THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF FOOD SCIENCE

GENOMIC CLASSIFICATION OF CLINICAL ESCHERICHIA COLI O157:H7 STRAINS IN THE STATE OF PENNSYLVANIA

ANNETTE HARTZELL
Spring 2010

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Food Science
with honors in Food Science

Reviewed and approved* by the following:

Edward Dudley
Assistant Professor of Food Science
Thesis Supervisor

Gregory Ziegler
Professor of Food Science
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Objective

The purpose of this study was to examine populations of *E. coli* O157:H7 in Pennsylvania using a DNA-based method. *E. coli* O157:H7 is a highly virulent foodborne pathogen, and is an organism of interest for food producers, food safety specialists, and researchers. There are two main lineages of *E. coli* O157:H7, termed Lineage I and Lineage II, which differ in their virulence capacity. Lineage I strains are more frequently implicated in cases of human disease than are Lineage II strains. A third group, termed Lineage I/II, may represent a subset of *E. coli* O157:H7 with increased virulence potential.

Methodology

E. coli O157:H7 lineages can be differentiated by DNA based methods. One of these methods, termed the Lineage Specific Polymorphism Assay (LSPA), uses PCR and gel electrophoresis to identify size differences in 6 loci in the bacterial DNA. Based upon the size differences of these 6 regions, lineage types can be assigned. In this study, LSPA was used to determine the lineages of 52 *E. coli* O157:H7 strains isolated from patients in Pennsylvania during 2007 and 2008. The strains were obtained from the Pennsylvania Department of Health. Six sets of primers were designed to flank lineage-specific polymorphisms, and PCR was used to amplify these regions. Polyacrylamide gel electrophoresis was used to view the products. Bands from the test strains were compared to those of control LSPA type 111111 (Lineage 1) and 222222 (Lineage 2) strains. Test strain bands which matched those of the LSPA 111111 control were given a designation of “1”, and those which matched the bands of the LSPA 222222 control were given a designation of “2”. Test strain bands

matching neither Lineage I nor Lineage II bands were given a designation of “3”. Lineage I/II strains were identified as having the LSPA type 211111.

Results

Of the 52 strains, 63% were found to be Lineage I, 4% were Lineage II, 31% were Lineage I/II, and 2% were LSPA type 311111.

Significance and Implications

These findings were compared to those from studies on clinical *E. coli* O157:H7 isolates from other parts of North America. Some differences were found, indicating that there are regional differences in the distribution of *E. coli* O157:H7 lineages. However, high levels of Lineage I/II strains were found in both Pennsylvania and Michigan, indicating that this virulent strain is may be widespread in North America. This study supports a new model of *E. coli* O157:H7 virulence, indicating that there are two separate lineages, Lineage I and I/II, which have the capacity to cause serious disease in humans.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It is an honor for me to thank those who made the completion of my Schreyer Honors Thesis possible. First, thank you to my research advisor Dr. Edward Dudley for his continuous guidance and support during my laboratory work and thesis writing process. I am also grateful to Stanley Reynolds, Division Director of the Pennsylvania Department of Health, for supplying me with the bacterial strains for my study. Thank you to my lab-mate Kuanqing Liu, who did a wonderful job of getting me started on my project and teaching me lab techniques. I would also like to thank my lab-mates Chun Chen and Amrita Puri, who offered their help in the lab and made it a fun place to work. Finally, thank you to lab-mate Carrie Lewis, who will be continuing work on this project.

LITERATURE REVIEW

Escherichia coli

Escherichia coli are a group of bacteria which are found primarily in the intestines of vertebrate animals. Most strains live commensally in their hosts, but some have acquired virulence factors which allow them to cause disease in humans. There are two types of pathogenic *E. coli,* the diarrheagenic *E. coli* and the extraintestinal *E. coli* (Kaper et al. 2004). The diarrheagenic *E. coli* are divided into six groups, based upon their mode of pathogenicity. These groups include the enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC). *E. coli* O157:H7 is part of the enterohaemorrhagic (EHEC) group of *E. coli,* which attach to intestinal cells and disrupt the microvillar structure (Kaper et al. 2004). The bacteria first adhere to the epithelial cell membrane, and accumulate polymerized actin directly beneath them (Nataro and Kaper 1998). Secreted toxins are then absorbed into the cell, often resulting in hemorrhagic colitis (Hayashi et al. 2008).

E. coli O157:H7 Epidemiology and Ecology

Escherichia coli O157:H7 is a foodborne pathogen which has the ability to cause serious human disease. The organism was first recognized as a foodborne pathogen in 1983, during a ground beef patty outbreak (Riley et al. 1983), which involved 4 states, more than 700 illnesses, and a recall of more than 250,000 hamburgers from fast food chains (Rangel et al. 1995). Since then, this organism has been of increasing concern to food processors, food safety specialists and researchers, and to consumers. Between the years 1982 and 2002, 49 states in the United States reported 354 outbreaks, comprising 8,598 individual cases. Of these cases, 52 % were found to be foodborne (Rangel et al.

2005). The CDC estimates that approximately 70,000 people are infected by *E. coli* O157:H7 each year in the United States (CDC 2010).

Cattle are believed to be a major reservoir for *E. coli* O157:H7. This organism is often present in the intestines of cattle, and generally does not negatively affect their health (Faith Et al. 1996). It is estimated that 1.8 to 16% of cattle per herd in North America and Europe carry the organism (Byrne et al. 2003). Infected cattle shed *E. coli* O157:H7 in their feces, which can contaminate their hides and the environment. Beef may become contaminated with *E. coli* O157:H7 from the hides of the cattle during the slaughter process, making ground beef one of the most common sources of *E. coli* O157:H7 infection (Oswald et al. 2004). The organism has also been found in raw milk, most likely also contaminated by cattle feces (Ibekwe et al. 2006). Produce, such as lettuce and spinach, may become contaminated with the organism from the soil, possibly due to contaminated irrigation water (Ibekwe et al. 2006).

Molecular Mechanisms Contributing to Disease

E. coli infection can cause several very serious symptoms in humans. Symptoms can range from vomiting and diarrhea to hemorrhagic colitis, which is characterized by an inflamed colon and bloody diarrhea (Franke et al. 1995 and Ibekwe et al. 2006). In very advanced cases, hemolytic uremic syndrome may occur, leading to hemolytic anemia and kidney failure (Banatvala et al. 2001). Between 1982 and 2002, *E. coli* infections in the United States resulted in 1,483 hospitalizations, 354 hemolytic uremic syndrome cases, and 40 deaths (Rangel et al. 2005).

E. coli O157:H7 is a member of the Shiga toxin-producing *E. coli* (STEC). Shiga toxin is a cytotoxin that inhibits protein synthesis in infected cells. It is absorbed into the infected cells, and carried by the blood stream to the kidneys, where it can cause damage and hemolytic uremic syndrome (HUS) (Melton-Celsa and O'Brian 1998). There are two

antigenically distinct types of Shiga toxin, Stx 1 and Stx 2, which share approximately 55% of their amino acids (Andreoli et al. 2002). Stx 1 is identical to the Shiga toxin found in the bacteria *Shigella dysenteriae* type 1 (Xu et al. 1999), while Stx 2 has been shown to be the major cause of disease in *E. coli* O157:H7 infections (Boerlin et al. 1999). The genetic information for the Shiga toxins is located in lamboid prophage DNA within the *E. coli* chromosome (Melton- Celsa and O'Brian 1998).

Another major contributor to *E. coli* O157:H7 virulence is the Locus of Enterocyte Effacement, or LEE. The LEE is responsible for the attaching and effacing activity of the *E. coli* on the epithelial cells of the gastrointestinal tract (Nataro and Kaper 1998). The LEE contains the type III secretion system (TTSS), which facilitate the attachment of the bacteria to the host cell. Several secreted effector molecules function as a part of the TTSS. The LEE also encodes the adhesin intimin and the translocated intimin receptor (Tir). The TTSS translocates intimin and the Tir into the plasma membrane of the infected cell, allowing the *E. coli* O157:H7 to infect the host (Delahay et al. 2001).

E. coli O157:H7 Subtyping Methods

Within *E. coli* O157:H7, genetic differences exist that can be used to differentiate strains. One of the first methods devised was called Octamer-based Genome Scanning (OBGS) (Kim et al. 1999). OBGS uses PCR to amplify pieces of the *E. coli* genome that contained over-represented oligomers, are were skewed toward one DNA strand. Fluorescent PCR primers are used that target six over-represented octamers on the leading strand of the DNA. Following PCR, fragment analysis is performed on automated sequencers. Binary files are then created based on the presence and absence of fluorescent bands. The binary files of different strains of *E. coli* O157:H7 can then be compared, detecting small, single nucleotide differences between them (Kim et al. 1999).

OBGS was able to distinguish two main lineages of *E. coli* O157:H7, termed Lineage I and Lineage II (Kim et al. 1999). Researchers found that Lineage I strains were more often found in human clinical isolates, while Lineage II strains were more often isolated from cattle. The researchers tested strain sets from several locations in the United States, and found that the two lineages were geographically widespread. Findings from this study indicated that Lineage I strains may be a more virulent to humans than Lineage II strains. Restriction fragment length polymorphism assays (RFLPs) suggested that lysogenization, excision, recombination, and other genetic events involving prophages were responsible for the split between the two lineages (Kim et al. 1999).

In 2004, the same research group developed a more refined genotyping method, known as the Lineage Specific Polymorphism Assay (LSPA) (Yang et al. 2004). LSPA is the method utilized in the current study, described in Chapter 2. The LSPA assay uses differences in six loci in the *E. coli* O157:H7 genome to distinguish between Lineage I and Lineage II strains. Primers are prepared which flank six loci in the genome, located within both coding and non-coding regions. These six regions are amplified using PCR, and then viewed using polyacrylamide gel electrophoresis. By viewing the pattern of the six bands on the gel, lineage types can be assigned to *E. coli* O157:H7 strains (Yang et al. 2004). Test strains are run along with known Lineage I and Lineage II strains in order to compare amplicon sizes. Bands of the test strains are matched with those of the control. Bands which match those of the Lineage I control are given a designation of 1, while bands which match those of the Lineage 2 control are given the designation of 2. Bands not matching Lineage I or Lineage II strains are given a numerical designation of 3. In this way, a six-digit code is given to each of the test strains. The most common designation for Lineage I strains is 111111, while the most common designations for Lineage II strains are 222222 and 222211 (Yang et al. 2004). More recently, comparative genomics suggested that a

third lineage exists, with the LSPA genotype 211111. This lineage type was termed Lineage I/II (Zhang et al. 2007).

