EVALUATION OF AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) ANALYSIS AS A METHOD FOR TRACKING *SALMONELLA ENTERICA* IN A FOODBORNE OUTBREAK

By

STEPHANIE KUNZ

Bachelor of Science in Investigative and Medical Sciences

Saint Louis University

Saint Louis, Missouri

2011

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December 2013

EVALUATION OF THE DISCRIMINATORY POWER OF AMPLIFIED FRAGMENT LENGTH POLYMORPHISM ANALYSIS METHOD FOR TRACKING *SALMONELLA ENTERICA* IN A FOODBORNE OUTBREAK

Thesis Approved:

Dr. Robert Allen

Thesis Adviser

Dr. Franklin R. Champlin

Dr. R. Thomas Glass

Name: STEPHANIE KUNZ

Date of Degree: DECEMBER 2013

Title of Study: EVALUATION OF AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) ANALYSIS AS A METHOD FOR TRACKING *SALMONELLA ENTERICA* IN A FOODBORNE OUTBREAK

Major Field: FORENSIC SCIENCE

Abstract: Foodborne disease causes illness and death in people around the world and costs nations millions of dollars and many lives. Currently in US, pulsed field gel electrophoresis (PFGE) is used to monitor and track foodborne outbreaks. However, it lacks the discriminatory power to identify two different sources of the same strain of *Salmonella enterica.* This study evaluated the discrimination, reproducibility, and accuracy of an alternative method, amplified fragment length polymorphism (AFLP), in conjunction with a simple electropherogram coding system. Comparing the codes prepared in triplicate of six clinical samples of *S. enterica* from the Oklahoma Department of Health to each other results were less reproducible and discriminatory than previous AFLP studies done on organisms in the *Serratia* and *Pseuodomonas* genera. When codes prepared in lab were compared to codes prepared *in silico* as a measure of accuracy, the codes were only 12-23% similar. Because of the low accuracy, reproducibility, and discrimination of AFLP when applied to *S. enterica*, AFLP in conjunction with the AFLP haplotype system are not suitable for tracking foodborne outbreaks of salmonellosis.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Foodborne illness comes at a great human and economic cost to the United States and the rest of the world. In fact, the Centers for Disease Control (CDC) estimates that 48 million Americans faced symptoms of foodborne illness within the last year, with most of them caused by bacteria including *Escherichia coli*, *Shigella*, and *Salmonella*[1,2]

Salmonellosis affects humans worldwide and can cause individuals without proper treatment to die of dehydration.[3,4] In addition to outbreaks caused by coincidental food contamination, scientists documented use of *Salmonella* in a terrorist plot within the United States.[5] In addition to encouraging proper washing of food at the site of production as well as consumption and proper hand washing to avert disease, the CDC works hard to prevent spreading of existing instances of the disease.[6]

To deter both accidental and intentional foodborne illness outbreaks from spreading, the CDC and other public health agencies monitor instances of these diseases using a combination of assays. Public health labs in the United States and other modernized nations most commonly utilize a technique called pulsed field gel electrophoresis (PFGE) to genetically type organisms potentially causing the outbreaks collected at infected sites. PFGE provides

useful information to a CDC-monitored network called PulseNet which is used to track these outbreaks. However, the process of obtaining and standardizing results requires more time and expertise than available in a particularly pernicious outbreak. More specifically relevant to foodborne outbreaks, PFGE has proven to be an insufficient technique for genotyping samples for determination of strain and source of *Salmonella*, the leading cause of foodborne illness in the United States.[7,8] Other methods, specifically amplified fragment length polymorphism (AFLP) analysis, have been suggested to provide a greater measure of discrimination between species and/or strains using resources and technology common in many public health labs.

By combining knowledge of restriction fragment length polymorphism (RFLP) analysis with the reproducibility and power of polymerase chain reaction (PCR) and capillary electrophoresis (CE), profiles created through AFLP have been shown to be distinct within the genera in *Pseudomonas* and *Serratia*. These profiles allowed for grouping of organisms by species or strain as well as discriminating between sources of the same strain.[9,10] Thus, AFLP holds potential to be a more robust genotyping method than PFGE and might allow agencies to more effectively investigate outbreaks of foodborne illness occurring naturally or as a result of criminal activity.

Previous studies utilizing AFLP were successful in distinguishing species and strains of bacteria that attack a wide array of organisms including plants and animals. Unfortunately, not enough data has been collected to identify *Salmonella* spp. to this level using AFLP alone. Utilizing multiple samples of three strains of *Salmonella* (as identified using API-20E panels and serotyping) at the Oklahoma Department of Health, comparisons were made to reveal the discriminatory power of a codified AFLP procedure in comparison with PFGE.[8] While previous studies using AFLP have utilized graphics software to compare data, demonstrating the reproducibility and discriminatory power (DP) using a more simple numeric code to compare AFLP profiles proved a simpler means to compare samples.[11,12] Thus, *Salmonella* strains were

"grouped" using PFGE and AFLP by the Oklahoma Department of Health and Oklahoma State University Center for Health Sciences respectively. By providing reliable data in an easily comparable format, AFLP may be a useful and discriminatory means of tracking the spread of intentional or unintentional foodborne illness.

Because the assay utilizes restriction enzymes to digest genomes, single nucleotide differences can result in a differing pattern of restriction fragment sizes, thereby altering a haplotype code. In addition to producing AFLP profiles in the laboratory, the genomic sequences of strains analyzed in the laboratory were analyzed using the GenBank repository information and bioinformatics software to produce the AFLP profile and haplotype code *in silico*.

The purpose of this study is to assess the power of AFLP and the AFLP haplotype coding system as a means of determining serotype and source of *Salmonella enterica* samples using indications of reproducibility, discriminatory power, and accuracy to do so.

CHAPTER II

LITERATURE REVIEW

Investigations of disease outbreak, by the Centers of Disease Control and Prevention (CDC) in the United States and the United Nations World Health Organization (WHO) internationally, prevent infectious illnesses from spreading within populations and eliminate sources of harmful pathogens. Though the CDC and other agencies worldwide have set up infrastructure for reporting of disease and source determination, the technique most commonly applied today, namely pulsed field gel electrophoresis (PFGE), lacks the discrimination to confidently determine species and source of *Salmonella* samples from outbreaks. While this may seem like a small caveat in the successful application of this technique, *Salmonella* accounts for most of the foodborne disease outbreaks within recent years in this nation. Without the application of more discriminatory methods, *Salmonella* outbreaks will cost Americans and world citizens not only their health, but also a considerable cost for treatment of those who contract salmonellosis.

Foodborne Disease

Individuals can contract many diseases through ingestion of contaminated food. Though respiratory diseases are more common, foodborne illnesses cause major illness in American citizens as well as the rest of the world. Outbreaks of foodborne illnesses come at a cost to not only the individuals who contract the disease but also the public health agency that investigates the outbreak. The CDC estimated that 48 million Americans experienced symptoms of foodborne illness often caused by bacteria including *Escherichia coli, Shigella*, and most commonly various species of *Salmonella* in 2010. Individuals and companies involved in processing the food prior

to ingestion are also negatively affected. In fact, species of *Salmonella* accounted for 10 of the 16 major foodborne illness outbreaks investigated by the CDC last year.[1,13]

Foodborne Illness Outbreak Detection

In America, foodborne outbreaks are detected through local public health offices that record the number of cases reported by clinicians. These numbers and the results of various tests are then passed on to the CDC so that the agency can monitor potential nationwide outbreaks. As food production has become centralized in this nation and many others, it has become more important for surveillance of foodborne outbreaks to happen on a national scale as contaminated products may be shipped between coasts in the matter of a few hours.[13] Ultimately, the CDC will deploy their teams of epidemiology and microbiology experts to investigate and attempt to contain outbreaks, minimizing damage.

On an international scale, the WHO's Global Foodborne Infection Network (GFN) provides information regarding outbreaks often limited by the resources and knowledge available to the nation of outbreak. To bridge this gap the GFN offers courses to member labs worldwide in basic microbiological as well as molecular techniques such as PCR and PFGE.[14] The WHO recognizes the disparities in testing abilities between nations and includes both classical microbiological susceptibility testing as well as molecular testing as part of their guidelines for investigating food disease outbreaks. However, the *WHO's Foodborne Disease Outbreaks: Guidelines for Investigation* lists PFGE and genetic sequencing as appropriate measures of surveillance when available.[15] Neither the training modules nor guidelines list AFLP as a method for surveillance.

Salmonellosis

Salmonellosis affects many individuals within developed and undeveloped nations, with the United States alone reporting thousands of the millions of cases reported from around the

globe to the World Health Organization (WHO).[3,4] The disease occurs due to ingestion of *Salmonella* species that normally reside in the intestines of humans and other animals as well as in the fecal material of animals. This bacterium often becomes associated with produce like mangoes, bean sprouts, or spinach, either out in the field or during processing. Humans then often ingest this raw, insufficiently washed produce, thereby ingesting *Salmonella*, and contracting salmonellosis.[16] After exposure to the bacteria, usually from ingestions of miniscule amounts of fecal material on food, the bacteria travel through the gastrointestinal system until it reaches the intestines. The bacteria then colonize the intestine and translocate into the intestinal epithelium, where they induce inflammation by releasing endotoxins after death. The endotoxins most commonly causing vomiting, diarrhea, cramps, and fevers in an infected individual which can lead to life threatening dehydration.[17]

Taxonomy of *Salmonella*

To understand how well laboratory techniques can distinguish between similar bacteria, the levels of organization of these organisms must be understood. The levels of organization of *Salmonella* as in most organisms are best defined using taxonomic approaches. Unfortunately, the taxonomy of the genus *Salmonella* changes often and different regulatory groups do not accept the same classifications. The CDC, the Judicial Commission of the International Committee on the Systematics of Prokaryotes, and the WHO all agree that *Salmonella* is a genus of the family Enterobacteriaceae. By the mid-20th century scientists identified different types of *Salmonella* using serological testing of O antigens on the cells' surface and H antigens on the cells' flagellae as a means of differentiation. Though scientists today know designations based upon these differences as serovars, scientists initially studying *Salmonella* named each serovar a different species. This system developed by Kauffmann is often known as the "one serotype one species" system.[18] Microbiologists named *Salmonella* species sometimes after location of origin, resultant disease, or antigens present on the organism.

In the late 1970s to early 1980s, due to the similarities among all species revealed through DNA hybridization tests, all *Salmonellae* were grouped in a single species, *Salmonella choleraesuis*, with the exception of a single out-group subspecies made of a single serovar *Salmonella choleraesuis* subspecies *bongori*, which was separated into its own species, *Salmonella bongori*.[19] In 1986, following a publication by Le Minor *et al.* [20] that discerned subspecies by DNA hybridization and phenotypic characteristics, the CDC accepted a proposal to separate the serovars into seven subspecies by Le Minor's standards and to change the name of *Salmonella choleraesuis* to *Salmonella enterica* to eliminate confusion regarding the previously named *Salmonella choleraesuis* serovar Choleraesuis.[20,21] However, the Judicial Commission of the International Committee of Systematic Bacteriology did not accept *Salmonella enterica* as the type species until 2002. Currently the CDC, the American Microbiological Association, and the WHO recognize two species of *Salmonella: Salmonella bongori* and *Salmonella enterica.* (Figure1)

S.enterica currently consists of six subspecies as originally determined by Le Minor: *enterica* or I, *salame* or II, *arizonae* or IIIa, *diarizonae* or IIIb, *houtenae* or IV, and *indica* VI. In the subspecies *enterica* as well as any other subspecies that named serovars prior to 1966, serovars retained the geographic or disease related names originally given to them. All other subspecies of *Salmonella enterica* as well as serovars within *Salmonella bongori* that lacked names prior to 1966 use the Kauffmann-White scheme which designates the serotype using the formula (O antigen): (H1 antigen):(H2 antigen).[21,22] As more studies regarding the common genetic nature of *Salmonella* serotypes continue, the debate regarding nomenclature of *Salmonella* develops as a subject for debate as it has for decades.

