomeTrakr Project: FDA-CFSAN MDP Escherichia coli survey from foods	SRS1570718	402.36	1725829	689837948	CFSAN046715
SRX3107169	Whole genome Illumina MiSeq sequence of Escherichia coli serovar O157:H7	Escherichia coli O157:H7	Illumina MiSeq	CFSAN	SRP101500	GenomeTrakr Project: Penn State University | Department of Food Science | Dudley Lab	SRS2442420	185.33	580162	280394130	Nextera XT library SEQ000061613

SRX1298757	Whole genome shotgun sequencing of Escherichia coli by Illumina MiSeq	Escherichia coli	Illumina MiSeq	FDA	SRP061878	GenomeTrakr Project: Texas Department of State Health Services	SRS1096293	246.04	863237	402078517	Escherichia coli Nextera XT shotgun
SRX5170949	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP065993	GenomeTrakr Project: New Mexico State University, Food Safety Laboratory	SRS4178582	377.29	1349576	564907718	Nextera XT library SEQ000084922
SRX5170961	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP065993	GenomeTrakr Project: New Mexico State University, Food Safety Laboratory	SRS4178593	456.94	1601412	694797163	Nextera XT library SEQ000084921
SRX3851958	Whole genome Illumina MiSeq sequence of Escherichia coli	Escherichia coli	Illumina MiSeq	CFSAN	SRP041719	GenomeTrakr Project: New York State Dept. of Health, Wadsworth Center	SRS3096178	485.79	1707926	805854998	Nextera XT library SEQ000071385
SRX1298759	Whole genome shotgun sequencing of Escherichia coli by Illumina MiSeq	Escherichia coli	Illumina MiSeq	FDA	SRP061878	GenomeTrakr Project: Texas Department of State Health Services	SRS1096295	214.89	791684	357157958	Escherichia coli Nextera XT shotgun
SRX3030534	Whole genome shotgun sequencing of Escherichia coli serovar O157:H7 by Illumina MiSeq	Escherichia coli	Illumina MiSeq	CFSAN	SRP083072	GenomeTrakr Project: State Hygienic Laboratory at the University of Iowa	SRS2379423	266.23	857852	418609734	Escherichia coli serovar O157:H7 Nextera XT shotgun
SRX3230830	DNA E. coli O157:H7 from packaged spinach, treated with nalidixic acid	Escherichia coli	Illumina MiSeq	USDA	SRP119076	Escherichia coli Genome sequencing and assembly	SRS2553171	293.17	1119256	503427151	06F00475Nal
SRX3230829	DNA E. coli O157:H7 from packaged spinach	Escherichia coli	Illumina MiSeq	USDA	SRP119076	Escherichia coli Genome sequencing and assembly	SRS2553170	249.11	990713	429509445	06F00475WT
SRX3230832	DNA E. coli O157:H7 from packaged lettuce, treated with nalidixic acid	Escherichia coli	Illumina MiSeq	USDA	SRP119076	Escherichia coli Genome sequencing and assembly	SRS2553173	327.35	1262958	564960667	7386Nal
SRX3230831	DNA E. coli O157:H7 from packaged lettuce	Escherichia coli	Illumina MiSeq	USDA	SRP119076	Escherichia coli Genome sequencing and assembly	SRS2553172	177.66	664150	298522199	7386WT
SRX1960705	FDA-CFSAN: microbial foodborne pathogen research	Escherichia coli	Illumina MiSeq	FDA/CFSAN	SRP078859	GenomeTrakr Project: FDA-CFSAN MDP Escherichia coli survey from foods	SRS1570724	307.6	1299774	486385934	CFSAN046720
SRX3107210	Whole genome Illumina MiSeq sequence of Escherichia coli serovar O157:H7	Escherichia coli O157:H7	Illumina MiSeq	CFSAN	SRP101500	GenomeTrakr Project: Penn State University | Department of Food Science | Dudley Lab	SRS2442461	239.54	747471	356841515	Nextera XT library SEQ000061612

Table 5
qPCR samples and their respective C_T values for detection of *Salmonella enterica* subsp. *enterica* DNA.