Another genotyping method that separates *E. coli* O157:H7 strains into nine distinct clades has been developed (Manning et al. 2008). This method uses real-time PCR to identify single-nucleotide polymorphisms (SNPs) between strains of *E. coli* O157:H7, and can be used to detect small differences between closely related strains of *E. coli* O157:H7. In the study which described this method, 528 strains of *E. coli* O157:H7 isolated from Michigan patients were screened for 32 SNPs using real-time PCR. Based upon the differences in the 32 SNPs, the strains could be placed into 9 distinct clades. Clades were found to differ in their Shiga toxin profile, their capability for causing disease, and in their frequency of isolation. The rates of hemolytic uremic syndrome (HUS) were analyzed among the patients infected with strains from the various clades. The results of this analysis suggested that patients with HUS were more likely to have been infected by a clade 8 strain of *E. coli* O157:H7 (Manning et al. 2008).

Researchers from the Dudley lab at Penn State University have developed a molecular subtyping method which can differentiate between *E. coli* O157:H7 strains using multilocus sequence typing (MLST). This study was performed in order to develop a more efficient alternative to the clade typing method described by Manning et al., and to determine whether there was a correlation between lineage classification of *E. coli* O157:H7 strains (Yang et al. 2004) and clade classifications (Manning et al. 2008). The study used MLST to analyze the rearrangement hot spot (*rhs*) genes of *E. coli* O157:H7 strains belonging to Lineages I and II, and 8 of the clades identified by Manning et al. Clade 5 was not included in this study because it had been suggested by previous research that clade 5 did not actually represent a distinct clade, but was instead based upon mixed cultures. This study showed that clades 1, 2, 3 and 4 fall into lineage I, while clades 6 and 8 fall into Lineage I/II, with the lineage classification 211111. Clade 7 strains were found to have the

lineage classifications 211111 and 222222, and clade 9 strains were found to have the lineage classifications 311111 and 212111 (Liu et al. 2009). This study was useful in providing a link between clade and lineage types, allowing studies using the two methods to be directly compared.

E. coli O157:H7 Subtype and Virulence Capacity

It has been shown that disease severity can vary among *E. coli* O157:H7 types. For example, two recent outbreaks in the United States, involving lettuce and spinach, resulted in unusually high percentages of patients developing HUS. The strain implicated was shown to be a part of clade 8, as designated by Manning et al.'s SNP clade typing method (Manning et al. 2008). This indicated that clade 8 strains may be more virulent to humans than other clades (Manning et al. 2008). Based upon the relationship between clade 8 and Lineage I/II strains demonstrated by Liu et al., it is suggested that clade 8 and Lineage I/II strains may represent an emerging group of *E. coli* O157:H7 with a heightened virulence capacity (Manning et al. 2008 and Liu et al. 2009).

Geography of *E. coli* O157:H7 Subtypes

A few studies have investigated the dissemination of clade and lineage types among groups of clinical *E. coli* O157:H7 strains in North America. Manning et al., the developers of the clade typing system, studied a group of more than 500 *E. coli* O157:H7 strains isolated from patients in Michigan. For the *E. coli* O157:H7 isolated in 2006, clades 2 and 8 were the most frequently identified. Approximately 38% of the strains studied belonged to clade 2, and approximately 46% of the strains belonged to clade 8. The rates of *E. coli* O157:H7 strains isolated from the other clades was much lower. Together, strains from other clades made up only 15% of the total number isolated (Manning et al. 2008). This study also looked at the changes in clade distribution over time, between 2001 and 2006. The increase in the percentage of clade 8 strains isolated is particularly notable. Between

2002 and 2006, the percentage of clade 8 strains isolated from clinical cases increased from 10% to 46% of total cases (Manning et al. 2008). This data suggests that this virulent clade may be increasing in frequency of isolation.

Another study investigated the LSPA types of *E. coli* O157:H7 strains isolated from 313 cattle and 203 humans, in northern and southern Alberta, Canada. This study showed that 72.2% of the *E. coli* O157:H7 strains isolated from cattle were LSPA typed 111111, 19.4% of the strains were LSPA typed 222222, and 1.6% were LSPA type 211111 (Sharma et al. 2008). Of the human clinical isolates, 90.1% were LSPA typed 111111, and 3.1% were LSPA typed 222222. The percent of Lineage I/II, LSPA type 211111, strains among the human clinical isolates was 3.9% (Sharma et al. 2008). This number was much lower than percentage reported by Manning et al., which stated that in 46% of the Michigan *E. coli* O157:H7 strains studied belonged to clade 8 (Manning et al. 2008). Based on the relationship between clades and lineages described by Liu et al., it can be assumed that these clade 8 strains would also be typed Lineage I/II by the LSPA assay (Liu et al. 2009). The differences between these two studies suggest geographic differences in the dissemination of clade 8 and Lineage I/II strains of *E. coli* O157:H7.

Current Study

The current study seeks to investigate clinical *E. coli* O157:H7 strains isolated from patients in Pennsylvania in order to determine their lineage types. The strains were collected during 2007 and 2008, and typed using the LSPA assay (Yang et al. 2004). The frequency of Lineage I, Lineage II, and Lineage I/II strains were compared with results from similar studies in North America. Given discrepancies between the Manning and Sharma studies, the researchers were especially interested in investigating the dissemination of Lineage I/II strains. There is evidence that Lineage I/II may be an especially virulent group of strains, which may be increasing in frequency of isolation in some parts of North America.

INTRODUCTION

E. coli O157:H7 was first identified as a foodborne pathogen in 1983, during an outbreak involving contaminated ground beef patties (Riley et al. 1983). Since that time, *E. coli* O157:H7 has become an increasingly common agent concern for food producers, food safety specialists, and researchers. This organism is highly virulent and can cause debilitating and deadly diseases, including hemorrhagic colitis and hemolytic uremic syndrome (HUS) (Riley et al. 1983).

E. coli O157:H7 is a member of the STEC, or Shiga toxin-producing group of *E. coli.* Shiga toxins act as inhibitors of protein production in the cells they infect. Members of the STEC produce one of both of two antigenically distinct forms of Shiga toxin, termed Stx1 and Stx2. Stx1 is identical to the Shiga toxin found in the bacteria *Shigella dysenteriae* type 1. Within the STEC, *E. coli* O157:H7 is a member of the subset EHEC, or enterohemhorrhagic *E. coli.* EHEC is the designation generally given to STEC strains that cause serious disease in humans (Xu et al. 1999). Besides the Shiga toxins, EHEC also carry other virulence genes, including the locus of enterocyte effacement (LEE) pathogenicity island (McDaniel et al. 1995).

E. coli O157:H7 is a normal part of the gut flora of many cattle, and the organism generally does not affect their health (Faith et al. 1996). In North America and Europe, it is estimated that 1.8 to 16% of cattle per herd carry the organism (Byrne et al. 2003). The cattle shed *E. coli* O157:H7 in their feces, which can contaminate the hides of the cattle and their environment (Faith et al. 1996). Ground beef is one of the most common sources of *E. coli* O157:H7, but other foods such as raw milk, lettuce, fruit juices, and spinach have also been implicated in outbreaks (Ibekwe et al. 2006 and Jay et al. 2007).

There are two main lineages of *E. coli* O157:H7, termed Lineage I and Lineage II, which can be differentiated DNA based methods (Yang et al. 2004 and Kim et al.

1999). Octamer-based genome scanning was the method that originally defined Lineage I and II strains (Kim et al. 1999). This method uses 8-bp primers in PCR reactions, generating a large number of amplicons, and differences between strains can be visualized by gel electrophoresis. It was noted previously that Lineage I strains were more frequently isolated from clinical cases than were Lineage II strains (Kim et al. 1999).

A more refined genotyping method, termed the Lineage Specific Polymorphism Assay (LSPA), was later developed by the same group (Yang et al. 2004). The LSPA assay can quickly and easily distinguish between strains belonging to Lineage I and Lineage II subpopulations of *E. coli* O157:H7. This assay uses PCR and gel electrophoresis to determine DNA size differences at 6 genomic loci, located within both coding and non-coding regions. Each of the genetic loci generally exhibit one of two amplicon sizes, and a 6-digit binary designation is accordingly assigned to strains. The most common designation for Lineage I strains is 111111. For Lineage II strains, LSPA genotypes 222222 and 222211 are commonly observed. Bands not matching Lineage I or Lineage II sizes are given a numerical designation of 3 (Yang et al. 2004). More recently, comparative genomic analysis suggested the presence of a third lineage, designated Lineage I/II (LSPA type 211111) (Zhang et al. 2007).

In addition to the lineage designation, a 32-locus, single nucleotide polymorphism (SNP) method has been described that separates clinical isolates of *E. coli* O157:H7 into one of 9 phylogenetic groups designated as clades. Strains between each clade may differ in their Shiga toxin profile, their capability for causing disease, and in their frequency of isolation (Manning et al. 2008).

Recent research has suggested a relationship between the clade and LSPA type of *E. coli* O157:H7 strains, with specific clades matching with specific lineages (Liu et al. 2009). The Dudley lab at Penn State University has reported on the correlations between the

lineage and clade designations (Liu et al. 2009). This work showed that clades 1, 2, 3, and 4 fall into Lineage I, while clades 6 and 8 fall into Lineage I/II (Liu et al. 2009). Clade 5 was not included in the study because previous research had suggested that the original classification of clade 5 was actually based upon mixed cultures, and did not constitute a distinct group (Liu et al. 2009). Clade 7 strains were found to have the lineage classifications 211111 and 222222, and clade 9 strains were found to have the lineage classifications 311111 and 212111 (Liu et al. 2009). Based upon the high virulence capacity of clade 8 strains of *E. coli* O157:H7 and the relationship between the clade 8 and Lineage I/II groupings, it is postulated that severe cases of disease in humans can be caused by both Lineage I and Lineage I/II strains (Liu et al. 2009).