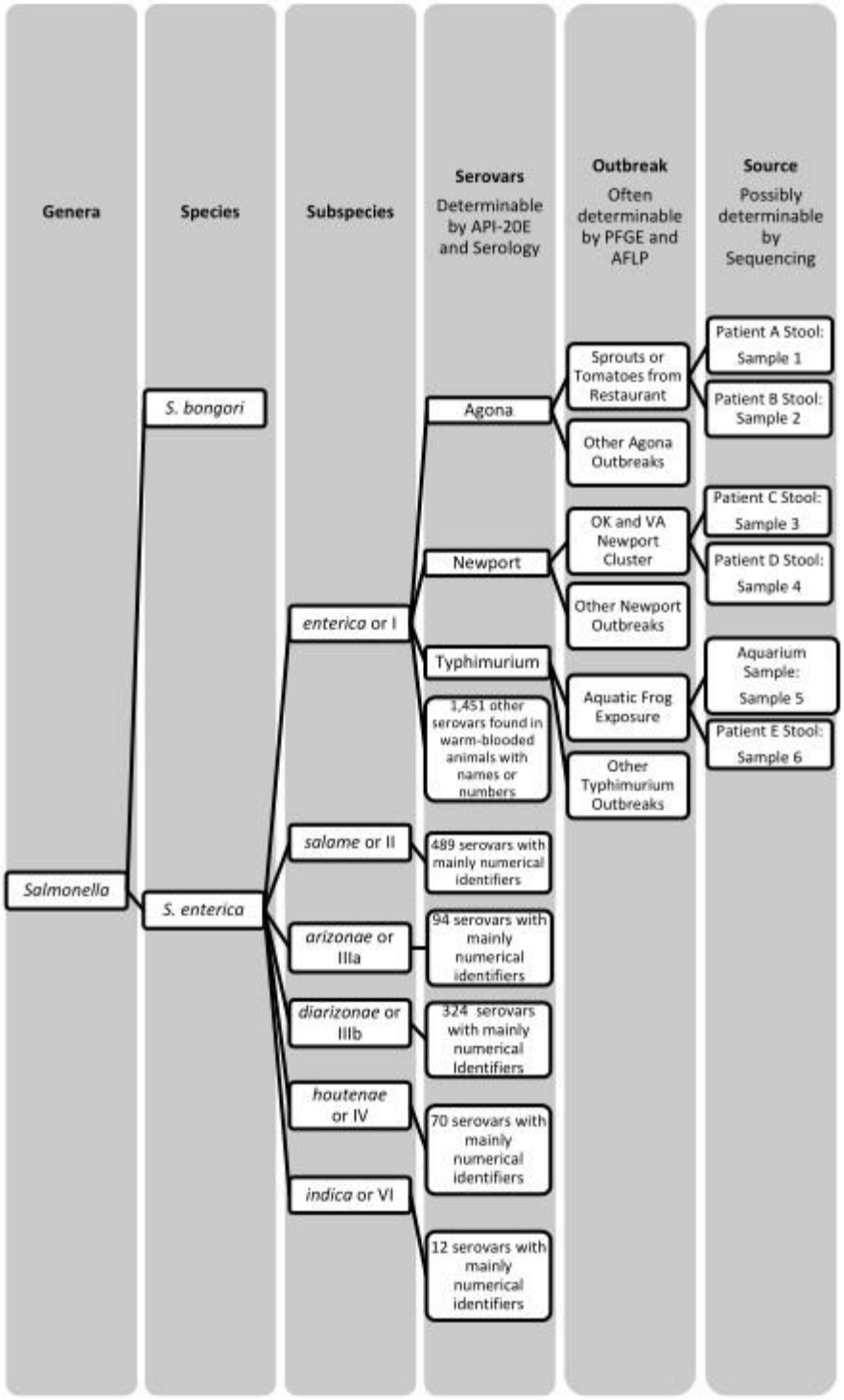


Figure 1: Current CDC accepted taxonomy of *Salmonella* and assays used to discriminate between taxa.

Effects of Salmonellosis

In most individuals salmonellosis causes severe dehydration with vomiting, diarrhea, and cramps and in immunocompromised individuals these symptoms can cause death. Due to the high level of available health care, the United States experiences very few salmonellosis related deaths with only 13 deaths reported in 2008, though global estimates put the yearly mortality at 155,000.[23,24] In addition to human loss, nations also face the economic burden of salmonellosis; citizens of the United States, for instance, spent over $2.7 million combined or $1,938 on average for each case in 2010 alone.[25] One study suggests that accounting for medical treatment, quality of life costs including pain and suffering, and mortality, the average case of non typhoidal *Salmonella* costs $11,086.[1]

Terrorist Use of *Salmonella*

Though aerosolized diseases like SARS and influenza pose the greatest threat to a population, *Salmonella* has been used in a terrorist application before. In 1984 after patients ate at local restaurants, The Dalles, Oregon experienced a rash of 751 cases of Salmonellosis. Local public health officials investigated the outbreak and found that all infected individuals had eaten at local salad bars. The local public health agency called in the CDC to investigate the outbreak. Through antibiotic susceptibility testing and serological tests, the CDC compared clinical samples to samples from salad bars. They found that a local cult known as the Rajneeshee had cultured *Salmonella* Typhimurium and poured samples of this culture onto salad bars and into coffee creamers. According to testimony by cult members they intended to keep individuals within the area from voting in an upcoming election.[5]

Public health offices and the CDC have since changed many surveillance, monitoring, and detection systems in response to this outbreak and the 2001 terrorist attacks. However, even

with the genotyping technologies and the systems employed in labs currently, an outbreak of Salmonellosis could still be difficult to type and track epidemiologically.

Current Outbreak Surveillance

Like all other foodborne illnesses, the CDC, in cooperation with state and local public health offices, tracks cases of foodborne illnesses to detect outbreaks, not only to eliminate natural occurrences of salmonellosis but also to monitor for possible bioterrorism applications.[5] In clinical settings, laboratorians utilize a combination of chemical tests such as the API 20E strip to discriminate between bacterial species in order to provide appropriate treatments.[26] These assays often have issues differentiating species of Enterobacteriaceae, making API 20E inappropriate for use in epidemiological analysis of salmonellosis or other Enterobacteriaceae-caused illnesses.[27]

Public health labs utilize various molecular techniques to group cases and determine sources with pulsed field gel electrophoresis (PFGE) being the most popular. PFGE utilizes currents of varying direction across a gel to separate restriction enzyme digested genomic DNA to create restriction fragment profiles visualized through staining.[28,29] The CDC collects PFGE results utilizing a successful computerized system called PulseNet to make them available to public health labs to help determine the source of contamination and eliminate it.[30] However, obtaining and processing PFGE gels for this purpose consumes a large amount of time and requires a high level of technical expertise.[31,32]

Alternative Methods for Genotyping

Newer methods for microbial genotyping, such as AFLP and multi-locus variable number of tandem repeats analysis (MLVT), are often applied in research to discriminate bacteria. These methods often offer less complication and higher levels of discrimination.[7,32] However serologic or physiologic tests, while often informative, have not been the national public health surveillance

system.[7,32] While MLVT requires knowledge of the genetic code of an organism to target housekeeping genes for sequencing, AFLP requires no knowledge regarding the genome of a microbe as it utilizes anonymous of restriction fragments of varying sizes to differentiate and group samples.[33] This characteristic makes AFLP a better fit for epidemiological surveillance because no knowledge of even the bacteria's species is necessary to perform the test.

Amplified Fragment Length Polymorphism

In foodborne illness outbreaks, quick and specific identification of bacteria is critical to early control. Whether the outbreak is food, hospital, or terrorist related, learning more about the organism that causes the outbreak is the best way to find the source, treat, and stop the disease from infecting more individuals. Very few analytical techniques are capable of producing quick, reliable, and discriminatory data to allow this. As microbial forensics gets more attention with the creation of the Department of Homeland Security and other agencies that monitor food outbreaks, it is imperative that labs be provided with a fast, reliable, discriminatory method. In addition, public health labs will require personnel and equipment in order to implement the method.

AFLP is a genetic analysis technique that can provide quick differentiation and characterization of bacteria in an outbreak whether natural or terrorist in nature. AFLP ultimately produces a DNA profile utilizing a combination of the sequence variability in the genome detected through RFLP analysis with the reliability and speed of PCR in order to produce DNA fingerprints that are highly specific for each sample. The process utilizes equipment found in many biological labs and relatively low cost reagents.

In this application of AFLP, *MseI* and *EcoRI* restriction enzymes digest genomic DNA extracted from bacterial cultures. Oligonucleotides corresponding to the restricted ends of genomic fragments are ligated to the digested fragments to provide target sites for PCR primers. Once all restricted fragments are ligated and non-selectively amplified, a second round of

amplification occurs. This round selectively targets fragments with a nucleotide or two following the primer binding site, thereby reducing the number of fragments amplified and subsequently analyzed. Using capillary electrophoresis the samples are genetically analyzed creating an electropherogram that can be converted into a numerical haplotype code that enables easy comparison between samples.

Previous Applications of AFLP to *Salmonella* Outbreaks

From 2006-2009 scientists at the University of Iowa collected *Salmonella* isolates from the mesentery and feces of infected pigs on 24 farms. The study found that fecal samples could be matched to each farm based upon AFLP results, while serotyping and MLVT were unable to do so. However, *Salmonella* samples from mesenteric lymph nodes could not be differentiated by farm, likely due to comingling in lairage pens.[34] Clinicians in Italy utilized a single enzyme variant of AFLP (SE-AFLP) to analyze three cases of *Salmonella enterica* serovar Rissen in comparison to separate outbreaks of the Rissen serovar for epidemiological purposes. They concluded that all three cases resulted from the same strain and were probably contracted nosocomially.

Sources of AFLP Error

Because of the long process of AFLP and somewhat blind approach to the genome tested, errors in the AFLP process are both inherent and difficult to pinpoint. Using the haplotype code system eliminates subjective scoring of restriction fragments of low intensity by analysts, a major source of error identified in previous studies.[11] However, differences and peak intensities between repetitions and slight shifts between peaks, also identified as a source of error, can affect the haplotype code obtained by shifting the scored peaks between bins or eliminating a peak due to a fluorescence just below the threshold.[35] Additionally AFLP faces issues of homoplasy, where multiple fragments of identical size, yet from different genomic sources can mask each another

by appearing as a single peak within an electropherogram instead of multiple peaks of identical sizes. This occurs often in the smaller fragment size ranges.[36]

AFLP *in silico*

With sequencing becoming more commonplace and agencies like the National Center for Biotechnology Information (NCBI) willing to share the information gained through research, scientists have access to a great variety of entire genomic sequences. Using this information in conjunction with computer software, scientists can predict and analyze AFLP results *in silico* and compare them with AFLP profiles produced in the laboratory.

Most recently Zhang used a program he created himself called AFLP Runner (AR), a modified KMP algorithm to search for the specific restriction sequence in the string of nucleotides, however, this method produced codes at only 88% accuracy and is not available for download.[37] A group at a French university also developed a homegrown analysis program available for download, but published no accuracy information.[38] Other options for AFLP *in silico* analysis include a web hosted program that creators encourage users to use for free up to five times which has no published accuracy values, and DNASTAR's Lasergene (DNASTAR, Inc., Madison, WI), a costly genomic sequence and protein analysis software suite that stands as the most frequently cited sequence analysis software and has produced AFLP in silico results at 99% accuracy.[39,40]

Summary

While the CDC and other agencies mainly utilize PFGE for the typing of clinical isolates and epidemiological tracking of *Salmonella*, the discriminatory powers of AFLP technology could provide epidemiologists with even more useful information that would enable them to better track the source of a disease and stop it from spreading. By acquiring information regarding the source, strain, and species of a pathogen earlier through the use of AFLP, the CDC and other

public health agencies can more effectively investigate and stem the spread of foodborne illness. However, very little information regarding AFLP's accuracy has been generated despite its known flaws. Using in silico methods to produce AFLP codes from sequenced samples can provide a means for evaluating AFLP codes produced in the laboratory against a benchmark target to establish accuracy and reliability of the method. Adding to the list of available effective testing methods can greatly assist health agencies track and half the spread of foodborne illness.

CHAPTER III

METHODOLOGY

Sample Collection and Identification

Samples of suspected *Salmonella* cases were sent to Oklahoma's public health lab for characterization as required by the Oklahoma state law.[41] The Oklahoma Public Health lab then identified the isolates using Biomeriux's API 20E testing and serological H and O antigen agglutination tests using antiserum from Denka Seiken and Difco.[8] Public health lab professionals then used the Kauffman-White Scheme to link H and O antigen types to a specific serotype name. PFGE was run on each of the samples, and each profile was uploaded to PulseNet for outbreak tracking by the CDC.[8]

Bacterial Cultures

Six samples of *Salmonella*, as listed in Table 1, were cultured and isolated on a Brain-Heart Infusion plate agar. The sample numbers were changed to deidentify patients associated with samples. Then a single colony of each sample was cultured in a 2.0 mL aliquot of Mueller-Hinton broth in a 37˚ C incubator overnight until growth was visible. Liquid cultures were centrifuged at 10,000 XG for 3 minutes at room temperature to create a cell pellet from which DNA could be extracted.