Sample[a]	**Inoculum size** (CFU/g)		**Average threshold cycle value** (C_T)[b]
	***S.* Saintpaul**	**EHEC**	
PCR positive control[c]	N/A	N/A	19.059
PCR negative control[d]	N/A	N/A	undetermined
MDA negative control[e]	N/A	N/A	undetermined
Enrichment positive control	10	10	18.286
Enrichment negative control	N/A	N/A	undetermined
EC0.1 SE0.1	0.1	0.1	23.543
EC1 SE1	1	1	18.142
EC0.1 SE1	1	0.1	18.028
EC1 SE0.1	0.1	1	23.515

[a]EC corresponds to *E. coli* and SE corresponds to *S. enterica*
[b]All results (except for EC0.1 SE0.1) are reported as the averages from three biological replicates
[c]Purified *Salmonella* Saintpaul DNA using the Qiagen DNeasy Blood and Tissue Kit
[d]Molecular grade water
[e]TSB used to resuspend bacterial pellets

Table 6

qPCR samples and their respective C_T values for detection of *E. coli* O157:H7 DNA.

Sample[a]	**Inoculum size** (CFU/g)		**Average threshold cycle value** (C_T)[b]
	***S.* Saintpaul**	**EHEC**	
PCR positive control[c]	N/A	N/A	16.131
PCR negative control[d]	N/A	N/A	undetermined
MDA negative control[e]	N/A	N/A	undetermined
Enrichment positive control	10	10	26.628
Enrichment negative control	N/A	N/A	undetermined
EC0.1 SE0.1	0.1	0.1	22.432
EC1 SE1	1	1	23.757
EC0.1 SE1	1	0.1	22.074
EC1 SE0.1	0.1	1	24.321

[a]EC corresponds to *E. coli* and SE corresponds to *S. enterica*

[b]All results (except for EC0.1 SE0.1 inoculum) are reported as the averages from three biological replicates

[c]Purified *E. coli* O157:H7 DNA using the Qiagen DNeasy Blood and Tissue Kit

[d]Molecular grade water

[e]TSB used to resuspend bacterial pellets

Table 7
Sequencing metrics for quasimetagenomic samples.

IMS bead	Sample name[a]	EC inoc. (CFU/g)	SE inoc. (CFU/g)	Output (Mb)	Coverage (%)	Depth ratio	N_{50}	Serotype	SE or EC reads (%)	SE or EC abundance (%)
EC	EC0.1	0.1	N/A	421	68.92	0.527	1068	Undetermined	3.09	11.05
EC	EC1	1	N/A	272	82.12	1.658	2020	O157:H7	9.82	18.67
EC	EC10	10	N/A	365	95.77	7.764	10189	O157:H7	43.71	70.71
SE	SE0.1	N/A	0.1	660	92.94	1.488	3740	Saintpaul	8.57	16.80
SE	SE1	N/A	1	440	97.47	8.521	6072	Saintpaul	51.44	74.91
SE	SE10	N/A	10	567	95.5	3.001	4720	Saintpaul	17.62	29.90
EC	EC1 SE1	1	1	267	97.52	9.760	17395	O157:H7	61.73	90.61
SE	EC1 SE1	1	1	539	89.1	1.169	2492	Saintpaul	6.96	12.64
EC	EC1 SE1	1	1	522	97.92	10.430	14066	O157:H7	60.04	91.07
SE	EC1 SE1	1	1	408	95.72	4.748	5147	Saintpaul	30.46	38.05
EC	EC1 SE1	1	1	428	69.45	0.517	1161	Undetermined	3.67	11.69
SE	EC1 SE1	1	1	502	69.25	0.384	1123	Saintpaul	2.38	5.82
EC	EC1 SE0.1	1	0.1	491	97.51	6.595	14329	O157:H7	37.70	56.72
SE	EC1 SE0.1	1	0.1	574	88.02	0.954	2346	Saintpaul	5.58	10.36
EC	EC1 SE0.1	1	0.1	475	59.47	0.326	1047	Undetermined	2.26	3.59
SE	EC1 SE0.1	1	0.1	530	21.86	0.058	No data	Undetermined	0.36	0.54
EC	EC1 SE0.1	1	0.1	536	94.94	2.505	9842	O157:H7	14.28	31.32
SE	EC1 SE0.1	1	0.1	404	8.77	0.045	583	Undetermined	0.83	1.10
EC	EC0.1 SE1	0.1	1	536	97.44	56.401	15712	O157:H7	35.65	52.90
SE	EC0.1 SE1	0.1	1	382	56.04	0.326	940	Undetermined	2.10	2.72
EC	EC0.1 SE1	0.1	1	460	98.41	10.125	19919	O157:H7	61.13	88.41
SE	EC0.1 SE1	0.1	1	578	96.43	4.768	5076	Saintpaul	28.07	37.62
EC	EC0.1 SE1	0.1	1	508	98.13	8.494	17494	O157:H7	48.76	74.93
SE	EC0.1 SE1	0.1	1	325	84.11	1.272	2023	Saintpaul	8.43	12.57
EC	EC0.1 SE0.1	0.1	0.1	357	46.16	0.251	841	Undetermined	1.73	3.64
SE	EC0.1 SE0.1	0.1	0.1	371	15.81	0.055	617	Undetermined	0.43	0.57
EC	EC0.1 SE0.1	0.1	0.1	109	82.79	5.538	2066	O157:H7	32.35	62.96
SE	EC0.1 SE0.1	0.1	0.1	271	4.77	0.024	545	Undetermined	0.21	0.42