There has been speculation that disease severity can vary among clinical isolates of *E. coli* O157:H7 strains typically associated with clinical symptoms (Manning et al. 2008). For example, two recent outbreaks in the United States, involving lettuce and spinach, resulted in unusually high percentages of patients developing Hemolytic Uremic Syndrome (HUS). The strain implicated in one of the outbreaks was part of a group of *E. coli* O157:H7 designated clade 8 by the 32 SNP-typing method (Manning et al. 2008). It was also found that the frequency of Michigan *E. coli* O157:H7 cases caused by clade 8 strains increased from 10% in 2002, to 46% in 2006. The study suggested that clade 8 may be an emerging group of *E. coli* O157:H7 which has a heightened virulence capacity (Manning et al. 2008).

The purpose of this study was to utilize the LSPA assay to examine 52 *E. coli* O157:H7 strains isolated from patients in the state of Pennsylvania. The frequency of Lineage I, Lineage II, and Lineage I/II strains was compared to results of studies from other locations in North America, in order to determine geographic differences in the dissemination of the various lineages. It was found that there were differences in the proportion of lineages found between the locations studied in North America. This

information was useful in understanding the emerging role of Lineage I/II strains of *E. coli* O157:H7 as important agents of infection.

MATERIALS AND METHODS

Bacterial strains and growth conditions. All *E. coli* O157:H7 strains used in this study are listed in Table 1. The 52 test strains were isolated from infected patients in Pennsylvania, and obtained from the Pennsylvania Department of Health. The Lineage I/II strains were clade-typed using a single-nucleotide polymorphism assay. All Lineage I/II strains except for three (8, 9 and 38) were found to be clade 8 (Chen, unpublished data).

The strains Sakai and FRIK920 were used as positive controls in the experiments. Sakai and FRIK920 are typical Linage I and II strains of *E. coli* O157:H7, respectively (Liu et al. 2009). Sakai was obtained from Dr. Wei Zhang from the Illinois Institute of Technology, and Frik920 was obtained from Dr. Andrew Benson from the University of Nebraska-Lincoln. All strains were stored at -80°C in 10% glycerol. When needed, strains were grown on sorbitol MacConkey agar plates made of Difco MacConkey agar base (Becton, Dickinson and Company, Sparks, MD), supplemented with 1% (wt/vol) D-sorbitol (Alfa Aesar, Ward Hill, MA). Streaked plates were incubated at 37°C for 18 to 24 hours.

TABLE 1: *E. coli* O157:H7 Test Strains

Strain #	County	Source	Collection Date	Serotype Result	Source Type
1	Allegheny	Stool	7/15/2008	*E. coli* O157:H7	Human
2	Butler	Stool	6/12/2008	*E. coli* O157:H7	Human
3	Berks	Stool	10/13/2007	*E. coli* O157:H7	Human
4	Berks	Stool	8/6/2007	*E. coli* O157:H7	Human
5	Northumberland	Stool	7/15/2007	*E. coli* O157:H7	Human
6	York	Stool	7/4/2007	*E. coli* O157:H7	Human
7	Lancaster	Stool	3/13/2007	*E. coli* O157:H7	Human
8	Lancaster	Stool	6/20/2007	*E. coli* O157:H7	Human

9	York	Stool	7/29/2008	*E. coli* O157:H7	Human
10	Crawford	Stool	8/4/2008	*E. coli* O157:H7	Human
11	Montour	Stool	10/20/2007	*E. coli* O157:H7	Human
12	Indiana	Stool	10/27/2007	*E. coli* O157:H7	Human
13	Berks	Stool	6/21/2008	*E. coli* O157:H7	Human
14	Westmoreland	Stool	4/16/2007	*E. coli* O157:H7	Human
15	York	Stool	3/31/2007	*E. coli* O157:H7	Human
16	Lancaster	Stool	10/17/2007	*E. coli* O157:H7	Human
17	Allegheny	Stool	11/6/2007	*E. coli* O157:H7	Human
18	York	Stool	6/8/2007	*E. coli* O157:H7	Human
19	Cambria	Stool	11/17/2007	*E. coli* O157:H7	Human
20	Lancaster	Stool	7/5/2008	*E. coli* O157:H7	Human
21	York	Stool	7/27/2008	*E. coli* O157:H7	Human
22	Bucks	Stool	5/16/2007	*E. coli* O157:H7	Human
23	York	Stool	7/7/2008	*E. coli* O157:H7	Human
24	Centre	Stool	8/29/2007	*E. coli* O157:H7	Human
25	Northumberland	Stool	7/29/2007	*E. coli* O157:H7	Human
26	Lancaster	Stool	7/14/2007	*E. coli* O157:H7	Human
27	Montgomery	Stool	9/12/2008	*E. coli* O157:H7	Human
28	Crawford	Stool	9/20/2008	*E. coli* O157:H7	Human
29	Wayne	Stool	7/25/2008	*E. coli* O157:H7	Human
30	Butler	Stool	7/20/2008	*E. coli* O157:H7	Human
31	Centre	Stool	4/4/2007	*E. coli* O157:H7	Human
32	Beaver	Stool	3/26/2007	*E. coli* O157:H7	Human
33	Centre	Stool	4/13/2007	*E. coli* O157:H7	Human

34	York	Stool	6/19/2007	*E. coli* O157:H7	Human
35	Berks	Stool	1/18/2007	*E. coli* O157:H7	Human
36	Allegheny	Stool	12/11/2006	*E. coli* O157:H7	Human
37	Bucks	Stool	1/21/2007	*E. coli* O157:H7	Human
38	Beaver	Stool	1/30/2007	*E. coli* O157:H7	Human
39	Centre	Stool	2/13/2007	*E. coli* O157:H7	Human
40	Philadelphia	Stool	4/8/2008	*E. coli* O157:H7	Human
41	York	Stool	8/15/2008	*E. coli* O157:H7	Human
42	Centre	Stool	7/11/2008	*E. coli* O157:H7	Human
43	York	Stool	5/28/2008	*E. coli* O157:H7	Human
44	Delaware	Stool	7/1/2008	*E. coli* O157:H7	Human
45	Allegheny	Stool	7/22/2008	*E. coli* O157:H7	Human
46	York	Stool	1/4/2008	*E. coli* O157:H7	Human
47	Berks	Stool	11/4/2008	*E. coli* O157:H7	Human
48	Berks	Stool	6/11/2008	*E. coli* O157:H7	Human
49	Chester	Stool	9/2/2008	*E. coli* O157:H7	Human
50	Centre	Stool	9/8/2008	*E. coli* O157:H7	Human
51	Chester	Stool	9/13/2008	*E. coli* O157:H7	Human
52	Delaware	Stool	7/31/2008	O157:H7	Human

Primer Design. Primers were designed according to Yang et. al, flanking lineage-specific polymorphisms (2004). Primers are shown in Table 2.

TABLE 2: PCR Primers (Yang et al. 2004)

Primer Name	Type	Sequence	Tm °C
folD-sfmA	Forward	TACGTAGGTCGAAGGG	51.0
	Reverse	CCAGATTTACAACGCC	51.5
Z5935	Forward	GTGTTCCCGGTATTTG	50.9
	Reverse	CTCACTGGCGTAACCT	50.3
yhcG	Forward	CTCTGCAAAAAACTTACGCC	50.3
	Reverse	CAGGTGGTTGATCAGCG	50.3
rbsB	Forward	AGTTTAATGTTCTTGCCAGCC	51.2
	Reverse	ATTCACCGCTTTTTCGCC	51.1
rtcB	Forward	GCGCCAGATCGATAAAGTAAG	51.3
	Reverse	GCCGTTGTAAACGTGATAAAG	50.3
arp-iclR	Forward	GCTCAATCTCATAATGCAGCC	51.7
	Reverse	CACGTATTACCGATGACCG	50.1

Tm: Melting Temperature

PCR and LSPA Assay. The LSPA assay was carried out using several primer combinations in order to maximize the efficiency of the amplifications. Reactions were performed in one multiplex PCR according to Yang et. al (2004). The six-amplicon multiplex did not always result in amplification of all target sites. Often, a 4-plex reaction using *yhcG, rbsB, rtcB,* and *arp-iclR* was performed, and *folD-sfma* and *Z5935* were placed into either 2-plex or single reactions.

Template DNA was prepared by suspending one colony from an overnight culture in water, making a final volume of 20 µl. For each 20 µl reaction, 1 µl of template was added. Concentrations of the PCR reagents and primers are shown in Table 4. A higher concentration of the primer pair *Z5935* was used as this was found to increase the amplification efficiency.

TABLE 3: PCR Reagent and Primer Concentrations

PCR Reagent/Primer	Final concentration
dNTPs	5 mM
10x ThermoPol Buffer	1x
Taq polymerase	0.02 U/ml
Primer pairs *folD-sfmA, yhcG, rbsB, rtcB, arp-iclR*	0.5 µM each
Primer pair *Z5935*	5 µM

PCR thermocycler conditions are shown in Table 4.

TABLE 4: LSPA PCR Thermocycler Conditions (Liu et al. 2009)

PCR Step	Temperature	Time	Repeats
Initialization	94°C	5 min	1x
Denaturation	94°C	30 sec	30x
Annealing	52.2°C	30 sec	30x
Extension	72°C	45 sec	30x
Final extension	72°C	10 min	1x
Final hold	4°C	∞	1x

After the PCR reaction was completed, loading dye (0.012% bromphenol blue-0.1 mM EDTA, [pH 8.0] in 100% formamide) was added. The PCR products were loaded onto a 6% polyacrylamide gel. When LSPA was performed in two or three reactions, PCR products from the reactions were added to the same well. The *E. coli* O157:H7 strains Sakai and FRIK920 were used as LSPA 111111 and LSPA 222222 controls, respectively. The gel was run at 80 volts for 999 minutes on a Bio-Rad DGGE. The DNA fragments were stained using ethidium bromide, and visualized using an EC3 500 BioImaging system (UVP, Upland, CA).