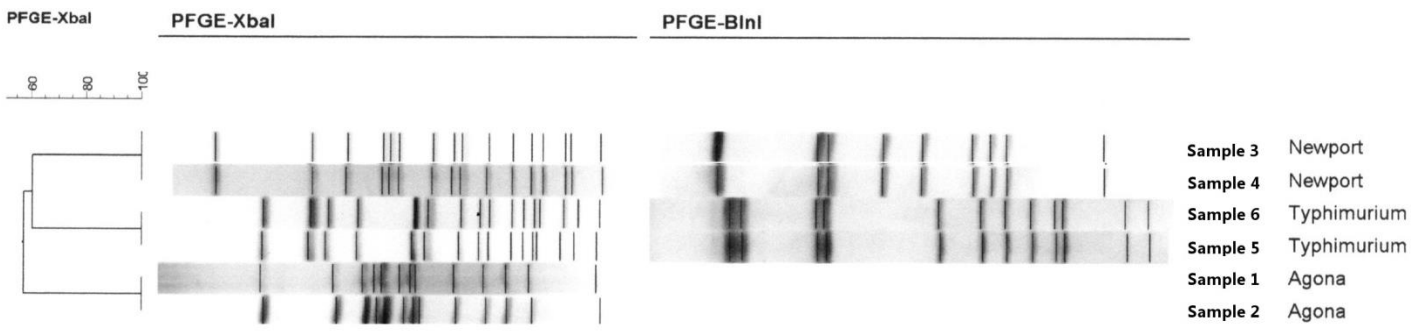
Table 1: Sample numbers, serotypes, and sources of *S. enterica* cultured for AFLP analysis.

Sample Number	Serotype	Outbreak	Source
1	Agona	Restaurant Sprouts/Tomatoes	Patient A Stool
2	Agona	Restaurant Sprouts/Tomatoes	Patient B Stool
3	Newport	Unknown OK & VA cluster	Patient C Stool
4	Newport	Unknown OK & VA cluster	Patient D Stool
5	Typhimurium	Aquatic Frog Exposure	Aquarium Sample
6	Typhimurium	Aquatic Frog Exposure	Patient E Stool

DNA Extraction

Cell pellets were resuspended in 250 µL of TNE (10 mM TRIS-Cl pH8.0 with 0.2 M NaCl and 1.0 mM EDTA) mixed with 15 mg/µL lysozyme and incubated at 37˚ C for an hour. A single mL of 2X DNA isolation solution was made with 910 µL TNE, 10µL RNase A (5 mg/ml), 40 µL Proteinase K (20 mg/mL in 10 mM TRIS-Cl pH 8.0 with 0.2 M KCl & 50% v/v Glycerol)

Figure 2. PFGE imaging of AFLP-tested organisms conducted by Oklahoma Public Health Laboratory.



and 50 μL 20% sodium dodecyl sulfate (SDS). 250 μL of the 2X isolation solution was added to each cell suspension and incubated for an hour at 65˚ C. An equal volume of phenol: chloroform: isoamyl alcohol mixture (9:0.96:0.4) was added to each sample and thoroughly mixed using vortexing. The samples were centrifuged for 3 minutes for 10,000 XG. The upper aqueous phase was collected by pipetting and deposited into new microcentrifuge tubes. To each sample an equal volume of chloroform: isoamyl alcohol (24:1 v/v) was added, mixed using vortex isolation for 10 seconds, and centrifuged for 10,000 XG for 3 min. Again, the upper phase was collected by pipetting and deposited into microcentrifuge tube with 2.0 volumes of 95% ethanol. The tubes were then gently inverted five times and DNA clots were removed using an inoculating loop and resuspended in 100-200 μL of TE^{-4}.

DNA Quantitation

The extracted DNA was quantified in 2.0 μL of the sample using a NanoDrop spectrophotometer (ND-1000) (NanoDrop Products). Spectrophotometric measurements with 260:280 ratio of at least 1.8, a reasonably purity, were kept and their concentrations recorded.

Enzyme Digestion

DNA was digested using 1 μL of 10X New England Biolabs Buffer 4 (New England Biolabs, Inc., MA), 1.0 μL of 10 U/μL *MseI* (New England Biolabs, Inc.), 5.5 μL of dH2O, and 2.5 μL of isolated DNA (adjusted depending on the initial concentration by dilution to a total amount of 300-500 ng of DNA) and added to a clean microcentrifuge tube for a total volume of 10 μL Samples were incubated at 37˚ C for one hour and then 65˚ for 5.0 minutes to denature the enzymes. Samples were then put on ice for 5.0 minutes to ensure that the reaction had stopped. To the first digestions, 1 μL containing 20 U of EcoR1, 2 μL of 10X EcoR1 buffer and 7 μL of H_2O were added. The samples were then incubated in a heat block at 37˚ C for one hour, then at 65˚ C for 5.0 minutes before placing on ice for 5.0 minutes.

To confirm a successful digestion, an aliquot of each sample was separated and analyzed on a 1% agarose electrophoresis gel in TAE buffer (10 mM Tris-acetate pH 8.3 with 1 mM EDTA).7.0 μL of each digested sample were mixed with 3.0 μL of 1X loading buffer (5X TAE with 1% ficoll 400 and 0.05% (w/v)
each of xylene cyanol & bromphenol blue). Samples were added to gel slots and electrophoresed at 50 V until dye entered the gel. The gel was then electrophoresed at 70V for 45 minutes or until the loading dye had reached 2/3 of the way down the gel. The gel was then stained using an ethidium bromide solution for 10 minutes, destained using deionized water for 10 minutes, and viewed using a UV gel box. Smears on the gel in each lane indicate successful digestion. (Fig. 3) When digestion was confirmed, 6 μL TE-4 was added to the remaining 4 μL of digested DNA. If the digested DNA was not used immediately, the sample was stored at 4˚ C.

DNA Ligation

The ends of the restriction fragments were ligated to adaptor pairs complementary to the *Mse I* and *EcoR1* sticky ends provided in the microbial AFLP genotyping kit (Applied Biosystems Inc., Foster City, CA) as per manufacturer's instructions. These adaptors hybridize to the sticky ends created by *MseI* and *EcoRI* digestion and contain additional sequences complementary to primers in both preselective and selective PCR reactions used in the AFLP typing process. Ligase Buffer (0.5 M NaCl, T4 DNA Ligase, and dH_2O) supplied in the AFLP Typing Kit (Applied Biosystems Inc.). 10-30 ng of digested DNA were utilized in each reaction. Each ligation reaction mixture was incubated in a thermocycler for 2.0 hours at 37˚ C. Following incubation, 190 μL of TE-4 was added to each sample and stored at 4˚ C if not immediately used.

Preselective Amplification

Amplification occurs twice during AFLP, with the first round amplifying nonspecifically and the second amplifying specifically. During preselective amplifications, DNA fragments that

have been successfully digested by *EcoRI* and/or *MseI* and ligated to their corresponding oligonucleotide adaptors are amplified. To amplify nonspecifically, two primers, one *MseI* preselective primer and one *EcoRI* preselective primer, bind to DNA fragment ends ligated to their corresponding adaptors. Because the *MseI* recognition site is only 4 basepairs long and the *EcoRI* recognition site is 6 basepairs long, a large number of fragments will be flanked by *MseI* sites or have an *MseI* site on one end and an *EcoRI* site on the other, while a small amount will

Figure 3: Ethidium bromide stained gel of successful of *Salmonella* DNA samples digests

have *EcoRI* sites on both ends. Pre-selective amplification will amplify all fragments that have any of these restriction sites on both ends of a fragment.

Preamplification reactions required 4.0 μL of diluted DNA with about 100 pg of ligated DNA to be mixed with 0.5 μL of *EcoRI* preselective primer, 0.5 μL of *MseI* preselective primer, and 15 μL of AFLP amplification core mix from the AFLP Microorganism Primer Kit (Applied Biosystems Inc.). The samples were then centrifuged briefly in a microcentrifuge at 1000 Xg, placed in the thermocycler, and amplified according to PCR setting recommended in the instructions provided with the AFLP genotyping kit (Table 2).

Following preselective amplification, 10 μL of the amplicon was mixed with 190 μL of TE-4 and stored at 4˚ C if not immediately used. The remaining 10 μL of PCR product was utilized to confirm that ligation and preselective PCR occurred. By mixing 10 μL of the product with 3.0 μL of agarose gel loading buffer and electrophoresing each sample on a 1% agarose gel at 75 V for about an hour, successful amplification and ligation was confirmed by the appearance of a bright streak in a gel track when stained with ethidium bromide and viewed using UV light (Fig. 4).

Table 2: Thermocycler parameters for preselective amplification.

HOLD	**CYCLE**			**HOLD**
	Each of 20 Cycles			
72 ˚ C	94 ˚ C	56 ˚ C	72 ˚ C	4 ˚ C
2 min.	20 sec.	30 sec.	2 min.	(∞)

Selective Amplification

After successful preselective amplification, AFLP requires selective amplification to reduce the number of DNA fragments that go through analysis. Selective amplification use primers that target only the preselectively amplified fragments that also contain one or two complementary nucleotide(s) within the restriction fragment distal to the end of the *MseI* and/or the *EcoRI* recognition sites. By selecting restriction fragments that include these one or two extra complementary nucleotides, the pool of segments amplified during this round is reduced as much as 16-fold. Furthermore, the EcoR1 selective primers are labeled with one of three fluorescent

Figure 4: Ethidium bromide stained gel of successful preselective amplification of four samples of *Salmonella*

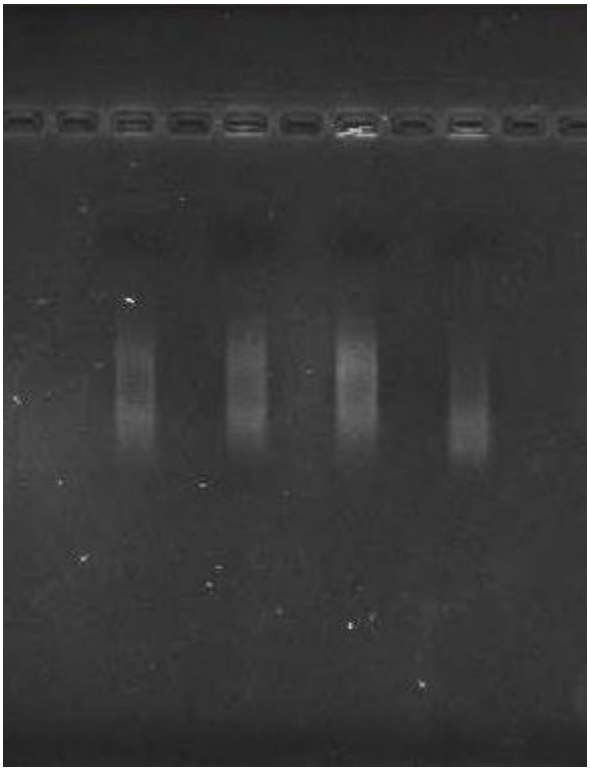


Table 3: Thermocycler parameters for selective amplification.

Hold	Cycle: Selective Amplification			Number of Cycles
72 ˚ C 2 min	94 ˚ C 20 sec	66 ˚ C 30 sec	72 ˚ C 2 min	2
	94 ˚ C 20 sec	65 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	64 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	63 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	62 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	61 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	60 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	59 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	58 ˚ C 30 sec	72 ˚ C 2 min	
	94 ˚ C 20 sec	57 ˚ C sec	72 ˚ C 2 min	
94 ˚ C 2 min	94 ˚ C 20 sec	56 ˚ C 30 sec	72 ˚ C 2 min	20
60 ˚ C 2 min				1
4 ˚ C ∞				1

dyes that indicate the selective nucleotide(s) it contains. AFLP analysis utilizes capillary electrophoresis. In this study the *MseI* selective primers extended beyond the adaptor by the nucleotide adenine (A) nucleotide at the 3' end, thereby amplifying only PCR amplicons produced non-selectively in the first reaction that has a thymidine (T) following the *MseI* recognition sequence. One of three fluorescently labeled *EcoRI* primers is utilized as the other primer in the selective PCR step, each with a different nucleotide extension: adenine (A), guanine (G), or cytosine (C). A green dye, JOE, is added to the *EcoRI*-G primers, a blue dye, FAM, to the *EcoRI*-A primers, and a yellow dye, NED, to the *EcoRI*-C primers. For each non-selectively amplified sample, three different amplifications are prepared, each using a different fluorescently labeled *EcoRI* primer in addition to the *MseI*-A primer. Samples were amplified using the following cycling conditions detailed in Table 3.

Each selective PCR reaction mixture contained 1.5 μL of diluted preselective amplification product with 7.5 μL AFLP core amplification mix, 0.5 μL *MseI*-A primer, and 0.5 μL of either *EcoRI*-A, *EcoRI*-G, or *EcoRI*-C primer. The reaction samples were vortexed, spun down in a microcentrifuge and placed in the thermocycler, and amplified according to settings provided with the kit (Table 3). When all amplification cycles were complete, the samples were stored at 4˚ C until their analysis.