[a]All samples were completed in triplicate except for EC0.1 SE0.1 samples, which were completed in duplicate

Table 8
SNP distance and loci for *Salmonella enterica*-targeted quasimetagenomic samples.

Sample#[a]	Quasimetagenomic sample	SNP distance	SNP loci
1	EC1 SE0.1	0	553
2	EC1 SE0.1	0	0
3	EC1 SE0.1	0	0
4	EC1 SE1	0	584
5	EC1 SE1	0	758
6	EC1 SE1	0	192
7	EC0.1 SE1	0	796
8	EC0.1 SE1	0	121
9	EC0.1 SE1	0	448
10	EC0.1 SE0.1	0	0
11	EC0.1 SE0.1	0	0
12	SE0.1	0	694
13	SE10	0	779
14	SE1	0	831

[a]Sample# refers to those listed in Figure 5

Table 9
SNP distance and loci for *E. coli*-targeted quasimetagenomic samples.

Sample#[a]	Quasimetagenomic sample	SNP distance	SNP loci
1	EC0.1	0	535
2	EC1	0	993
3	EC10	1	1696
4	EC0.1 SE0.1	0	1126
5	EC1 SE1	0	1730
6	EC0.1 SE1	0	1746
7	EC1 SE0.1	1	1721
8	EC1 SE1	0	1717
9	EC0.1 SE1	1	1729
10	EC0.1 SE1	0	1759
11	EC1 SE0.1	2	1669
12	EC0.1 SE0.1	0	275
13	EC1 SE0.1	0	602
14	EC1 SE1	0	823