Data analysis. Bands from the test strains were compared to those of the LSPA 111111 and 222222 controls. Test strain bands which matched those of the LSPA 111111 control were given a designation of "1", and those which matched the bands of the LSPA 222222 control were given a designation of "2". Test strain bands matching neither Lineage I nor II bands were given a designation of "3". In this manner, a 6 digit code was developed for each test strain. The order of the bands on the gel was, top to bottom, *yhcG, arp-iclR, rtcB, rbsB, folD-sfmA, Z5395.* The order of bands in the coding system was *folD-sfmA, Z5935, yhcG, arp-iclR, rtcB, rbsB*.

RESULTS

The objective of this study was to use the Lineage Specific Polymorphism Assay (LSPA) to determine the lineages of 52 *E. coli* O157:H7 strains isolated from patients in Pennsylvania, and compare results to studies performed in other locations in North America. Manning et al. had shown that 46% of *E. coli* O157:H7 strains isolated from Michigan patients in 2006 belonged to clade 8 (Manning et al. 2008). It was later shown that clade 8 strains fall into Lineage I/II (Liu et al. 2009). Together, this data suggested a notable role of Lineage I/II strains in causing human disease.

E. coli O157:H7 lineages can be differentiated between by DNA based methods. The Lineage Specific Polymorphism Assay (LSPA) uses PCR and gel electrophoresis to identify size differences in 6 loci in the bacterial DNA. Based upon the size differences of these 6 regions, lineage types can be assigned. In this study, LSPA was used to determine the lineages of 52 *E. coli* O157:H7 strains isolated from patients in Pennsylvania.

52 clinical *E. coli* O157:H7 strains were obtained from the Pennsylvania Department of Health, which had been collected from patients during 2007 and 2008. Figure 1 shows a map of the 67 counties in Pennsylvania, and the number of Lineage I strains included in this study from each county. Figure 2 shows the number of Lineage I/II strains isolated from the counties. Clusters of *E. coli* O157:H7 infections were observed around areas with the highest population density, such as Pittsburgh and Philadelphia.

FIGURE 1: *E. coli* O157:H7 Lineage I strains isolated from counties in Pennsylvania

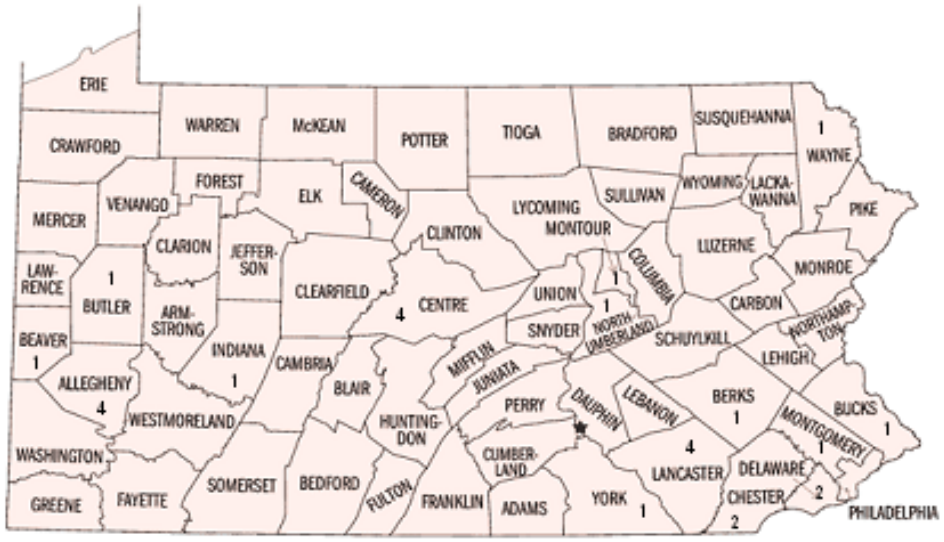

Figure 1: Distribution of E. coli O157:H7 Lineage I strains included in this study by Pennsylvania county.

FIGURE 2: *E. coli* O157:H7 Lineage I/II strains isolated from counties in Pennsylvania

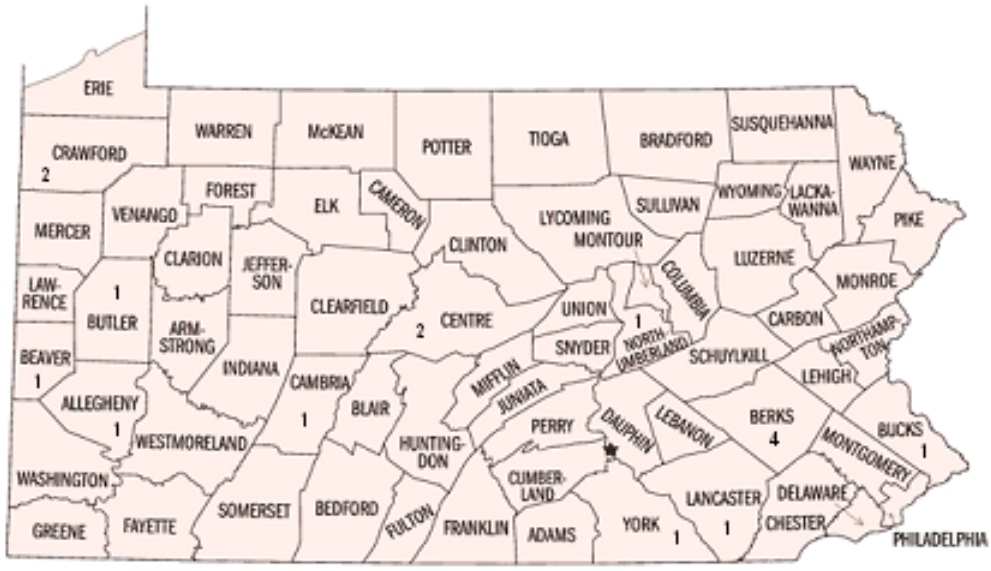

Figure 2: Distribution of E. coli O157:H7 Lineage I/II strains included in this study by Pennsylvania county.

LSPA typing of 52 Pennsylvania strains. LSPA types were determined by comparing the sizes of bands amplified in the test strains to those amplified in the Lineage I

and II controls, Sakai and FRIK 920. A sample gel, illustrating the lineage typing method, is shown in Figure 3. The bands observed, from largest to smallest, were those amplified by the primers *yhcG, arp-iclR, rtcB, rbsB, folD-sfmA,* and *Z5395.* Band sizes are also shown.

FIGURE 3: Sample gel for Lineage classification

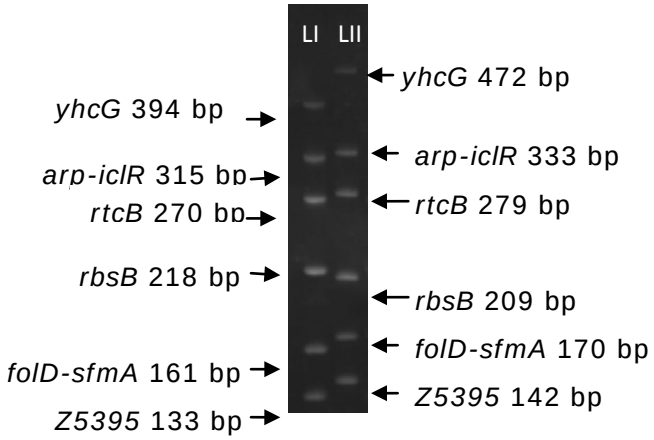

Figure 3: Lineage I and Lineage II LSPA products of *E. coli* O157:H7 appear on a polyacrylamide gel. The names of the primers used and the sizes of the bands are shown.

Test strain bands were given numerical designations, based upon whether they matched with bands the Lineage I or Lineage II control, or neither. All strains were given a six number designation. The strain characteristics are shown in Table 5. Lineage classification and *Xba*I and *Bln*I PFGE patterns are listed. Strains with the same lineage type or PFGE pattern are color coded. Most of the strains had different PFGE patterns, but those that did have the same PFGE pattern also had the same lineage type. An example of this would be strains 5 and 37. Some strains were shown to be the same using one restriction enzyme, and different using the other. Strains 9 and 39 are an example of this, representing one of the major drawbacks of PFGE typing. PFGE typing indicates that some of the strains tested may be parts of possible outbreaks. For example, strains 5, 31, 14, and 15 may be part of an outbreak, according to *BlnI* restriction. There were 32 Lineage I, 15 Lineage I/II, three Lineage II, and one Other strain, with the lineage type 311111 found. References are also

given for appendix figures, where photographs of the gels showing the lineage types of the strains are shown.

TABLE 5: *E. coli* O157:H7 Lineage Classifications of 52 Test Strains

Strain #	County	Lineage Class.	PFGE XbaI Pattern	PFGE BlnI Pattern	Append. Ref. Fig.	Strain #	County	Lineage Class.	PFGE XbaI Pattern	PFGE BlnI Pattern	Append. Ref. Fig.
1	Allegheny	111111	EXHX01. 2612	EXHA26. 2618	4	**27**	Montgomery	111111	EXHX01. 0238	EXHA26. 1806	3
2	Butler	211111	EXHX01. 0110	EXHA26. 0536	6, 7	**28**	Crawford	211111	EXHX01. 1516	EXHA26. 0420	6, 7
3	Berks	211111	EXHX01. 0047	EXHA26. 0015	6, 7	**29**	Wayne	111111	EXHX01. 3249	EXHA26. 0628	6, 7
4	Berks	111111	EXHX01. 4241	EXHA26. 0238	4	**30**	Butler	111111	EXHX01. 0248	EXHA26. 0569	6
5	Northumb erland	111111	EXHX01. 0797	EXHA26. 0838	1	**31**	Centre	111111	EXHX01. 0495	EXHA26. 0838	3
6	York	111111	EXHX01. 0902	EXHA26. 1348	1	**32**	Beaver	111111	EXHX01. 1348	EXHA26. 2159	3
7	Lancaster	111111	EXHX01. 4004	EXHA26. 2150	1	**33**	Centre	111111	EXHX01. 0008	EXHA26. 0569	3
8	Lancaster	211111	EXHX01. 2224	EXHA26. 0718	1	**34**	York	211111	EXHX01. 0224	EXHA26. 0536	6, 7
9	York	211111	EXHX01. 0224	EXHA26. 0536	1	**35**	Berks	211111	EXHX01. 1486	EXHA26. 0071	6, 7
10	Crawford	211111	EXHX01. 1531	EXHA26. 2830	1	**36**	Allegheny	111111	EXHX01. 3776	EXHA26. 1354	3
11	Montour	111111	EXHX01. 0079	EXHA26. 1080	1	**37**	Bucks	111111	EXHX01. 0797	EXHA26. 0838	3, 5
12	Indiana	111111	EXHX01. 0090	EXHA26. 2240	1	**38**	Beaver	211111	EXHX01. 0272	EXHA26. 2129	5
13	Berks	211111	EXHX01. 0125	EXHA26. 0570	4	**39**	Centre	211111	EXHX01. 0224	EXHA26. 0742	5
14	Westmore land	111111	EXHX01. 0495	EXHA26. 0838	4	**40**	Philadelphia	222211	EXHX01. 0352	EXHA26. 1515	3, 5
15	York	111111	EXHX01. 0495	EXHA26. 0838	4	**41**	York	222222	EXHX01. 4553	EXHA26. 2881	5