Capillary Electrophoresis

Capillary electrophoresis was performed using the ABI Prism 3130 genetic analyzer (Applied Biosystems Inc.). The ABI Prism 3130 genetic analyzer utilizes four capillaries filled with a filtering polymer, Performance Optimizing Polymer (POP4) connected to a high voltage energy source to power electrophoresis which allows for separation DNA fragments based on size. A laser and detection camera within the machine located at the end of the capillary detects DNA labeled with fluorescent dyes. Each of the selective PCR amplified samples, labeled with a different fluorescent dye, was added to a different well in a 96 well plate. Samples were prepared

by adding 2.0 μL of PCR product to 15 μL of a mixture of Hi-Di formamide (Applied Biosystems Inc.) and LIZ 500 labeled size standard (7 μL of standard added to 1000 μL of Hi-Di formamide) to each well (Applied Biosystems, Inc.). Samples were electrophoresed for 24 minutes at 60˚ C. Each sample yielded three electropherograms, one for each fluorescently labeled dye. FAM labeled fragments appeared blue, JOE fragments appeared green, and NED, the yellow primer, appeared black on the electropherogram to aid visualization. Typical AFLP results are shown in Figures 5A, 5B, and 5C which display the three electropherogram for *Salmonella enterica* subspecies Newport from sample 3.

Each of the three electropherogram (one for each fluorescently labeled primer) provides a visual representation of fragments amplified gathered from the genetic analysis of a single biological sample, the AFLP profile.

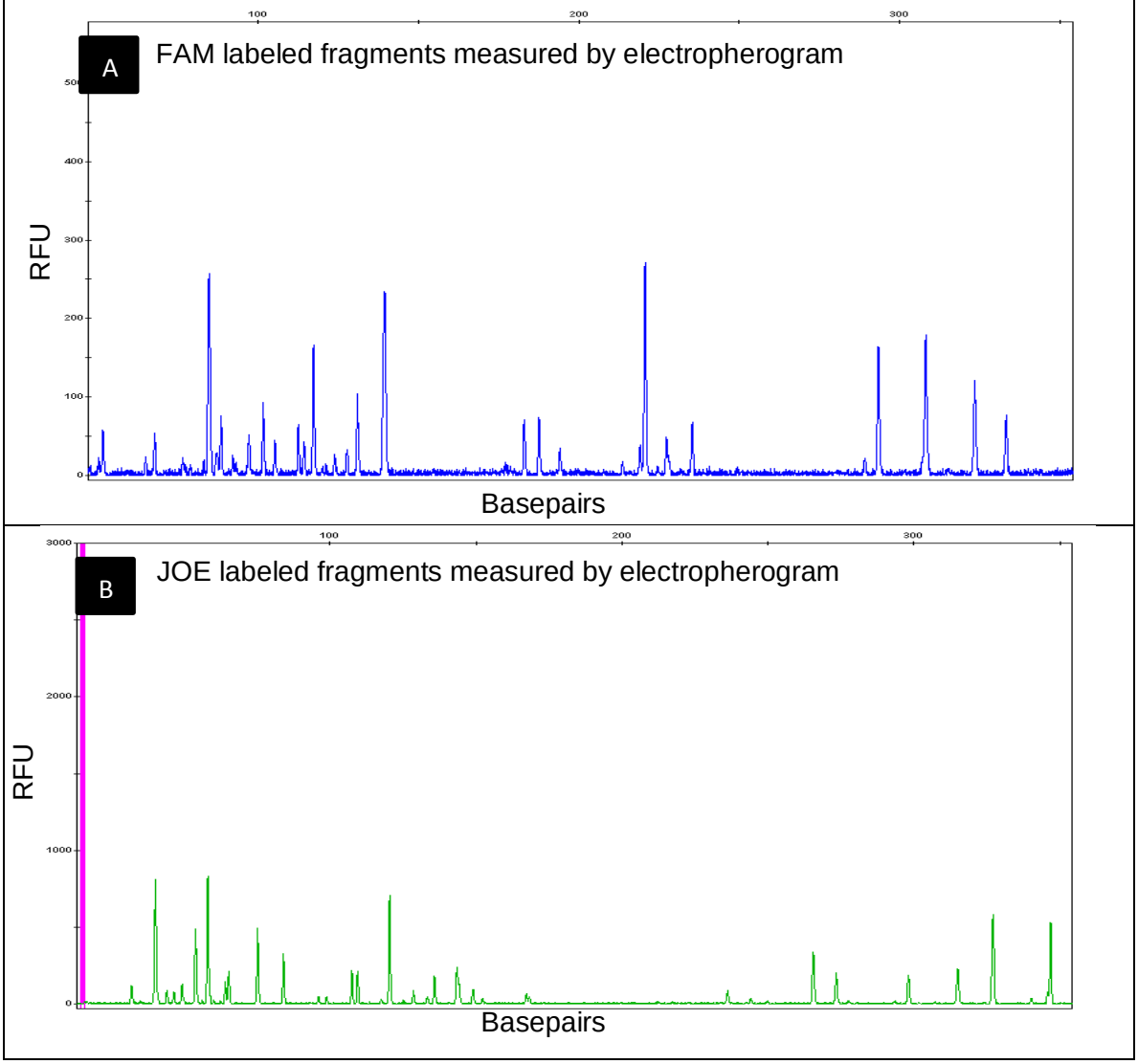
Haplotype Coding

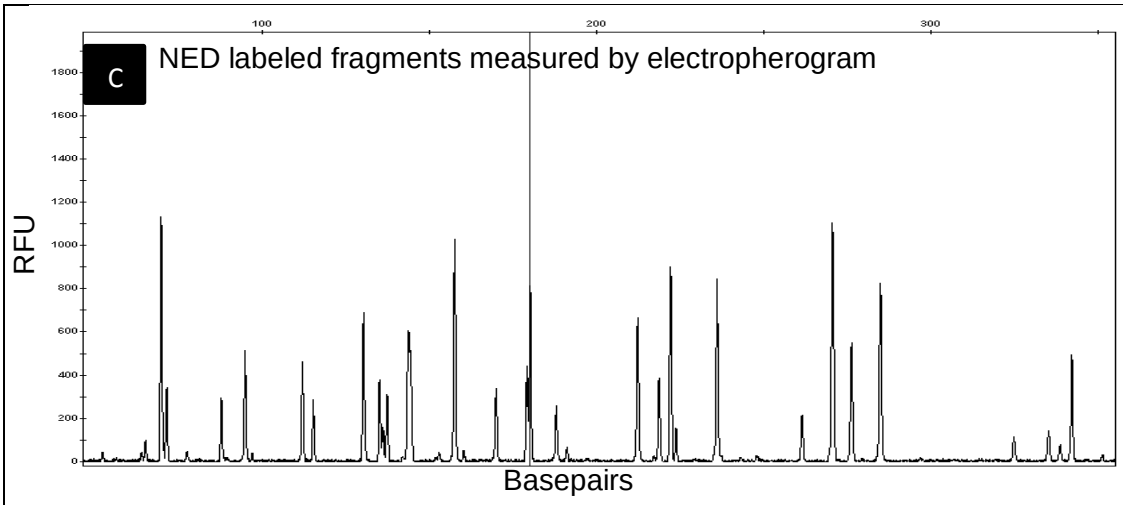
Because the AFLP profile differs between samples, a numerical representation of these differences was developed to compare and contrast the genetic composition of samples. The numerical haplotype code was developed to describe the profile considering the fragments that lie within 70 to 350 basepairs. This range is then divided into 28 bins, each spanning 10 nucleotides. Each bin in each electropherogram is scored for the number of peaks in that bin detected above an established fluorescence threshold. The threshold for peak scoring was set at half of the average relative fluorescence units of peaks in the 70-350 bp size range to normalize runs and increase reproducibility. The number of peaks above the threshold in a given bin created the code for that bin. Each strain analyzed consists of three separate codes, one each for FAM, JOE, and NED electropherograms created from each sample.

Selecting Genomes for *in silico* Analysis

To measure the accuracy of the AFLP codes generated in lab from the *Salmonella* samples, in silico analysis was performed using genomes available through NCBI's RefSeq. Genomes.[42–44] Genomes were selected based upon corresponding strain types and numbers of contigs, or segments of the genome that have been created by overlapping fragments of sequence

Figure 5: *S. enterica* subspecies Newport from sample 3 electropherogram.





The Y-axis of the histogram indicates relative fluorescence units (RFU) and the X-axis indicates fragment size in basepairs. A peak on each histogram indicates an amplified DNA restriction fragment.

Table 4: Genomes accessioned from NCBI's RefSeq for *in silico* analysis.

Organism	Origin	Genome Size	Contigs	Accession	PID	Genes
S. Agona str. SL483	Wisconsin Human Culture 2003 through BEI Resources[42]	4.84 MB	2	CP001138	20063	4,508
S. Typhimurium str. 14028S	Decades old derivative of CDC 60-6516 taken from Chicken hearts and livers in 1960 in US[43]	4.76 MB	2	CP001363	33067	4,653
S. Typhimurium str. D23580	Malawian Adult Blood Culture 2004[44]	4.88 MB	1	FN424405	40625	4,804
S. Newport str. SL254	Minnesota human culture 2000 through BEI Resources[42]	4.83 MB	3	CP001113	18747	4,710

Genomes listed with the same strain as those tested in lab and 3 or less contigs were chosen and accessioned from RefSeq (Table 4). Origins of accessioned genomes were listed to enable origin comparison to the AFLP tested organisms.[42–44]

AFLP *in silico*

Using the genomes accessioned from GenBank, AFLP was performed *in silico* to compare experimental haplotype results to theoretical haplotype results based upon genome analysis. Using Bikandi's online tool, AFLP fragments of the reference genomes (Table 4) that would be amplified as amplified by the Mse1-A and EcoRI-A (FAM), EcoRI-G (JOE), and EcoRI-C primers were obtained.[45] By adding the 33 bp lengths of both adapters to each fragment, the size of the fragment as measured by AFLP was obtained. Fragments were separated into 10 bp bins and tabulated as previously described for experimental AFLP. When multiple fragments of the same length occurred using the same primers, these fragments were counted as a single peak as would appear on an electropherogram obtained experimentally.

CHAPTER IV.

FINDINGS

Evaluating the validity of the AFLP coding process to track foodborne disease requires consideration of the reproducibility, discriminatory power, and accuracy of the method used to create AFLP codes. To evaluate these individual elements, AFLP codes were produced in triplicate from AFLP analyses of *Salmonella* isolates from different sources as well as codes produced *in silico* via sequence analysis. Analyses of these data provides insight regarding reproducibility, discriminatory power and accuracy of haplotype codes to better understand the value and limitations of AFLP analysis and its application for eliminating foodborne outbreaks and bioterrorism.

Reproducibility

In order for a genetic assay to be reliable enough for use in a public health or forensic settings, the results must be reproducible. If the results of a test are to be shared with another lab, scientists must feel confident that they would gain those same results every time they tested that same sample in that same way. Incorporating a codification system into the AFLP typing process should facilitate sharing of results by decreasing subjectivity in the analysis of AFLP electropherograms, an established cause of error in the assay.[35] When comparing AFLP results, only peaks with relative fluorescence (RFUs) above a threshold will contribute to the AFLP

electropherograms, the codification process used here obviates the need for sophisticated software by implementing a simple mathematical method for identifying peaks. However, because of the method used to establish a threshold for peak calling in the AFLP haplotype coding system, individual peaks in the electropherogram may or may not be scored among replicates due to a single Relative Fluorescence Unit (RFU) (Fig. 3). These occasional errors find their way into the haplotype code and can generate dissimilarities between replicates.

To measure reproducibility of haplotype codes generated from AFLP profiles, codes generated from multiple runs of the same samples were compared. To estimate the reproducibility of coding a given sample, percentage similarity was calculated by dividing the number of matching bins between two codes by the total number of bins analyzed.

$$Percentage\ Similarity = \frac{Number\ of\ Matching\ Bins}{Total\ Number\ of\ Bins} \times 100$$

To generate a measurement of reproducibility for a sample, percentage similarities between all three samples were averaged. These reproducibility values are represented in Table 5. Reproducibility was evaluated for the overall haplotype code compiled of all three colors (Table 5a) as well as for each FAM, JOE, & NED (Tables 5b, 5c, & 5d respectively).