[a]Sample# refers to those listed in Figure 6

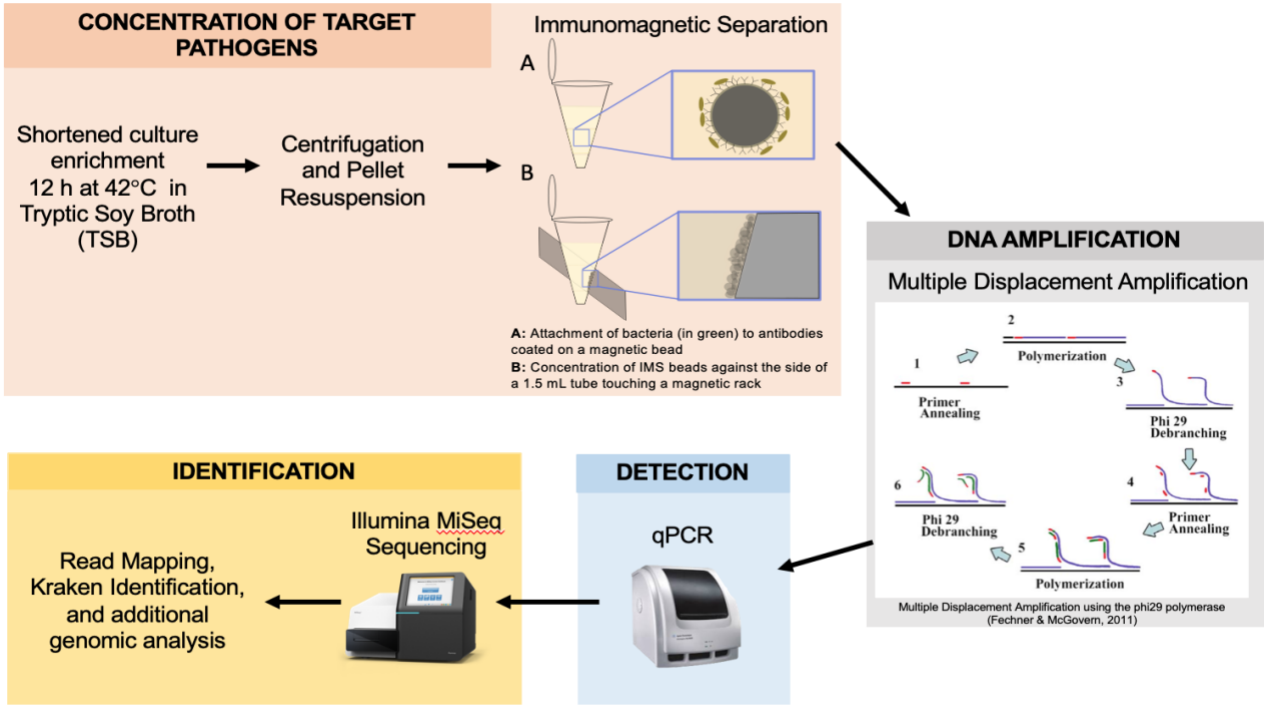

Figure 1. Overview of the quasimetagenomics sequencing workflow.