16	Lancaster	111111	EXHX01.0074	EXHA26.1349	4	**42**	Centre	111111	EXHX01.1343	EXHA26.0621	2
17	Allegheny	111111	EXHX01.0248	EXHA26.0014	4	**43**	York	111111	EXHX01.0225	EXHA26.0621	5, 2: LI Control
18	York	111111	EXHX01.2382	EXHA26.0569	4	**44**	Delaware	111111	EXHX01.1271	EXHA26.0842	5, 2: LI Control
19	Cambria	211111	EXHX01.0154	EXHA26.0556	6, 7	**45**	Allegheny	111111	EXHX01.0263	EXHA26.1835	7
20	Lancaster	111111	EXHX01.2357	EXHA26.2735	6, 7	**46**	York	111111	EXHX01.0349	EXHA26.0508	7
21	York	111111	EXHX01.3248	EXHA26.2799	4	**47**	Berks	211111	EXHX01.0047	EXHA26.0015	7
22	Bucks	211111	EXHX01.4108	EXHA26.2182	7	**48**	Berks	311111	EXHX01.1899	EXHA26.2180	5
23	York	111111	EXHX01.2358	EXHA26.0257	4	**49**	Chester	111111	EXHX01.1731	EXHA26.2904	5
24	Centre	211111	EXHX01.0574	EXHA26.1422	6	**50**	Centre	111111	EXHX01.0450	EXHA26.2814	5
25	Northumberland	211111	EXHX01.0124	EXHA26.0015	6	**51**	Chester	111111	EXHX01.4577	EXHA26.0576	5
26	Lancaster	111111	EXHX01.4213	EXHA26.2325	3	**52**	Delaware	111111	EXHX01.0008	EXHA26.0549	5

Lineage types listed as percentages are shown in Figure 4. 63% of the strains in the study were Lineage I, 4% were Lineage II, 31% were Lineage I/II, and 2% were designated as Other due to the presence of a band which did not match with any bands typical of Lineage I or II strains. This strain determined to be LSPA type 311111.

FIGURE 4: Distribution of Lineage classifications

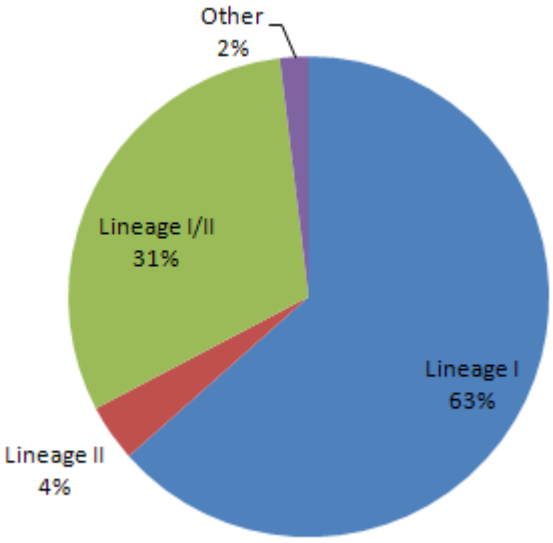

Figure 4: Percentages by lineage type of the 52 strains studied. Lineage I represents LSPA type 111111, Lineage II represents LSPA type 222222, Lineage I/II represents LSPA type 211111, and Other represents LSPA type 311111

DISCUSSION

The LSPA assay is used to distinguish between lineages of *E. coli* O157:H7. PCR is used to amplify 6 genomic loci, which are then viewed by gel electrophoresis. Each band is labeled with a 1 or a 2, depending on whether it matches with that of a known Lineage I or Lineage II strain (Yang et al. 2004). Lineage I has been found to more frequently cause human disease than Lineage II (Yang et. al 2004). Recently, however, strains with the designation 211111, termed Lineage I/II strains, have been implicated in serious cases of disease (Manning et al. 2008 and Liu et al. 2009). The purpose of this study was to use the LSPA assay to assign lineage classifications to 52 *E. coli* O157:H7 strains recently isolated from patients in the state of Pennsylvania. It was discovered that 63% of the strains were Lineage I, 4% were Lineage II, 31% were Lineage I/II, and 2% had the designation 311111. Differences were found across studies of *E. coli* O157:H7 lineage prevelance in North America. This indicates that there are possible geographic differences in the prevalence of disease-causing lineages of *E. coli* O157:H7.

In this study, the greatest number of strains were typed as Lineage I. This was expected, as Lineage I is the most common lineage to be identified in clinical isolates (Kim et al. 1999). Two Lineage II strains were identified among the 52 strains studied. It is unusual for Lineage II strains to cause infection in humans, but not unheard of. The researchers who developed the LSPA assay did find a small (~1%) number of Lineage II strains in their pool of human clinical isolates (Yang et al. 2004) and researchers studying strains from Alberta, Canada, found that between 1.4 and 4.8% of isolates from patients were Lineage II (Sharma et al. 2008). There was one strain found in the current study which had the lineage type 311111. This strain was determined to be sorbitol negative by growth on Sorbitol MacConkey Agar (SMAC) plates, and in M9 minimal media supplemented with sorbitol as the sole carbon source (Data not shown). Sorbitol negativity is one of the

key markers of pathogenic *E. coli* O157:H7. It is unusual to find sorbitol negative bacteria of this lineage type.

Of the strains analyzed, 31% belonged to Lineage I/II. The Lineage I/II were clade typed, and 12 of the 15 strains were found to be clade 8, or 23% of the total number of strains. These numbers were slightly lower than the percentage found in Michgan in 2006, with 46% of strains reported by Manning et al. to be clade 8 (Manning et al. 2004). Clade 8 strains are included in the lineage type I/II (Liu et al. 2009). Therefore, it can be assumed that the clade 8 strains identified by Manning et al. belong to Lineage I/II. The 31% and 23% figures derived from the current study are, however, notably higher than Manning et al.'s 2001 percentage, which placed the number of clade 8 strains isolated from Michigan patients at 10% (Manning et al. 2008). One question our study does not answer is whether Lineage I/II and/or Clade 8 strains have been increasing in frequency among clinical cases in Pennsylvania, and such experiments will be planning in the future.

Not all other studies on the prevalence of the various lineages of *E. coli* O157:H7 agree with the current one. Researchers studying clinical isolates from Northern and Southern Alberta, Canada have found that a much smaller percentage of strains fell into Lineage I/II (Sharma et al. 2008). This study found that 3.2% of the strains from Northern Alberta, and 4.2% of the strains from Southern Alberta were Lineage I/II strains. In addition, the percentage of Lineage I strains these researchers reported, 88.7% from Northern Alberta and 92.9% from Southern Alberta, was also higher than the 63% reported in the current study (Sharma et al. 2008).

There are several possible reasons for the differences in the proportion of lineages found between Pennsylvania, Michigan and Alberta, Canada. The discrepency could be due to random geographic differences in the lineage types most prevalent in different parts of the country. For instance, Lineage I/II strains of *E. coli* O157:H7 may be more common

overall in Pennsylvania and Michigan, than in Alberta, Canada. Alternatively, virulence differences may exist within the Lineage I/II strains found in Pennsylvania and Michigan, allowing them to develop traits which make them more virulent than those found in Alberta, Canada. Comparative genomics of spatially distinct Lineage I/II strains might help to determine whether the number of observed differences in distribution is related to changes in genetic makeup. Screening strains for virulence factors could further help to determine whether these strains carry different packages of virulence genes.

CONCLUSIONS AND FUTURE DIRECTIONS

This study determined that Lineage I/II is a notable source of human *E. coli* O157:H7 infection in Pennsylvania. By looking at similar studies performed elsewhere in North America, it was determined that geographical differences exist regarding the dissemination of *E. coli* O157:H7 strains, especially those belonging to Lineage I/II.

To determine why regional differences exist in *E. coli* O157:H7 populations, several future experiments can be conducted. First, the number of Pennsylvania strains screened should be increased in order to obtain more complete data on the lineage percentages throughout the state. Approximately 500 strains should be screened, to remain in line with the Michigan study performed by Manning et al. (Manning et al. 2008). The strains studied should be chosen based upon a set time frame, in order to gain an accurate representation of strain concentration and type throughout the state.

Second, similar studies should be conducted throughout North America and the rest of the world. More data points are necessary in order to gain a better understanding of how regional differences may affect the distribution of *E. coli* O157:H7 lineage types. To the best knowledge of the authors, only three studies, including this one, have investigated the distribution of either clade or lineage types in North America. It is recommended that research be carried out in each state in the United States, and in each territory in Canada. Lineage distributions throughout the rest of the world should also be investigated.

Third, genetic screening for virulence factors could be performed on Lineage I/II strains. Virulence genes found could be compared to those present in Lineage I and Lineage II bacteria. Screening for Stx and LEE genes could help to determine whether changes in the bacteria's DNA are linked to virulence in Lineage I/II strains of *E. coli* O157:H7.