The standard error of reproducibility was calculated conservatively using the reproducibility value closest to 0.5 as a baseline.

$$Standard\ Error = \sqrt{\frac{2 \times E\ \times (1 - E)}{Total\ number\ of\ digits\ in\ comparison}}$$

$$Where\ E = Value\ closest\ to\ 0.5\ of\ all\ reproducibility\ values\ compared$$

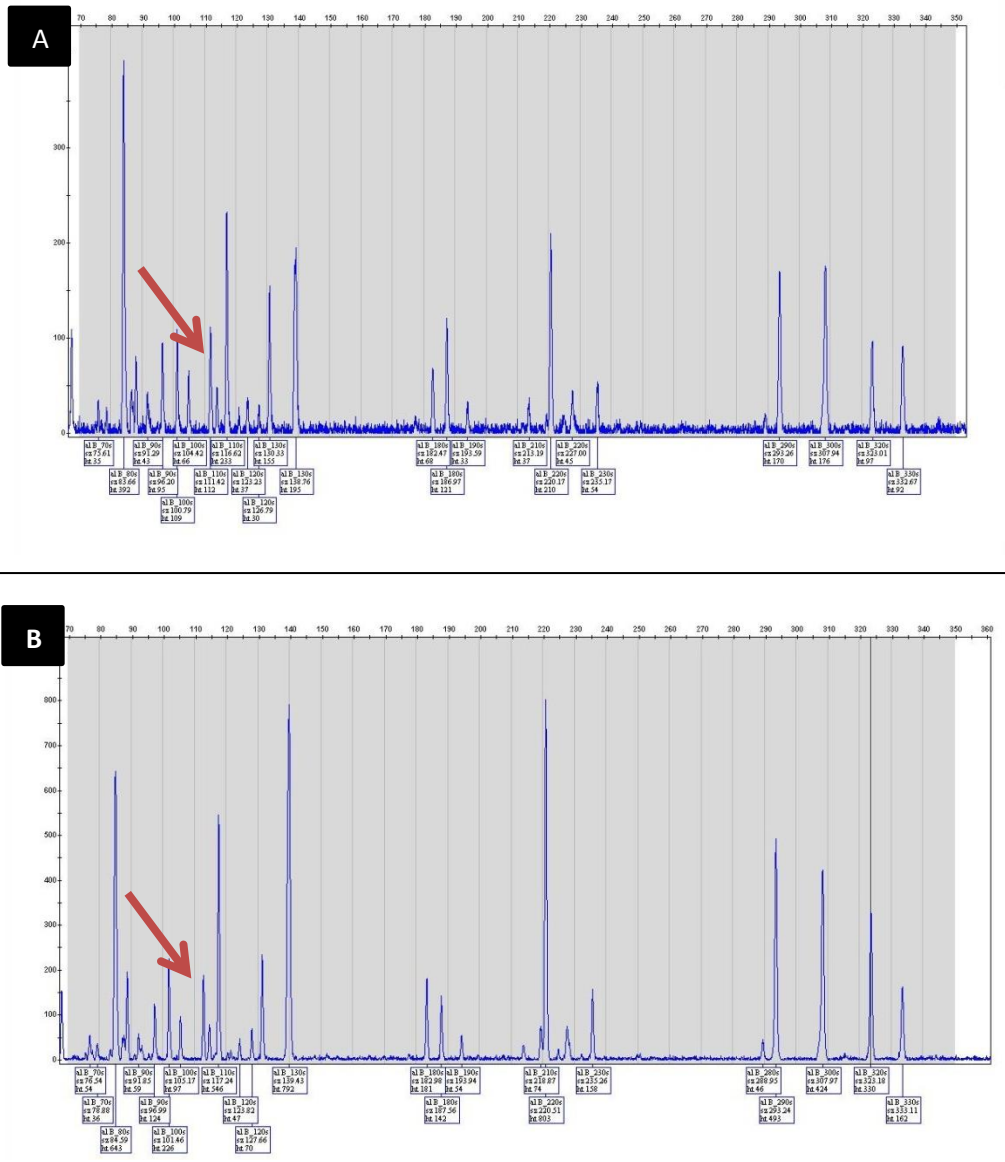


Figure 6: Electropherograms of sample demonstrating effect of threshold on reproducibility.

Electropherograms A and B represent results from AFLP analysis using the FAM dye on sample 3 a sample of *Salmonella enterica* Newport analyzed on two separate occasions. Despite having similar peak patterns in the 110 bp bin, only the code created from electropherogram A incorporated the first peak into the haplotype code. The peak in the same position on electropherogram B was not incorporated into the haplotype code because of falling 2 RFU's below the established threshold for that run.

Table 5a. Comparison of reproducibility of overall haplotype code between samples

Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
87.619^{AB}	86.469^{B}	91.270^{AB}	89.048^{AB}	92.857^{A}	85.317^{B}

Two mean percentage similarities of overall haplotype code with the same letter are not significantly different at α=0.05 (LSD=6.2, mean = 88.537)

Table 5b. Comparison of reproducibility of FAM labeled fragments of haplotype code between samples.

Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
79.762^{A}	75.714^{A}	88.095^{A}	80.714^{A}	85.714^{A}	85.714^{A}

Two mean percentage similarities of FAM (EcoRI-A) haplotype code with the same letter are not significantly different at α=0.05 (Mean= 80.612, LSD=13.0)

Table 5c. Comparison of reproducibility of JOE labeled fragments of haplotype code between samples.

Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
95.833^{A}	97.143^{A}	92.857^{A}	94.286^{A}	79.762^{B}	95.238^{A}

Two mean percentage similarities of JOE (EcoRI-G) haplotype code with the same letter are not significantly different at α=0.05 (Mean=94.082, LSD=12.1)

Table 5d. Comparison of reproducibility of NED labeled fragments of haplotype code between samples.

Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
86.310^{B}	90.00^{AB}	92.857^{AB}	92.143^{AB}	90.476^{AB}	97.619^{A}

Two mean percentage similarities of NED (EcoRI-C) haplotype code with the same letter are not significantly different at α=0.05 (Mean=90.918, LSD=10.4)

Standard error was then multiplied by 1.96 to find the LSD for $\alpha=0.05$. Values that differ by a margin larger than the LSD were deemed significantly different.

The overall haplotype code similarities between repetitions of the same bacterial isolate were averaged to evaluate overall reproducibility in all tested samples. These percentage similarities between repetitions for all samples were averaged for a mean reproducibility of all *Salmonella* tested as 88.537%. Overall haplotype reproducibility ranged from 92.857 for sample 5 to 85.317 for sample 6, representing the only significant difference between sample reproducibilities. Previous studies using similar methods to test both *Pseudomonas* and *Serratia* species had overall reproducibility rates of about 97% and 87%, respectively.[9,10]

Reproducibility was also assessed for each color of the haplotype code. The FAM labeled EcoRI-A portion of the code had the lowest average overall reproducibility of 80.612%, the lowest of the three fluorescent labels, with no statistically significant difference in reproducibility between samples with a range of 75.714% to 88.095%. Standard error for FAM labeled code reproducibility was the largest of all three labels tested at 6.63%. The haplotype code created JOE labeled EcoRI-G portion of the code had the highest overall reproducibility at 94.082% at a range of 79.762 to 97.143%. The reproducibility of JOE labeled sample 5 had the only significant difference in reproducibility at 79.762% with a calculated standard error of 6.17%. The NED labeled haplotype code had an average reproducibility of 90.918% between samples at a range of 86.310 to 97.619%. With a standard error of 5.306%, the only significant difference of reproducibility values between the samples was between sample 1 and sample 7 on either end of the range.

Discriminatory Power

Traditionally and presently in most hospitals, classical microbiological assays involving selective and differential culturing provides a means for identification of *Salmonella* spp. However, in current American public health labs, laboratorians use PFGE to discriminate between outbreaks and share epidemiological information between labs. While this assay can determine whether an organism is part of a given outbreak relatively well, the process is cumbersome and often lacks the power to discriminate between sources, especially in the genus *Salmonella*.

To determine the discriminatory power of AFLP as performed here, the mean similarities between haplotype codes of the paired samples from the three different strains of *S. enterica obtained* from the Oklahoma Department of Public Health's lab were compared. No samples were tested representing the same strain from different outbreaks as characterized by the Oklahoma Department of Health so no comparisons based from outbreak to outbreak can be made.

Discrimination between strains was demonstrated if the mean percentage similarities were significantly different between taxa as calculated using the conservative standard error and LSD estimates previously used to compare reproducibility. Comparisons were made on the sample to sample and strain to strain level for each possible comparison of the six samples tested. Calculations evaluated both the overall haplotype code (Tables 6a & 7a) as well as for each color individually. (Tables 6b-d & 7b-d) to determine the discriminatory effectiveness of each element of the code as well as the codes of each individual color.

Discrimination Between Sources

To determine whether two samples of the same strain and outbreak could be differentiated using AFLP haplotype coding, the percentage similarities between samples was

compared to their reproducibility values. These comparisons were carried out on the overall haplotype code as well as for each section of the code created by a different label (FAM, JOE, or NED). The LSD used for comparing within sample and between sample percentage similarities for the overall haplotype code was 4.4. Using the overall code, both samples 1 and 2 taken from two different patients infected by the Agona strain the Oklahoma Department of Health associated with tomatoes or sprouts from a restaurant could not be differentiated using AFLP.

Table 6a. Comparison of mean percentages of similarity of overall haplotype code between two samples within the same strain.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona	87.619^{A}	87.434^{A}				
Sample 2 Agona	87.434^{A}	86.469^{A}				
Sample 3 Newport			91.270^{A}	86.607^{B}		
Sample 4 Newport			86.607^{A}	89.048^{A}		
Sample 5 Typhimurium					92.857^{A}	87.037^{B}
Sample 6 Typhimurium					87.037^{A}	85.317^{A}

Mean percentage similarities labeled with the same letter in the same row are not significantly different at α=0.05 (LSD= 4.4)

Table 6b. Comparison of mean percentages of similarity of FAM labeled EcoRI-A) haplotype code between two samples within the same strain.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona	79.762^{A}	76.984^{A}				
Sample 2 Agona	76.984^{A}	75.714^{A}				
Sample 3 Newport			88.095^{A}	81.845^{A}		
Sample 4 Newport			81.845^{A}	80.714^{A}		
Sample 5 Typhimurium					85.714^{A}	82.540^{A}
Sample 6 Typhimurium					82.540^{A}	85.714^{A}

Mean percentage similarities labeled with the same letter in the same row are not significantly different at α=0.05 (LSD= 10.7)

Table 6c. Comparison of mean percentages of similarity of JOE labeled (EcoRI-G) haplotype code between two samples within the same strain.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona	95.833^{A}	96.627^{A}				
Sample 2 Agona	96.627^{A}	97.143^{A}				
Sample 3 Newport			92.857^{A}	89.881^{A}		
Sample 4 Newport			89.811^{A}	94.286^{A}		
Sample 5 Typhimurium					79.762^{A}	86.905^{A}
Sample 6 Typhimurium					86.905^{A}	95.238^{A}

Mean percentage similarities labeled with the same letter in the same row are not significantly different at α=0.05 (LSD= 9.7)

Table 6d. Comparison of mean percentages of similarity of NED labeled (EcoRI-C) haplotype code between two samples within the same strain.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona	86.310^{A}	88.690^{A}				
Sample 2 Agona	88.690^{A}	90.000^{A}				
Sample 3 Newport			92.857^{A}	88.095^{A}		
Sample 4 Newport			88.095^{A}	92.143^{A}		
Sample 5 Typhimurium					90.476^{A}	91.667^{A}
Sample 6 Typhimurium					91.667^{A}	97.617^{A}

Mean percentage similarities labeled with the same letter in the same row are not significantly different at α=0.05 (LSD= 9.2)

For both of the Newport samples taken from stools of sick patients in Oklahoma and Virginia that the State Health Department linked and the both of the samples of the Typhimurium strain aquarium outbreak taken from frog and stool, a significant difference between reproducibility of samples from the same outbreak could be found in half of the comparisons. In the Newport samples, 3 and 4, the similarity of sample 3 to itself was significantly different from the similarity between samples 3 and 4. However, the similarity of sample 4 to itself could not be deemed significantly different from sample 4's similarity to sample 3. A similar situation occurred in the Typhimurium samples 5 and 6, as the similarity of samples 5 to itself was significantly higher than the similarity between samples 5 & 6. However, the similarity of the overall haplotype code of samples 6 to itself was not significantly different from the overall haplotype code similarity between sample 5 and sample 6. Past studies proved able to differentiate between sources of the same strain of *Pseudomonas* using the overall AFLP haplotype code, however these organisms were cultured and maintained in separate labs for several years.[10]

Each of the labeled portions of the code, FAM, JOE, and NED were also analyzed using comparisons of reproducibility values with mean percentage similarities of samples from the same outbreak and strain. Each of the three dyes could not individually distinguish between the same sample and a sample of the same outbreak in any of the three pairs of *Salmonella* tested. Each of the three dyes tested had similar standard error values for this comparison at 5.5, 5.0, and 4.7 for FAM, JOE, and NED respectively. Therefore, though each code individually cannot make any differentiations, combining the three labeled elements of the haplotype code can establish some minor ability to differentiate between samples of the same outbreak.