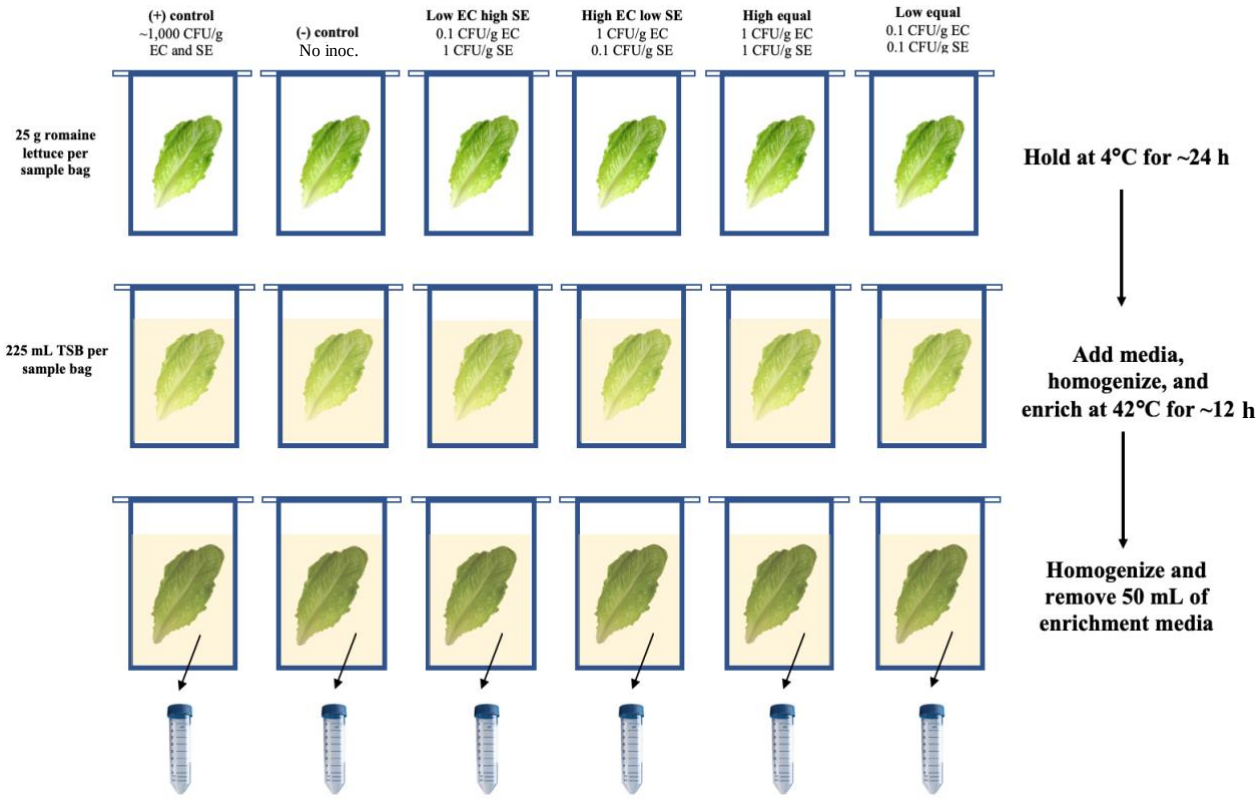

Figure 2. Overview of a single biological replicate enrichment with six varied, mixed inocula.