APPENDIX

Figure S.1

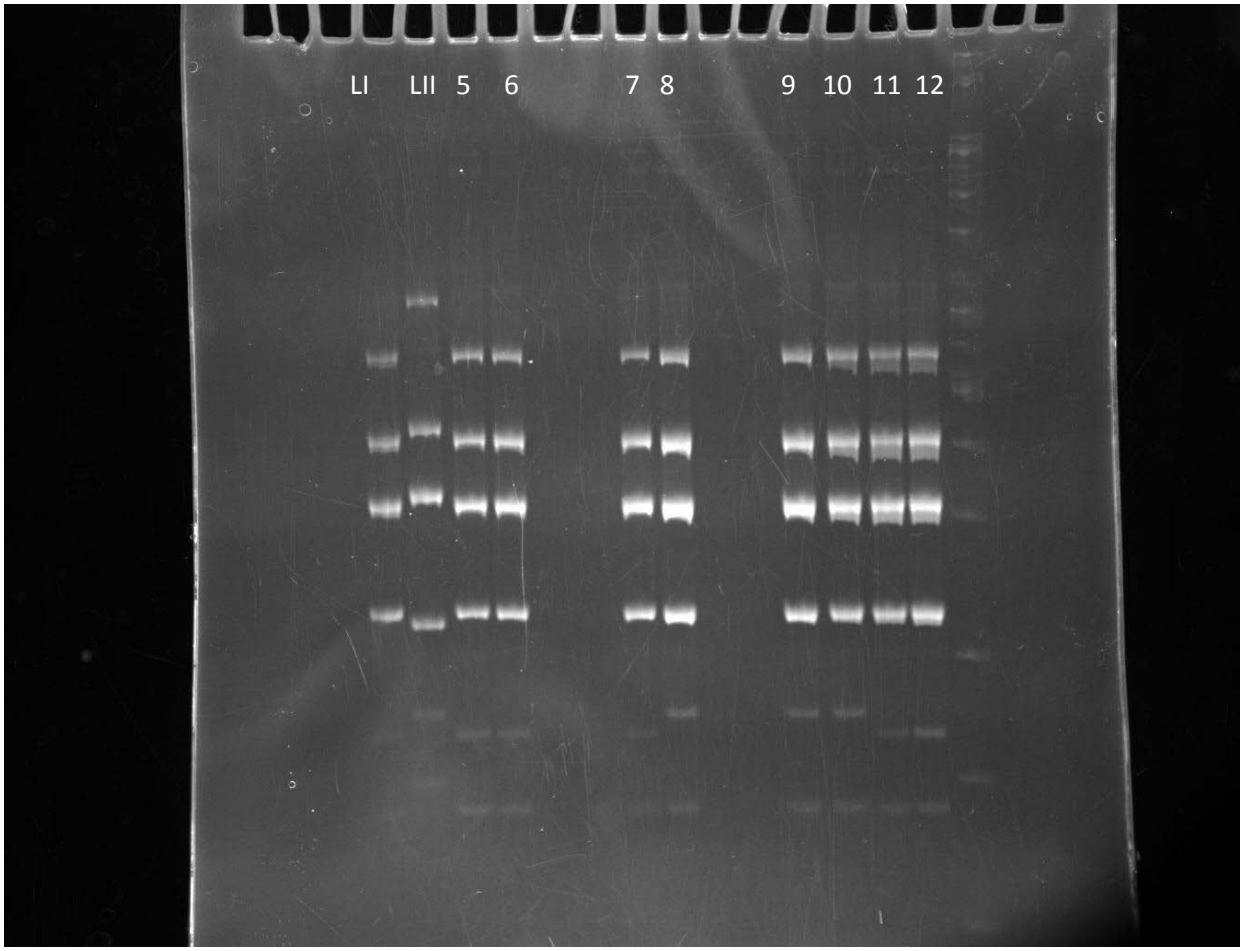

Figure S 1: Banding patterns of strains 5, 6, 7, 8, 9, 10, 11 and 12

Figure S.2

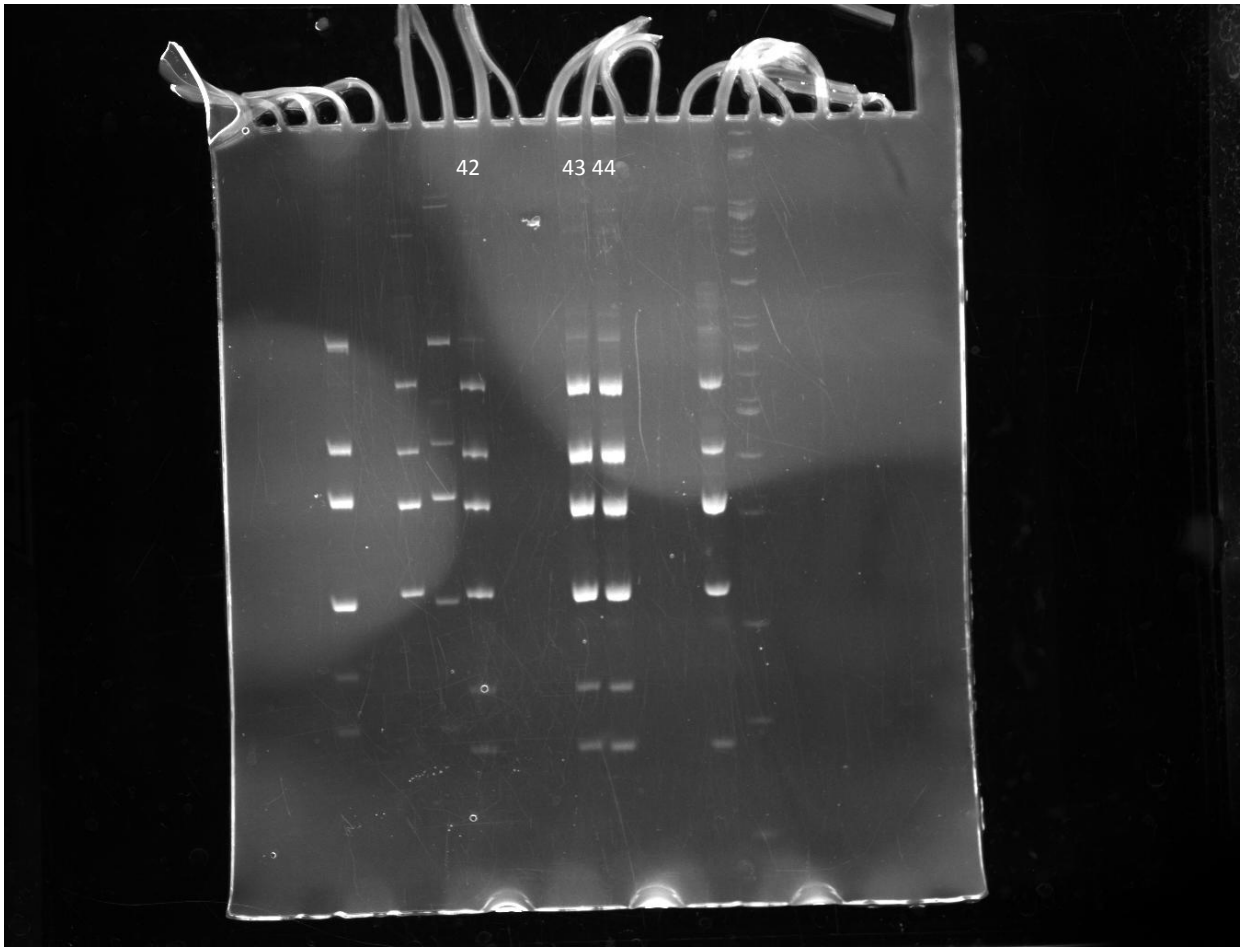

Figure S 2: Banding patterns of strains 42, 43, and 44.

Figure S.3

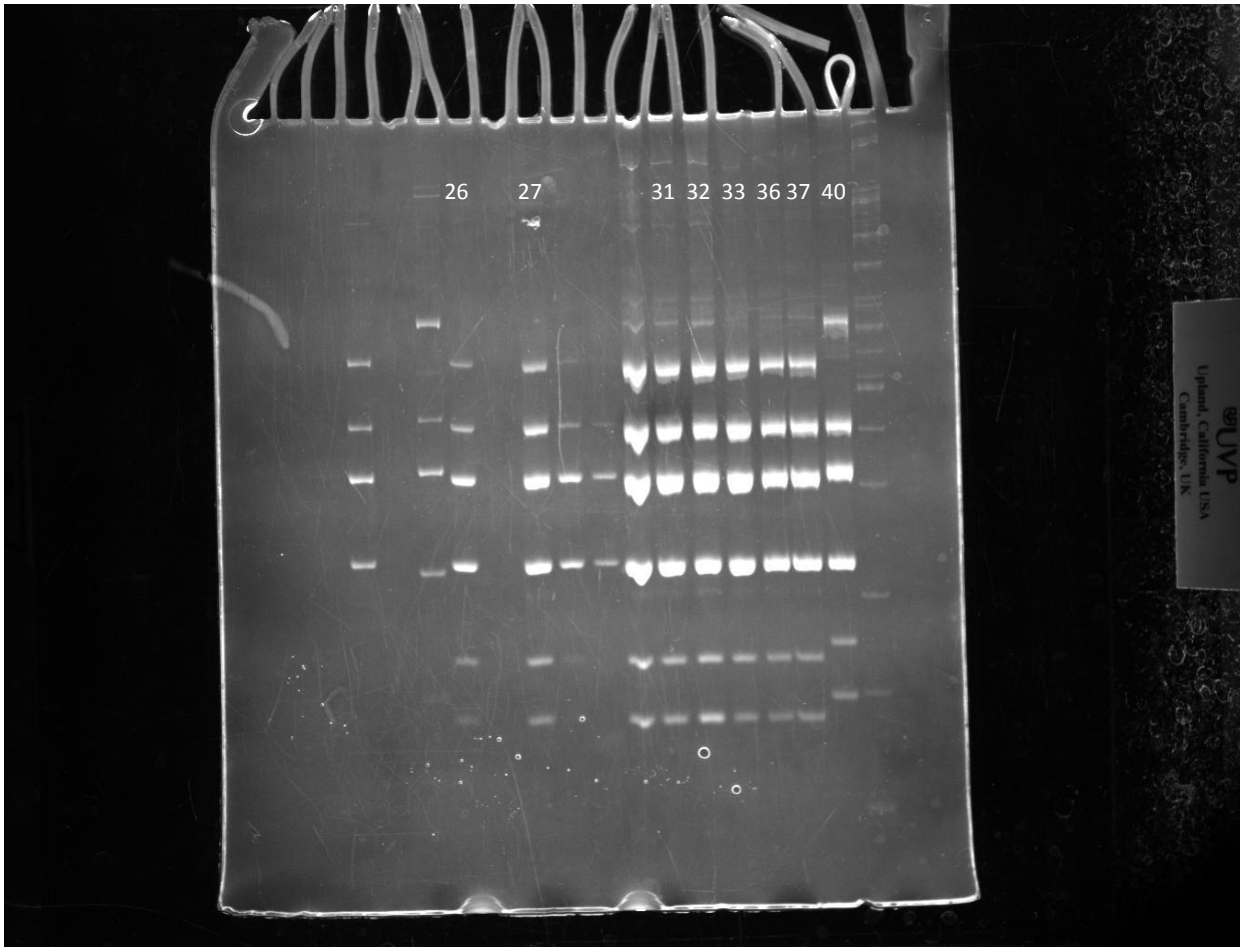

Figure S 3: Banding patterns of strains 26, 27, 31, 32, 33, 36, 37, and 40.