Table 7a. Comparisons of mean percentage similarities for overall haplotype code among samples of all three strains.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona		87.434^{A}	81.994^{B}	79.762^{B}	80.357^{B}	79.960^{B}
Sample 2 Agona	87.434^{A}		80.754^{BC}	82.500^{B}	81.448^{BC}	78.373^{C}
Sample 3 Newport	81.994^{B}	80.754^{B}		86.607^{A}	84.656^{AB}	82.672^{AB}
Sample 4 Newport	79.762^{B}	82.500^{B}	86.607^{A}		83.333^{AB}	80.357^{B}
Sample 5 Typhimurium	80.357^{C}	81.448^{BC}	84.656^{AB}	83.333^{ABC}		87.037^{A}
Sample 6 Typhimurium	79.960^{BC}	78.373^{C}	82.6721^{B}	80.357^{BC}	87.037^{A}	

Mean percentage similarities in the same row labeled with the same letter are not significantly different at α=0.05 (Mean=82.414, LSD=4.1)

Table 7b. Comparisons of mean percentage similarities for FAM (EcoRI-A) haplotype code among samples of all three strains.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona		76.984^{AB}	80.357^{A}	74.643^{AB}	75.000B^{AB}	71.429^{B}
Sample 2 Agona	76.984^{A}		74.702^{A}	77.857^{A}	72.321^{A}	74.107^{A}
Sample 3 Newport	80.357^{A}	74.702^{A}		81.845^{A}	82.540^{A}	76.984^{A}
Sample 4 Newport	74.643^{A}	77.857^{A}	81.845^{A}		76.488^{A}	74.107^{A}
Sample 5 Typhimurium	75.000^{AB}	72.321^{B}	82.540^{A}	76.488^{AB}		82.540^{A}
Sample 6 Typhimurium	71.429^{B}	74.107^{B}	76.984^{AB}	74.107^{B}	82.540^{A}	

Mean percentage similarities in the same row labeled with the same letter are not significantly different at α=0.05 (Mean=76.573, LSD=7.9)

Table 7c. Comparisons of mean percentage similarities for JOE (EcoRI-G) haplotype code among samples of all three strains tested using AFLP.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona		96.672^{A}	93.750^{AB}	88.750^{BC}	86.607^{C}	88.988^{BC}
Sample 2 Agona	96.672^{A}		95.238^{A}	92.321^{AB}	87.798^{B}	91.071^{AB}
Sample 3 Newport	93.750^{AB}	95.238^{A}		89.881^{AB}	87.302^{B}	92.063^{AB}
Sample 4 Newport	88.750^{AB}	92.321^{A}	89.881^{A}		83.333^{B}	87.500^{AB}
Sample 5 Typhimurium	86.607^{A}	87.798^{A}	87.302^{A}	83.333^{A}		86.905^{A}
Sample 6 Typhimurium	88.988^{A}	91.071^{A}	92.063^{A}	87.500^{A}	86.905^{A}	

Mean percentage similarities in the same row labeled with the same letter are not significantly different at α=0.05 (Mean=90.192, LSD=6.5)

Table 7d. Comparisons of mean percentage similarities for NED (EcoRI-C) haplotype code among samples of all three strains tested using AFLP.

	Sample 1 Agona	Sample 2 Agona	Sample 3 Newport	Sample 4 Newport	Sample 5 Typhimurium	Sample 6 Typhimurium
Sample 1 Agona		88.690^{A}	71.726^{C}	75.893^{BC}	78.274^{BC}	80.655^{AB}
Sample 2 Agona	88.690^{A}		72.321^{B}	77.321^{B}	75.000^{B}	79.167^{B}
Sample 3 Newport	71.726^{C}	72.321^{C}		88.095^{A}	78.175^{BC}	84.921^{AB}
Sample 4 Newport	75.893^{B}	77.321^{B}	88.095^{A}		81.250^{AB}	88.393^{A}
Sample 5 Typhimurium	78.274^{B}	75.000^{B}	78.175^{B}	81.25^{B}		91.667^{A}
Sample 6 Typhimurium	80.655^{BC}	79.167^{C}	84.921^{ABC}	88.393^{AB}	91.667^{A}	

Mean percentage similarities in the same row labeled with the same letter are not significantly different at α=0.05 (Mean=80.477, LSD=7.9)

Differentiation of *Salmonella* Strains

To determine whether AFLP haplotype coding can be used to differentiate between *Salmonella* strains, mean similarity values for the overall haplotype code as well as for each label between samples of the same strain and samples of different strains were compared.

The average percentages of similarity of all of the *Salmonella* samples is 82.424% when compared using the overall haplotype code. The standard error was calculated to be 2.1% with LSD of 4.1%. Previous studies using similar methods with *Pseudomonas aeruginosa* found similarity values among various strains of the species of 54%.[10](Table 7a) Both samples Agona 1 and 2 can be differentiated from other strains as the mean percentage similarity of the overall haplotype code between samples of the same strain and outbreak. However, mean similarity percentages of the overall haplotype code between either Typhimurium and Newport and the Agona samples were not statistically different.

Samples 3 and 4 of the Newport strain could not be differentiated from the other two strains using the mean similarity percentages of the overall haplotype code. Sample 3 was not statistically more similar to sample 4 than to sample 6, a Typhimurium strain. Sample 4 was not statistically more similar to sample 3 than to sample 5, a Typhimurium strain when comparing the overall haplotype code. (Table 7a) In addition, overall haplotype codes from neither sample 3 nor 4 were statistically more similar to the Typhimurium and Agona. Similarly, samples 5 and 6 of the Typhimurium strain could not be differentiated from the other two strains. The overall haplotype code of sample 5 was not statistically more similar to sample 6, of the same strain, than to sample 3 or to sample 4 of the Newport strain. Sample 6 was not statistically more similar to sample 5, of the same strain, than to samples 3 or 4, of the Newport strain when comparing complete haplotype codes. Furthermore mean similarity percentages of the overall haplotype code between either Newport or Agona and the Typhimurium samples were not statistically different.

Using only the FAM (EcoRI-A) labeled code, none of the strains could be successfully differentiated. When averaged, the average mean similarity percentage for the FAM haplotype was 76.573 with a standard error of 4.0% and LSD of 7.9%. No set of comparisons allowed for statistically significant differentiations of degrees of relatedness between different strains using only the FAM code. (Table 7b)

Utilizing the JOE (EcoRI-G) labeled portion of the haplotype code alone also allowed for no successful differentiation between strains of *Salmonella* tested. The average mean percentage similarities between all *Salmonella* samples for the JOE haplotype was 90.192 with a standard error at 3.3% and LSD of 6.5. The JOE labeled portion of the haplotype code like the FAM portion was unable to differentiate between strains using mean percentage similarities. (Table 7c)

The NED portion of the haplotype code was the most successful at discriminating between strain types, however could not perform this task in all instances. The average of all mean percentage similarities of NED haplotypes for all *Salmonella* samples compared was 80.477 with a standard error of 4.0% and LSD of 7.9%. (Table 7d)

AFLP *in silico* Similarities

Using Bikandi's online AFLP-PCR *in silico* [45] genomes of *Salmonella* Agona, *Salmonella*, Newport, and *Salmonella* Typhimurium were analyzed to produce fragment lengths for the different bacteria that would be labeled by FAM (EcoRI-A), JOE (EcoRI-G) or NED (EcoRI-C) primers. The *in silico* restricted fragments were then compiled into the AFLP haplotype codes using fragments in the 70-350 bp size range described previously. Though reproducibility values cannot be obtained from this *in silico* method, theoretical percentage similarities in AFLP haplotype code can easily be generated (Table 8).

Table 8: Percentages of similarity between overall AFLP haplotype profiles produced *in silico* from reference genomes.

	Agona SL483	Newport SL254	Typhimurium 14028S	Typhimurium D23580
Agona SL483		50%	53%	54%
Newport SL254	50%		58%	59%
Typhimurium 14028S	53%	58%		90%
Typhimurium D23580	54%	59%	90%	

The percentages of similarity of the overall haplotype codes generated from *Salmonella* genomes of different serotypes range from 50-59%. Interestingly, this value is similar to that described by Weinbrecht et al for strains of *Pseudomonas.*[12] However, the comparison made between two *in silico* haplotype codes from genomes of the same Typhimurium serotype, the percentage is higher at 90%.

Comparison of AFLP *in silico* and AFLP *in vitro*

The accuracy of AFLP code results obtained in the lab were measured by comparison to results prepared *in silico*. Because the sample code for each organism sample's code was not highly reproducible, each AFLP code created in lab was compared to the *in silico* code individually. A percentage similarity between the *in silico* code and each replication of each sample was calculated. The mean of the percentage similarities within a sample were calculated and noted in Table 9. The mean percentage similarities between lab created codes and their *in silico* counterparts of the same strain range from 15-20%. However, comparisons between lab generated overall haplotype codes and overall *in silico* haplotype codes of different strains had a

Table 9: Mean percentages of similarity between overall AFLP haplotype profiles produced *in silico* in comparison to codes generated *in vitro.*

		Prepared *in silico*			
		Agona SL483	Newport SL254	Typhimurium 14028S	Typhimurium D23580
Prepared in lab	Sample 1 Agona	19%	14%	18%	19%
	Sample 2 Agona	20%	17%	19%	20%
	Sample 3 Newport	23%	15%	17%	18%
	Sample 3 Newport	21%	15%	18%	19%
	Sample 4 Typhimurium	18%	14%	19%	17%
	Sample 5 Typhimurium	21%	12%	18%	15%

similar range of percentages with some percentages even higher than comparison within a strain. Notably, many of the digits in the codes that were the same between *in silico* and *in vitro* codes were zeroes, indicating no peaks within a given 10 bp range. In general, the *in silico* haplotype codes counted more peaks per bin than the *in vitro* haplotype codes, especially in bins of smaller fragment sizes. Thus, where we would score a single peak within a bin, there might actually be multiple, overlapping peaks in reality.

CHAPTER V

CONCLUSION

Genotyping bacteria through using AFLP has been considered as a technique to supplement or possibly replace PFGE to trace source and strain of foodborne outbreaks including *Salmonella*. Previous studies have successfully used the AFLP haplotype coding system to increase reproducibility and normalize AFLP results of other types of bacteria including species in the genera *Pseudomonas* and *Serratia*.[9,12] This study addressed whether using AFLP in conjunction with the AFLP haplotyping system on samples of *Salmonella* obtained from the Oklahoma Department of Health would create reproducible, discriminatory, and accurate results. AFLP haplotype codes created from *Salmonella* samples and their comparison to *in silico* developed haplotype codes demonstrate that the technique lacks in reproducibility, ability to discriminate adequately between strains and sources, and accuracy when compared to *in silico* produced codes. Thus, the weaknesses associated with molecular and serological strain typing methods used routinely are also apparent with the AFLP typing procedure.

Because human lives are affected quickly in *Salmonella* foodborne outbreaks, scientists want to ensure that the results obtained from AFLP are reproducible. Due to the importance of reproducibility in instances such as these, AFLP in conjunction with the AFLP haplotype coding system is not suitable for outbreak tracking. With a mean reproducibility of 88.537% and

Fisher's least standard difference of 6.2%, the reproducibility values for the AFLP haplotyping system when applied to *Salmonella* are below an acceptable percentage for use in a public health or forensic setting. Furthermore, investigation of each colored portion of the haplotype coding system demonstrated similar reproducibility values of 90.918, 80.612, and 94.082 for FAM, JOE, and NED, respectively. This indicated that while certain dyes produce more reproducible results than others, no one dye's results are as reproducible as those described in previous studies with other genera of bacteria.[9,12]

Percentages of similarity among different strains of *Salmonella* tested were generally much higher than percentages similarity found between different *Salmonella* strains of *Serratia* and *Pseudomonas*. Percentage similarities for two different strains ranged from 78.373 to 84.656% while in *Pseudomonas* the mean of percentage similarities between strains was 54%.[10] While some of this difference in similarity is likely due to errors discussed here within the assay itself, these percentages likely reflect a very high level of similarity among strains of *Salmonella* in comparison to other bacterial species.[46,47] That said, , the *in silico* results indicate that AFLP profiling and scoring of fragments between 70 and 350 bp into 10 bp bins theoretically yields AFLP code similarities between serotypes of *Salmonella enterica* at somewhere between 50-60%.(Table 8) These between serotype comparisons correspond to similarity values between *Pseudomonas* and *Serratia* strains reported in previous studies by Weinbrecht and Beauman.[9,10] The *in silico* values also indicate even when each sized fragment created through AFLP is accounted for in the haplotype code, only small differences in similarity are present between samples of the same serotype, indicating the inability of AFLP's to distinguish between source or outbreak(Table 8). Given that the *in silico* AFLP testing was still not powerful enough to discriminate between serotypes, it is possible that even nucleotide sequencing would reveal only small differences among the *Salmonella* strains in comparison to other bacteria.[46]