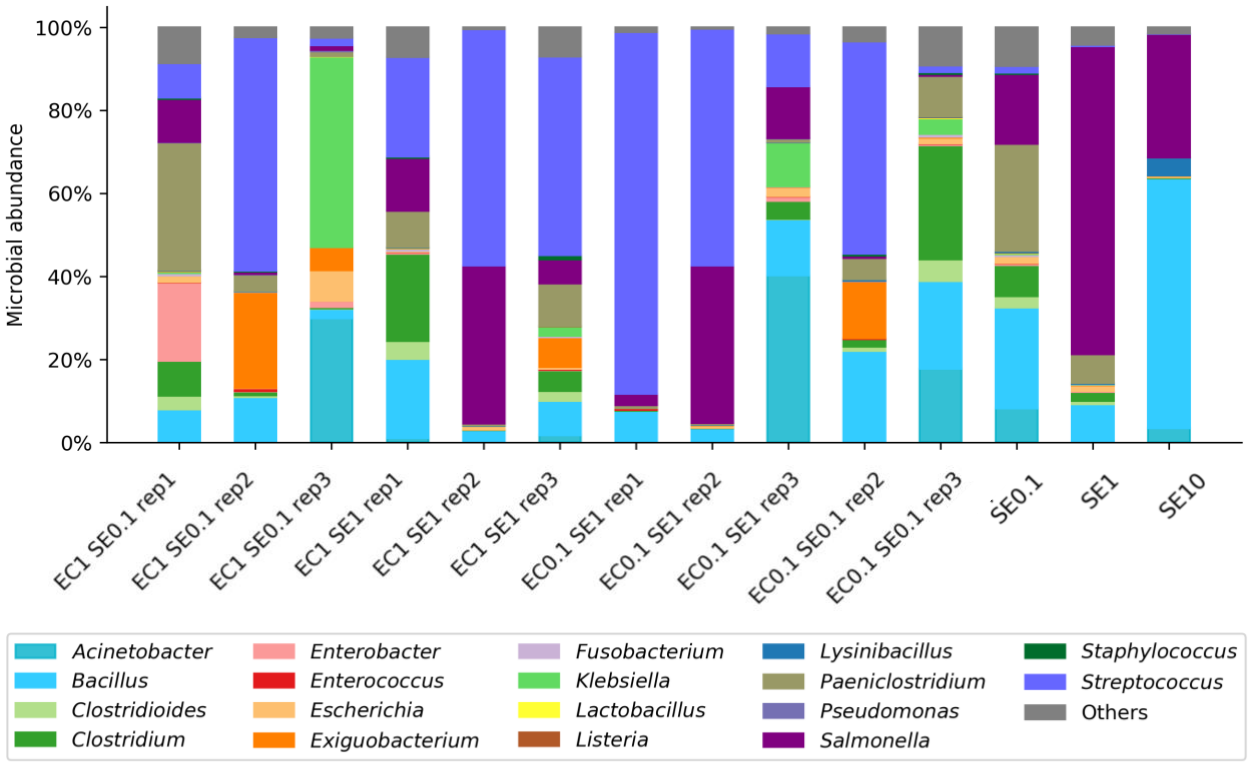

Figure 3. Stacked bar chart displaying the percent microbial abundance for each *Salmonella*-targeted quasimetagenomic sample.

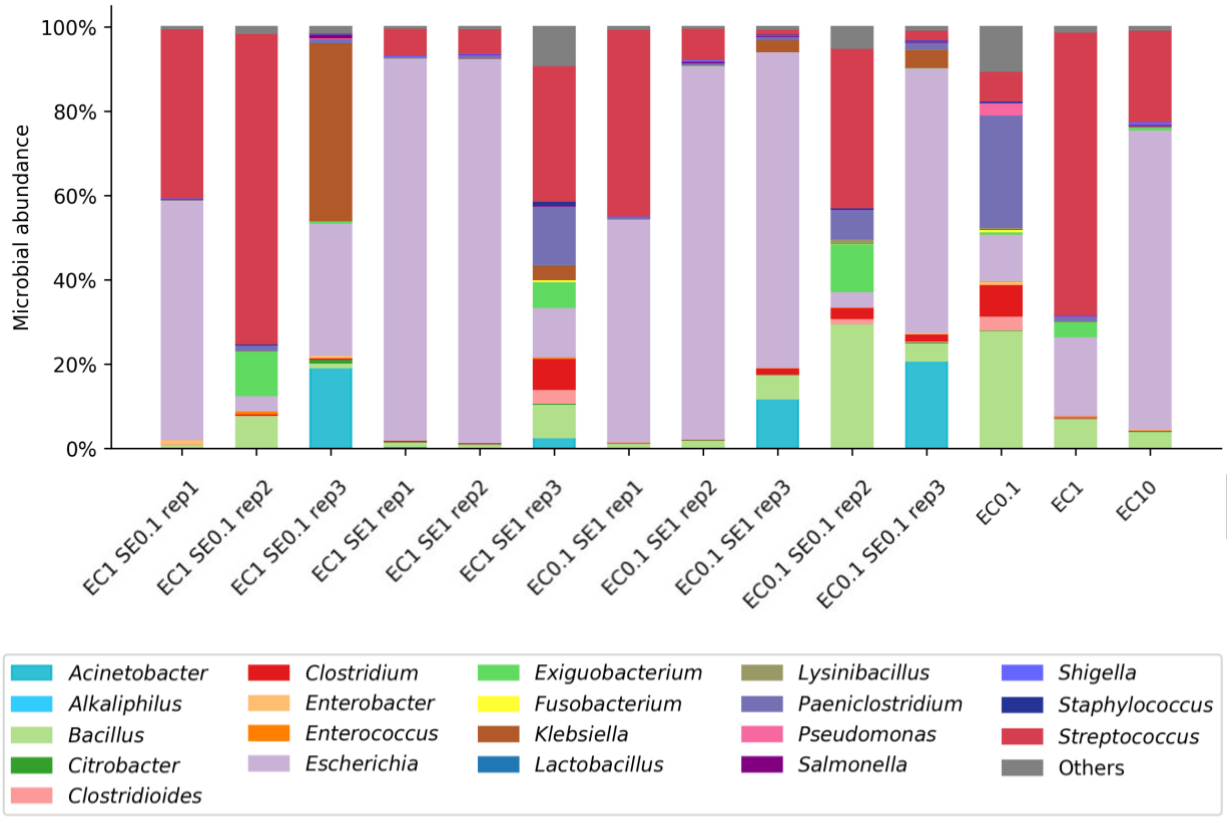

Figure 4. Stacked bar chart displaying the percent microbial abundance for each *E. coli* O157:H7-targeted quasimetagenomic sample.