Figure S.4

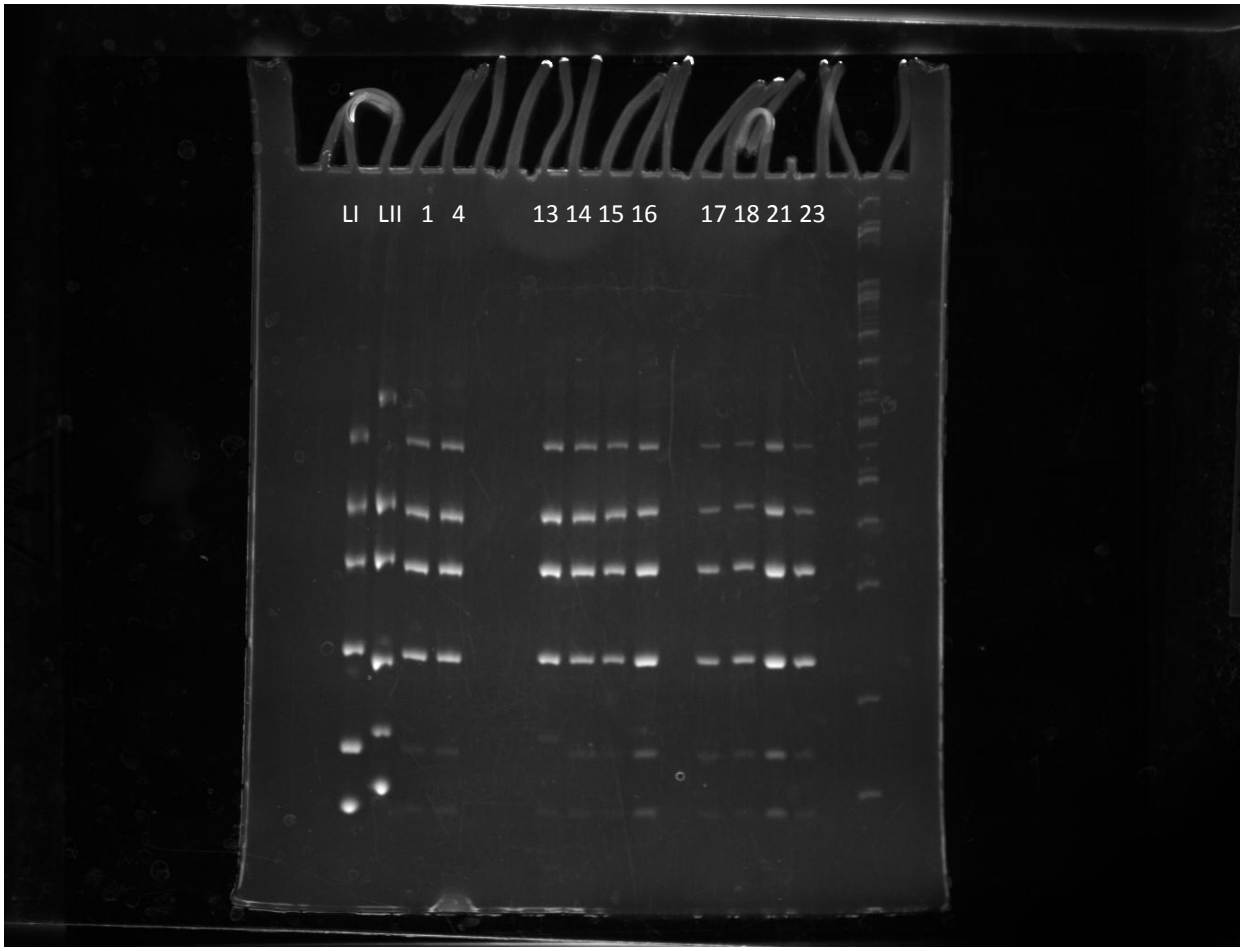

Figure S 4: Banding patterns of strains 1, 4, 13, 14, 15, 16, 17, 18, 21, and 23.

Figure S.5

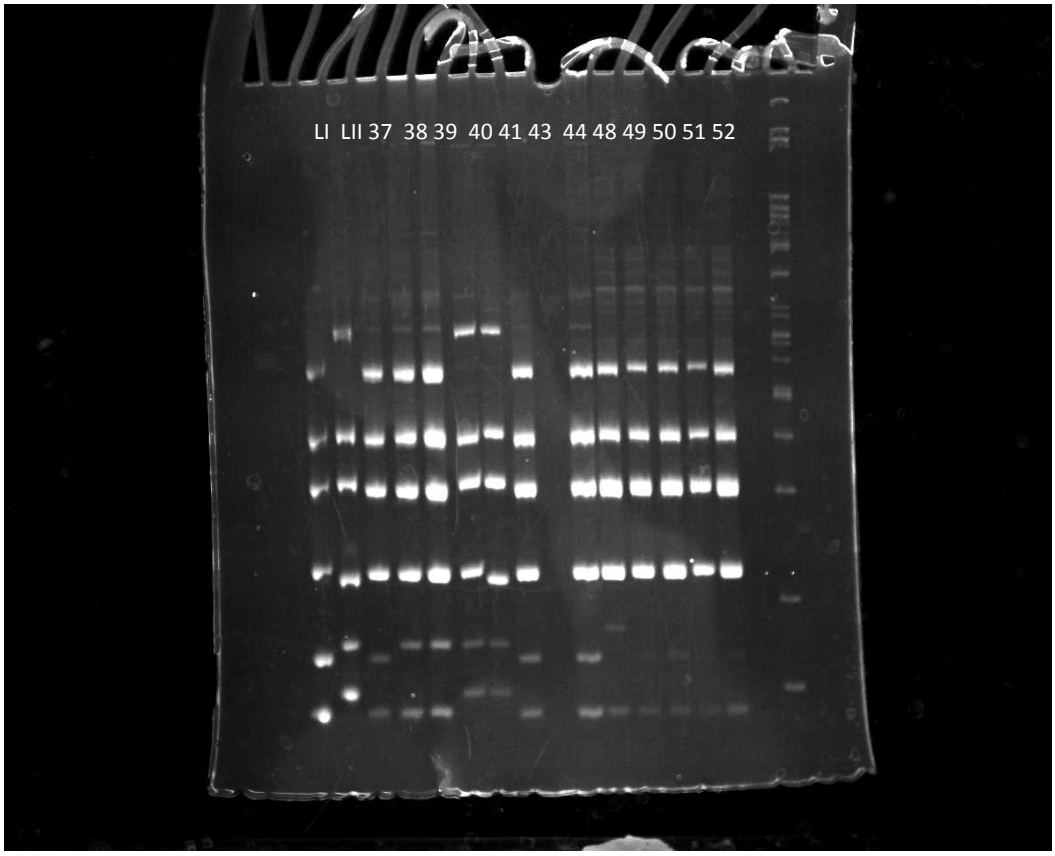

Figure S 5: Banding patterns of strains 37, 38, 39, 40, 41, 43, 44, 48, 49, 50, 51, and 52.