Causes for the error in reproducibility and differences between *in vitro* prepared and *in silico* generated AFLP codes abound. Using the threshold as established by the haplotype coding method eliminates peaks to increase the ratio of signal to noise. However, the current threshold of half of the average of all peaks in a given electropherogram not only requires additional calculation and thereby additional source of error, it eliminates peaks that not only could increase the discriminatory power of the assay but also contribute to reproducibility of results (Figure 3). AFLP studies in other fields, namely molecular ecology, set thresholds at a given RFU, usually not more than 75, and run blank samples of water along with their replicates to obtain a baseline RFU measurement with error rates less than 2%.[48,49] Calculated thresholds in this study as well as studies by Weinbrecht and Beauman were much higher and thereby neglected useful information that other AFLP studies used with great success.[12,48,49]

Despite the indications of reproducibility from the *in vitro* results, the *in silico* results and the *in vitro* results varied by a large margin, indicating inaccuracy in the result. The causes of these inaccuracies likely pertain to both the experimental and analytical aspects of preparing an AFLP code. Many of the bins in the AFLP code that did match between techniques had zero to two peaks present in them and those that did not often differed because the *in silico* method indicated that more peaks should be present. Aside from scoring differences due to threshold, one source of error here is likely what is termed "PCR selection".[50] PCR selection occurs during the primer annealing step of any multiplexed PCR reaction. Due to differing binding affinities and thermodynamics within the PCR reaction, primers and adapters bind to certain segments of nucleotide more often than others due to either simple chemical interactions or secondary structure. As the PCR cycles continue, the reaction mixture becomes crowded with replicates of the fragments with higher binding affinities and the likelihood of the other fragments with lower binding affinities to become replicated decreases.[49, 50] PCR selection occurring inside of AFLP reactions of similar organisms could create contribute to inaccurate profiles, like the results seen

for *Salmonella*. To possibly eliminate this problem, MseI could be replaced with an enzyme that makes fewer cuts in order to avoid the plethora of fragments within the reaction mixture, or primers with more than one selective nucleotide should be used to increase the specificity thereby increasing the binding affinity of fluorescently-labeled primers used in AFLP and increasing reproducibility of the AFLP code.[51]

Though making these changes to the AFLP procedure might yield more accurate, reproducible results, the technique of AFLP even when performed with high levels precision and accuracy yield less useful information than sequencing techniques that have already been used to track and effectively end outbreaks, both foodborne and otherwise. Next-gen sequencing provides clinical labs the ability to identify, track, and better treat outbreaks; by accessing the genomic information of the pathogen, clinicians can access knowledge useful in both clinical and forensic applications including the antibiotics that the pathogen is susceptible or resistant to, how often the pathogen mutates, and how it might be transferred.

At this moment the resources required to implement next generation sequencing technology, which employ massively parallel sequencers to increase the speed of the traditional Sanger or capillary electrophoresis based sequencing processes, in all public health labs in America include a high setup cost, the cost to run the tests, and experienced personnel to run and analyze the data. Agencies that monitor outbreaks in forward-thinking, modern nations including Public Health England are already seriously considering implementation of whole- genome sequencing in all public health labs as a cost-effective means to initially supplement current methods and then transition to whole genome sequencing investigate outbreaks. Investigations performed through whole genome sequence analysis have been able to track both widespread disease and nosocomially spread disease outbreaks in locations where the technology has been implemented.[26,52,53] The cost to set up a desktop sequencer can initially seem expensive and cost prohibitive. However, when processing a large amount of samples, as a hospital or public health

lab would, the cost would be absolutely miniscule in comparison to the cost of alternative assays in combination with the cost of treating additional individuals who have been infected by the pathogen while those tests were being run. The remaining barriers for implementing whole genome sequencing, genomic analysis and information sharing databases for use in public health labs in highly developed nations lay in the area of bioinformatics, whereas barriers for implementing AFLP as an epidemiological tracking method still abound in areas of technique, analysis, and efficiency.

Evaluation of the reproducibility, accuracy, and discriminatory ability of the AFLP technique as applied to *Salmonella* outbreaks indicates that it would be an inadequate replacement of PFGE for forensic or epidemiological purposes. Though the AFLP technique as applied here has room for improvement, even the most accurate, reproducible, and discriminatory AFLP results possible could never supply the information that whole genome sequencing could at a reasonable price and much smaller time interval. Until genetic sequencing becomes cost effective enough to be present in all types of molecular biology labs, the future of AFLP lies in ecological and non-human population genetics research rather than in arena in which lives are at stake.

REFERENCES

1. Scharff RL. Economic Burden from Health Losses Due to Foodborne Illness in the United States. *J Food Prot*. 2012;75(1):123–31.

2. CDC - Foodborne Outbreak Investigations - Controlling an Outbreak. Available at: http://www.cdc.gov/outbreaknet/investigations/control.html. Accessed April 17, 2012.

3. CDC. Foodborne Diseases Active Surveillance Network (FoodNet). *FoodNet surveillance report for 2010 (final report)*. Atlanta, Georgia: U.S. Department of Health and Human Services, CDC; 2011.

4. WHO | Drug-resistant *Salmonella*. *WHO*. Available at: http://www.who.int/mediacentre/factsheets/fs139/en/. Accessed September 14, 2012.

5. Török TJ, Tauxe RV, Wise RP, et al. A Large Community Outbreak of Salmonellosis Caused by Intentional Contamination of Restaurant Salad Bars. *JAMA J Am Med Assoc*. 1997;278(5):389–395. doi:10.1001/jama.1997.03550050051033.

6. CDC - Prevention - *Salmonella*. Available at: http://www.cdc.gov/*Salmonella*/general/prevention.html. Accessed November 18, 2012.

7. Ross IL, Heuzenroeder MW. Use of AFLP and PFGE to discriminate between *Salmonella* enterica serovar Typhimurium DT126 isolates from separate food-related outbreaks in Australia. *Epidemiol Infect*. 2005;133(4):635–644.

8. McDermott M. 2012 *Salmonella* Strain Information. 2013.

9. Beauman C. Evaluating the suitability of AFLP technology for genotyping strains of Serratia marcescens. 2007.

10. Weinbrecht KD. Amplified fragment length polymorphism (AFLP) analysis as a DNA typing tool for microbial forensics. 2011.

11. Meudt HM, Clarke AC. Almost Forgotten or Latest Practice? AFLP applications, analyses and advances. *Trends Plant Sci*. 2007;12(3):106–117. doi:10.1016/j.tplants.2007.02.001.

12. Weinbrecht K, Taylor A, Beaumann C, Allen RW. Exploration of amplified fragment length polymorphism analysis as an effective tool for discriminating pathogenic strains of bacteria. *Res Reports Forensic Med Sci*. 2012;2:9–17.

13. CDC - Outbreaks - OutbreakNet Team. Available at: http://www.cdc.gov/outbreaknet/outbreaks.html. Accessed September 24, 2012.

14. WHO | Microbiology Modules. *WHO*. Available at: http://www.who.int/gfn/training/TCLevel1/en/index.html. Accessed July 1, 2013.

15. WHO | Foodborne disease outbreaks: Guidelines for investigation and control. *WHO*. Available at: http://www.who.int/foodsafety/publications/foodborne_disease/fdbmanual/en/. Accessed July 1, 2013.

16. Fatica MK, Schneider KR. *Salmonella* and produce: Survival in the plant environment and implications in food safety. *Virulence*. 2011;2(6):573–579. doi:10.4161/viru.2.6.17880.

17. f C. *Salmonella* Infections. In: Viola Magni M, ed. *Detection of Bacteria, Viruses, Parasites and Fungi*. NATO Science for Peace and Security Series A: Chemistry and Biology. Springer Netherlands; 2010:125–138. Available at: http://dx.doi.org/10.1007/978-90-481-8544-3_6.

18. Kauffmann F. The Bacteriology of Enterobacteriaceae. Collected Studies of the Author and his Co-Workers. *Bacteriol Enterobact Collect Stud Author His Co-Work*. 1966.

19. Crosa JH, Brenner DJ, Ewing WH, Falkow S. Molecular Relationships Among the Salmonelleae. *J Bacteriol*. 1973;115(1):307–315.

20. Le Minor L, Veron M, Popoff M. The taxonomy of *Salmonella*. In: *Annales de microbiologie*.Vol 133.; 1982:223.

21. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B. *Salmonella* Nomenclature. *J Clin Microbiol*. 2000;38(7):2465–2467.

22. Scientific Nomenclature - Emerging Infectious Disease journal - CDC. Available at: http://wwwnc.cdc.gov/eid/pages/scientific-nomenclature.htm. Accessed July 23, 2013.

23. CDC Data & Statistics | Feature: Foodborne Disease Outbreaks Are Deadly Serious – What You Can Do to Avoid Them. Available at: http://www.cdc.gov/features/dsFoodborneOutbreaks/. Accessed September 23, 2012.

24. Majowicz SE, Musto J, Scallan E, et al. The global burden of nontyphoidal *Salmonella* gastroenteritis. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2010;50(6):882–889. doi:10.1086/650733.

25. ERS/USDA Data Foodborne Illness Cost Calculator. 2011. Available at: http://webarchives.cdlib.org/wayback.public/UERS_ag_1/20120311151624/http://ers.usda.gov/Data/FoodborneIllness/salm_Intro.asp. Accessed September 14, 2012.

26. Reuter S, Ellington MJ, Cartwright EP, et al. RApid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology. *JAMA Intern Med*. 2013;173(15):1397–1404. doi:10.1001/jamainternmed.2013.7734.

27. Appelbaum PC, Arthur RR, Parker ME, Shugar GL, von Kuster LC, Charache P. Comparison of three methods for identification of Enterobacteriaceae. *Eur J Clin Microbiol*. 1982;1(2):76–81.

28. Herschleb J, Ananiev G, Schwartz DC. Pulsed-field gel electrophoresis. *Nat Protoc*. 2007;2(3):677–684. doi:10.1038/nprot.2007.94.

29. Swaminathan B, Barrett TJ, Hunter SB, Tauxe RV. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. *Emerg Infect Dis*. 2001;7(3):382–389. doi:10.3201/eid0703.010303.

30. Boxrud D, Monson T, Stiles T, Besser J. The role, challenges, and support of pulsenet laboratories in detecting foodborne disease outbreaks. *Public Heal Reports Wash DC 1974*. 2010;125 Suppl 2:57–62.

31. Sandt CH, Krouse DA, Cook CR, Hackman AL, Chmielecki WA, Warren NG. The Key Role of Pulsed-Field Gel Electrophoresis in Investigation of a Large Multiserotype and Multistate Food-Borne Outbreak of *Salmonella* Infections Centered in Pennsylvania. *J Clin Microbiol*. 2006;44(9):3208–3212. doi:10.1128/JCM.01404-06.

32. Bergamini F, Iori A, Massi P, Pongolini S. Multilocus variable-number of tandem-repeats analysis of *Salmonella* enterica serotype Gallinarum and comparison with pulsed-field gel electrophoresis genotyping. *Vet Microbiol*. 2011;149(3–4):430–436. doi:10.1016/j.vetmic.2010.12.002.

33. Maiden MCJ, Bygraves JA, Feil E, et al. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci*. 1998;95(6):3140 –3145.

34. Wang B, Wang C, McKean JD, et al. *Salmonella* enterica in swine production: assessing the association between amplified fragment length polymorphism and epidemiological units of concern. *Appl Environ Microbiol*. 2011;77(22):8080–8087. doi:10.1128/AEM.00064-11.

35. Bonin A, Bellemain E, Bronken Eidesen P, Pompanon F, Brochmann C, Taberlet P. How to track and assess genotyping errors in population genetics studies. *Mol Ecol*. 2004;13(11):3261–3273. doi:10.1111/j.1365-294X.2004.02346.x.

36. Paris M, Bonnes B, Ficetola GF, Poncet BN, Després L. Amplified fragment length homoplasy: in silico analysis for model and non-model species. *BMC Genomics*. 2010;11(1):287. doi:10.1186/1471-2164-11-287.

37. Zhang M, Liang S, Huang L, Sun Y, Zhou J, Wang C. AFLP runner (AR): A software for prediction and analysis Of AFLP. In: *2010 3rd International Conference on Biomedical Engineering and Informatics (BMEI)*.Vol 5.; 2010:2150–2153. doi:10.1109/BMEI.2010.5639293.

38. Paris M, Després L. In Silico Fingerprinting (ISIF): A User-Friendly In Silico AFLP Program. In: Pompanon F, Bonin A, eds. *Data Production and Analysis in Population Genomics*. Methods in Molecular Biology. Humana Press; 2012:55–64. Available at: http://link.springer.com/protocol/10.1007/978-1-61779-870-2_4. Accessed June 30, 2013.