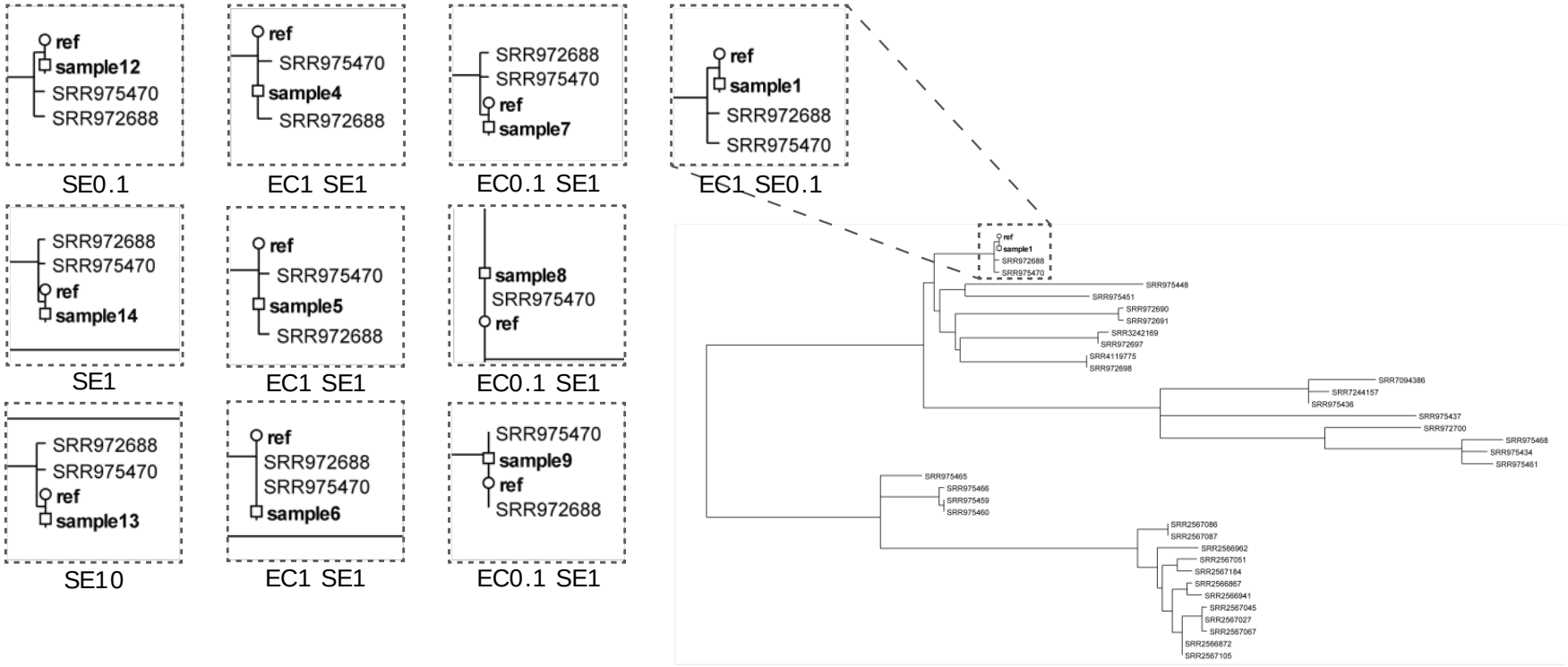

Figure 5. Phylogenetic clustering of *Salmonella*-targeted quasimetagenomic samples (sample#), *Salmonella* Saintpaul isolates, and reference strain (ref) used for sample inoculation. Samples SRR975470 and SRR972688 are *Salmonella* Saintpaul strains from the 2008 jalapeno-associated outbreak, which is also the reference strain. Each box contains part of the SNP-phylogeny for particular quasimetagenomic samples of varying inocula.

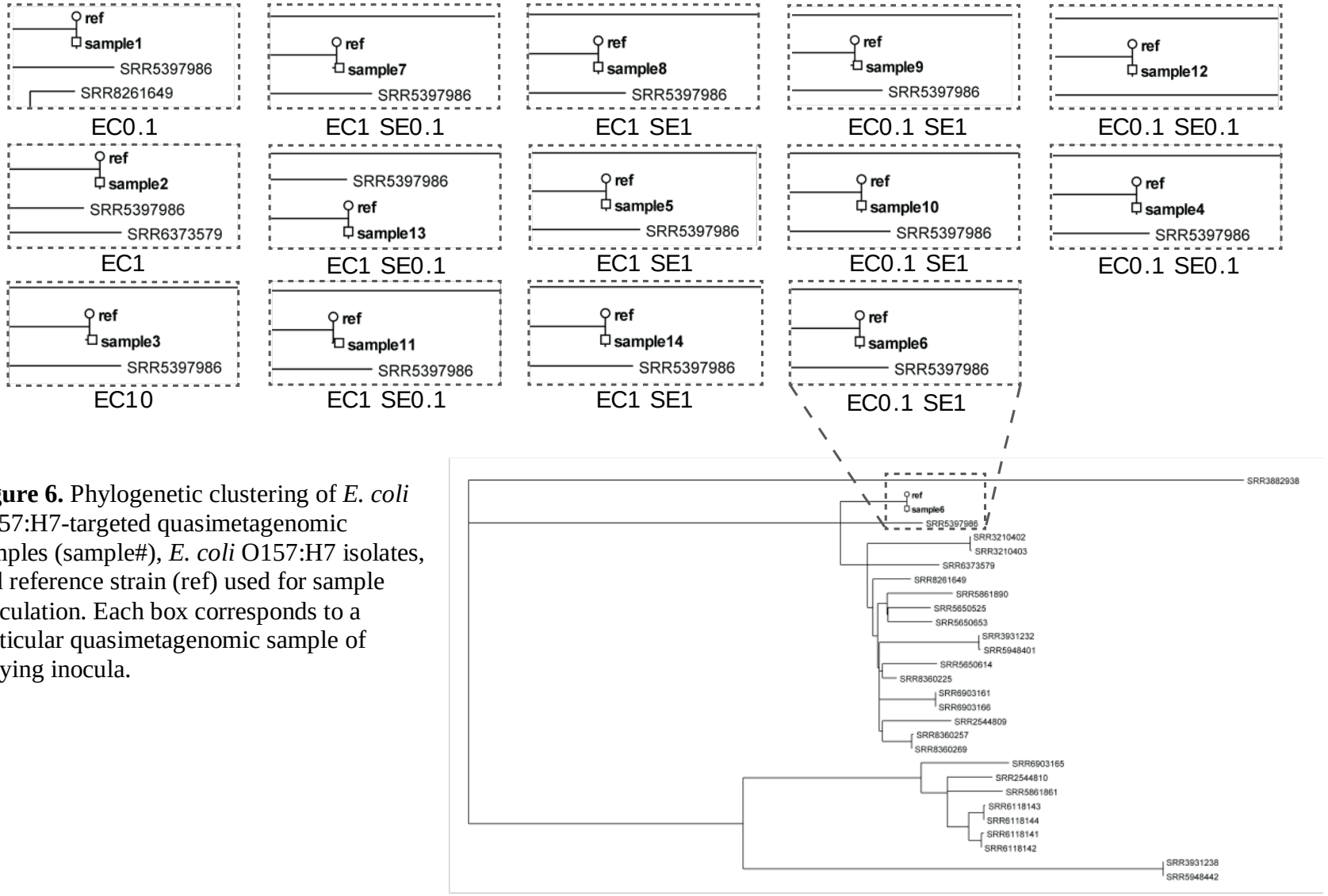

Figure 6. Phylogenetic clustering of *E. coli* O157:H7-targeted quasimetagenomic samples (sample#), *E. coli* O157:H7 isolates, and reference strain (ref) used for sample inoculation. Each box corresponds to a particular quasimetagenomic sample of varying inocula.