39. Rombauts S, Peer YV de, Rouzé P. AFLPinSilico, simulating AFLP fingerprints. *Bioinformatics*. 2003;19(6):776–777. doi:10.1093/bioinformatics/btg090.

40. Arnold C, Metherell L, Clewley JP, Stanley J. Predictive modelling of fluorescent AFLP: a new approach to the molecular epidemiology of E. coli. *Res Microbiol*. 1999;150(1):33–44. doi:10.1016/S0923-2508(99)80044-8.

41. *Communicable disease and injury reporting effective 7/12/2012*.; 2012:1–4. Available at: http://www.ok.gov/health2/documents/Chapter%20515%20Final%20Rules%202012.pdf.

42. Fricke WF, Mammel MK, McDermott PF, et al. Comparative genomics of 28 *Salmonella* enterica isolates: evidence for CRISPR-mediated adaptive sublineage evolution. *J Bacteriol*. 2011;193(14):3556–3568. doi:10.1128/JB.00297-11.

43. Jarvik T, Smillie C, Groisman EA, Ochman H. Short-term signatures of evolutionary change in the *Salmonella* enterica serovar typhimurium 14028 genome. *J Bacteriol*. 2010;192(2):560–567. doi:10.1128/JB.01233-09.

44. Okoro CK, Kingsley RA, Connor TR, et al. Intracontinental spread of human invasive *Salmonella* Typhimurium pathovariants in sub-Saharan Africa. *Nat Genet*. 2012;44(11):1215–1221. doi:10.1038/ng.2423.

45. Bikandi J, Millán RS, Rementeria A, Garaizar J. In silico analysis of complete bacterial genomes: PCR, AFLP–PCR and endonuclease restriction. *Bioinformatics*. 2004;20(5):798–799. doi:10.1093/bioinformatics/btg491.

46. Jacobsen A, Hendriksen R, Aaresturp F, Ussery D, Friis C. The *Salmonella* enterica Pan-genome. *Microb Ecol*. 2011;62(3):487–504. doi:10.1007/s00248-011-9880-1.

47. Kiil K, Binnewies TT, Willenbrock H, et al. Comparative genomics of Pseudomonas. *Pseudomonas Model Org Pathog Cell Fact Weinh Wiley VCH*. 2008:1–24.

48. Keyghobadi N, Crawford LA, Maxwell SA. Successful analysis of AFLPs from non-lethally sampled wing tissues in butterflies. *Conserv Genet*. 2009;10(6):2021–2024. doi:10.1007/s10592-009-9884-2.

49. Whitlock R, Hipperson H, Mannarelli M, Butlin RK, Burke T. An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Mol Ecol Resour*. 2008;8(4):725–735. doi:10.1111/j.1755-0998.2007.02073.x.

50. Wagner A, Blackstone N, Cartwright P, et al. Surveys of Gene Families Using Polymerase Chain Reaction: PCR Selection and PCR Drift. *Syst Biol*. 1994;43(2):250. doi:10.2307/2413465.

51. Han TH, Eck HJ van, Jeu MJD, Jacobsen E. Optimization of AFLP fingerprinting of organisms with a large-sized genome: a study on Alstroemeria spp. *Theor Appl Genet*. 1999;98(3-4):465–471. doi:10.1007/s001220051093.

52. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet*. 2012;13(9):601–612. doi:10.1038/nrg3226.

53. Mellmann A, Harmsen D, Cummings CA, et al. Prospective Genomic Characterization of the German Enterohemorrhagic Escherichia coli O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. *PLoS ONE*. 2011;6(7):e22751. doi:10.1371/journal.pone.0022751.

APPENDICES

Appendix 1. FAM labeled AFLP haplotype codes prepared in lab.

Sample	Replicate	FAM 70s	FAM 80s	FAM 90s	FAM 100s	FAM 110s	FAM 120s	FAM 130s	FAM 140s	FAM 150s	FAM 160s	FAM 170s	FAM 180s	FAM 190s	FAM 200s	FAM 210s	FAM 220s	FAM 230s	FAM 240s	FAM 250s	FAM 260s	FAM 270s	FAM 280s	FAM 290s	FAM 300s	FAM 310s	FAM 320s	FAM 330s	FAM 340s
1	1	0	1	1	1	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
	2	1	2	2	2	2	0	2	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0
	3	0	1	2	2	2	0	2	0	1	0	0	1	0	0	0	1	0	0	1	0	0	1	1	1	0	1	1	0
2	1	1	2	2	2	2	1	2	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
	2	1	2	2	2	2	1	2	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	1	0	1	1	0
	3	1	1	1	2	2	0	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	1	0
3	1	0	1	0	1	1	0	1	0	0	0	0	2	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	0
	2	0	1	1	2	2	0	2	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
	3	0	1	1	1	2	0	2	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
4	1	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0
	2	0	1	1	2	2	0	2	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
	3	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0
	4	0	1	1	2	2	0	1	0	0	0	0	2	0	0	0	1	0	0	0	0	0	0	1	1	0	1	1	0
5	1	0	1	1	2	1	1	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
	2	0	1	1	2	1	1	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
	3	0	1	1	2	1	0	1	0	0	0	0	2	0	0	0	1	1	0	0	0	0	0	1	1	0	1	1	1
6	1	0	2	1	2	2	1	2	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
	2	0	2	0	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	1	0
	3	0	1	0	1	1	1	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0

Appendix 2. JOE labeled AFLP haplotype codes prepared in lab.

Sample	Replicate	JOE 70s	JOE 80s	JOE 90s	JOE 100s	JOE 110s	JOE 120s	JOE 130s	JOE 140s	JOE 150s	JOE 160s	JOE 170s	JOE 180s	JOE 190s	JOE 200s	JOE 210s	JOE 220s	JOE 230s	JOE 240s	JOE 250s	JOE 260s	JOE 270s	JOE 280s	JOE 290s	JOE 300s	JOE 310s	JOE 320s	JOE 330s	JOE 340s
1	1	1	1	0	1	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	1	1	1	1	2	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	1	1	0	1
	3	1	1	0	1	1	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
2	1	1	1	0	1	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	1	1	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	3	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
3	1	1	1	0	2	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	1	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	3	1	1	0	1	0	1	1	2	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	1	1	0	1
4	1	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	1
	3	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	1
	4	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
5	1	1	1	0	2	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	1	0	1	1	2	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	3	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
6	1	1	1	0	2	0	1	1	2	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	2	1	1	0	2	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1
	3	1	1	0	2	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	1	1	0	1	0	1	1	0	1

Appendix 3. NED labeled AFLP haplotype codes prepared in lab.

Sample	Replicate	NED 70s	NED 80s	NED 90s	NED 100s	NED 110s	NED 120s	NED 130s	NED 140s	NED 150s	NED 160s	NED 170s	NED 180s	NED 190s	NED 200s	NED 210s	NED 220s	NED 230s	NED 240s	NED 250s	NED 260s	NED 270s	NED 280s	NED 290s	NED 300s	NED 310s	NED 320s	NED 330s	NED 340s
1	1	0	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	2	0	0	1	0	0	0
	2	0	0	1	0	1	0	1	1	1	0	2	2	0	0	2	2	0	0	0	0	0	1	0	0	0	0	0	0
	3	0	0	1	0	1	0	1	1	1	0	2	2	0	0	2	2	1	0	0	0	2	2	0	0	1	0	1	0
2	1	0	0	1	0	1	0	1	1	1	0	2	2	0	0	2	2	1	0	0	0	2	2	0	0	1	0	1	1
	2	0	0	1	0	1	0	1	1	1	0	2	2	0	0	2	2	1	0	0	0	2	2	0	0	1	0	1	1
	3	1	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	2	0	0	1	0	1	0
3	1	2	1	1	0	2	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	1	0	0	0	0	0	1
	2	1	1	1	0	2	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	1	0	0	0	0	0	1
	3	1	1	1	0	2	0	2	1	1	1	1	2	0	0	2	1	1	0	0	0	2	1	0	0	0	0	0	1
4	1	1	0	1	0	2	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	1	0	0	0	0	0	1
	2	2	0	1	0	2	0	2	1	1	0	2	2	0	0	2	1	1	0	0	0	2	1	0	0	0	0	0	1
	3	1	0	1	0	2	0	2	1	1	0	2	2	0	0	2	2	1	0	0	0	2	1	0	0	0	0	0	1
	4	1	0	1	0	2	0	2	1	1	0	2	2	0	0	2	1	1	0	0	1	2	1	0	0	0	0	0	1
5	1	0	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	1
	2	0	0	1	0	1	0	2	2	1	0	2	2	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	2
	3	0	0	1	0	1	0	2	2	1	0	1	1	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	1
6	1	0	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	1
	2	1	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	1
	3	1	0	1	0	1	0	2	1	1	0	2	2	0	0	2	1	0	0	0	0	2	1	0	0	0	0	0	1

Appendix 4. AFLP haplotype codes labeled with FAM (blue), JOE (green), and NED (yellow) generated from reference genomes *in silico.*

	70s FAM	80s FAM	90s FAM	100s FAM	110s FAM	120s FAM	130s FAM	140s FAM	150s FAM	160s FAM	170s FAM	180s FAM	190s FAM	200s FAM	210s FAM	220s FAM	230s FAM	240s FAM	250s FAM	260s FAM	270s FAM	280s FAM	290s FAM	300s FAM	310s FAM	320s FAM	330s FAM	340s FAM
Agona SL483	1	2	6	6	4	2	3	4	4	2	0	0	3	5	0	2	4	3	2	4	2	3	0	2	1	1	3	0
Newport SL254	1	2	5	5	4	4	4	3	3	2	0	1	3	5	1	2	3	3	2	2	1	3	0	1	1	1	2	1
Typhimurium 14028S	2	2	5	6	4	3	5	3	3	2	0	0	3	5	0	2	3	2	2	3	1	3	0	1	1	0	1	1
Typhimurium D23580	1	2	5	6	4	3	5	3	3	2	0	0	3	5	0	2	3	2	2	3	1	3	0	1	1	0	2	1

	70s JOE	80s JOE	90s JOE	100s JOE	110s JOE	120s JOE	130s JOE	140s JOE	150s JOE	160s JOE	170s JOE	180s JOE	190s JOE	200s JOE	210s JOE	220s JOE	230s JOE	240s JOE	250s JOE	260s JOE	270s JOE	280s JOE	290s JOE	300s JOE	310s JOE	320s JOE	330s JOE	340s JOE
Agona SL483	6	3	3	2	2	4	4	4	4	3	1	0	0	2	2	1	3	1	3	1	0	1	3	1	3	1	3	0
Newport SL254	8	3	4	2	2	6	2	5	2	3	2	1	0	2	2	2	2	1	4	2	0	1	3	0	3	2	1	1
Typhimurium 14028S	7	3	4	2	3	4	4	4	2	4	2	1	0	3	0	1	2	1	3	2	0	1	4	1	3	2	2	1
Typhimurium D23580	7	3	4	3	3	4	3	4	2	4	2	1	0	3	0	1	2	1	3	2	0	2	4	1	3	2	1	1

	70s NED	80s NED	90s NED	100s NED	110s NED	120s NED	130s NED	140s NED	150s NED	160s NED	170s NED	180s NED	190s NED	200s NED	210s NED	220s NED	230s NED	240s NED	250s NED	260s NED	270s NED	280s NED	290s NED	300s NED	310s NED	320s NED	330s NED	340s NED
Agona SL483	4	1	3	2	2	3	0	1	4	2	2	2	4	2	2	1	4	2	2	1	1	2	3	0	1	0	0	1
Newport SL254	4	4	4	2	2	3	0	1	3	2	2	2	4	3	2	1	6	1	2	1	0	3	2	1	1	1	1	1
Typhimurium 14028S	3	3	3	3	1	3	0	2	3	2	2	2	4	2	2	1	5	1	2	1	1	3	2	0	1	0	0	1
Typhimurium D23580	3	3	3	3	1	3	0	1	4	2	2	2	4	2	2	1	5	1	2	1	1	3	2	0	1	0	0	1

VITA

Stephanie Kunz

Candidate for the Degree of

Master of Science

Thesis: EVALUATION OF AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP) ANALYSIS AS A METHOD FOR TRACKING *SALMONELLA ENTERICA* IN A FOODBORNE OUTBREAK

Major Field: Forensic Science

Biographical:

Education:

Completed the requirements for the Master of Science in Forensic Science at Oklahoma State University College of Health Sciences, Tulsa, Oklahoma in December 2013.

Completed the requirements for the Bachelor of Science in Investigative and Medical Sciences at Saint Louis University, St. Louis, Missouri in 2011.

Experience:
Completed 100 hours of practicum experience in the Biology and DNA section of the St. Louis Metropolitan Police Crime lab from October 2010 to May 2011