Figure S.6

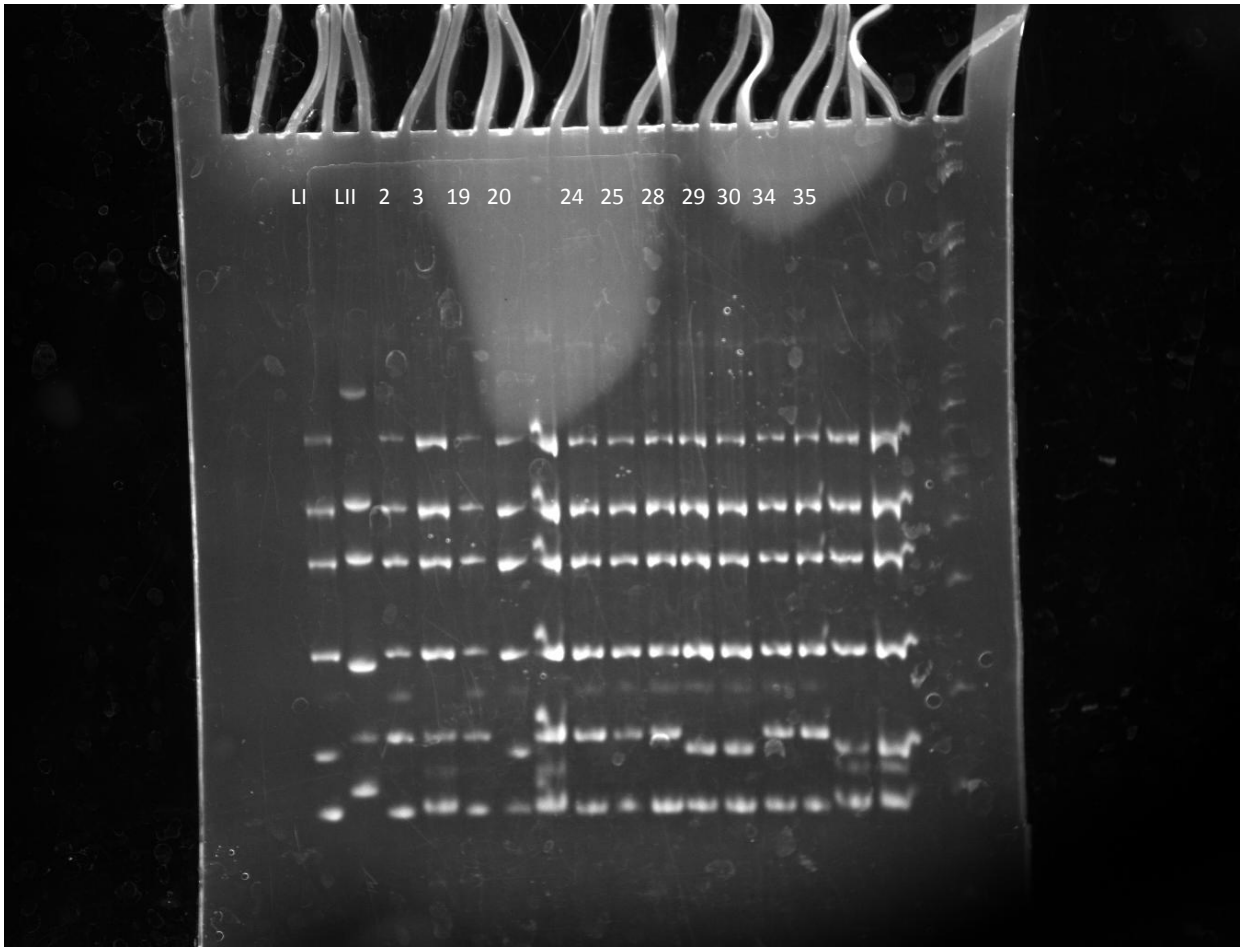

Figure S 6: Banding patterns of strains 2, 3, 19, 20, 24, 25, 28, 29, 30, 34, and 35.

Figure S.7

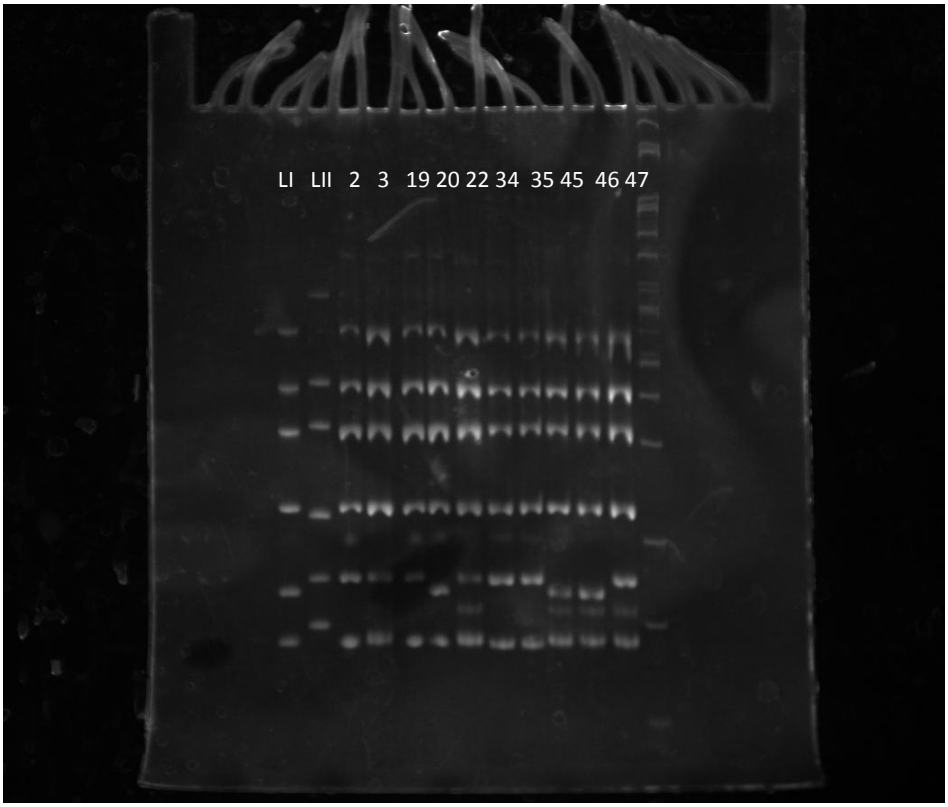

Figure S 7: Banding patterns of strains 2, 3, 19, 20, 22, 34, 35, 45, 46, and 47.

REFERENCES

Andreoli, S.P., H. Trachtman, D.W. Acheson, R.L. Siegler, and T.G. Obrig. 2002. Hemolytic Uremic Syndrome: Epidemiology, Pathophysiology, and Therapy. Pediatric Nephrology. 17:293-298.

Banatvala, N. P.M. Griffin, K.D. Greene, T.J. Barrett, W.F. Bibb, J.H. Green. 2001. The United States National Prospective Hemolytic Uremic Syndrome Study: Microbiologic, Serologic, Clinical, and Epidemiologic Findings. Journal of Infectious Disease 183: 1063-1070.

Boerlin, P., S.A. McEwen, I Boerlin-Petzold, J.B. Wilson, R.P. Johnson, C.L. Gyles. 1999. Associations Between Virulence Factors of Shiga toxin-producing *Esherichia coli* and Disease in Humans. Journal of Clinical Microbiology 37:497-503.

Byrne, C.M., I. Erol, J.E. Call, C.W. Kaspar, D.R. Buege, C.J. Hiemke, P.J. Fedorka-Cray, A.K. Benson, F.M. Wallace, and J.B. Luchansky. 2003. Characterization of *Escherichia coli* O157:H7 from Downer and Healthy Dairy Cattle in the Upper Midwest Region of the United States. Applied and Environmental Microbiology. 4683-4668.

Centers for Disease Control and Prevention. 2010. *Escherichia coli* O157:H7. Retrieved from http://www.cdc.gov/nczved/divisions/dfbmd/diseases/ecoli_o157h7/.

Delahay, R.M., G. Frankel, Knutton, S. 2001. Intimate Interactions of Enteropathogenic *Escherichia coli* at the Host Cell Surface. Current Opinions in Infectious Disease 14:559-565.

Faith, N.G. J.A. Shere, R. Brosch, K.W. Arnold, S.E. Ansay, M.S. Lee, J.B. Luchansky and C.W. Kaspar. 1996. Prevalence and Clonal Nature of *Escherichia coli* O157:H7 on Dairy Farms in Wisconsin. Applied and Environmental Microbiology. 1519-1525.

Franke, S. D. Harmsen, A. Caprioli, D. Pierard, L.H. Wieler, H. Karch. 1995. Clonal Relatedness of Shiga-like Toxin-producing *Escherichia coli* O101 of Human and Porcine Origin. Journal of Clinical Microbiology 33: 3174-3178.

Hayashi, T., T. Ooka, Y. Ogura, Asadulghani. 2008. Genomic View on the Evolution of Enterohemorrhagic *Escherichia coli*. Evolutionary Biology of Bacterial and Fungal Pathogens. Washington DC: ASM Press.

Ibekwe, A.M., P.J. Shouse, C.M. Grieve. 2006. Quantification of Survival of *Escherichia coli* O157:H7 on Plants Affected by Contaminated Irrigation Water. Engineering in Life Sciences 6(6): 566-572.

Jay, M.T., M. Cooley, D. Carychao, G.W. Wiscomb, R.A. Sweitzer, L. Crawford-Miksza, J.A. Farrar, D.K. Lau, J. O'Connell, A. Millington, R.V. Asmundson, E.R. Atwill, and R.E. Mandrell. 2007. *Escherichia coli* O157:H7 in Feral Swine near Spinach Fields and Cattle, Central California Coast. Emerging Infectious Diseases 13(12).

Kaper, J.B., J.P. Nataro, H.L.T. Mobley. 2004. Pathogenic *Escherichia coli.* Microbiology 2:123-138.

Kim, J. , J. Nietfeldt, and A. Benson. 1999. Octamer-based Genome Scanning Distinguishes a Unique Subpopulation of *Escherichia coli* O157:H7 Strains in Cattle. Proceedings of the National Academy of Sciences 96(33): 13288-13293.

Liu, K., S.J. Knabel, and E.G. Dudley. 2009. *rhs* Genes Are Potential Markers for Multilocus Sequence Typing of *Escherichia coli* O157:H7 Strains. Applied and Environmental Microbiology 75(18): 5853-5862.

Manning, S.D., A.S Motiwala, A.C. Springman, W. Qi, D.W. Lacher, L.M. Oullette, J.M. Mladonicky, P. Somsel, J.T. Rudrik, S.E. Dietrich, W. Zhang, B. Swaminathan, D. Aland, and T.S. Whittham. 2008. Variation in Virulence Among Clades of *Escherichia coli* O157:H7 Associated with Disease Outbreaks. Proceedings of the National Academy of Sciences 105(12): 4868-4873.

McDaniel, T.K., K.G. Jarvis, M.S. Donnenberg, and J.B. Kaper. 1995. A Genetic Locus of Enterocyte Effacement Conserved Among Diverse Enterobacterial Pathogens. Proceedings of the National Academy of Sciences USA. 92:1664-1668.

Melton-Celsa, A.R., A. O'Brian. 1998. Structure, Biology, and Relative Toxicity of Shiga Toxin Family Members for Cells and Animals, p. 121-1248. *In* J.B. Kaper and A.D. O'Brian (eds.) *Escherichia coli O157:H7 and Other Shiga Toxin-Producing E. coli Strains.* Washington DC: ASM Press.

Nataro, J.P., J.B. Kaper. 1998. Diarrheagenic *Escherichia coli.* Clinical Microbiology Reviews 11:142-201.

Rangel, J.M., P.H. Sparling, C. Crowe, P.M. Griffin and D.L. Swerdlow. Epidemiology of *Escherichia coli* O157:H7 Outbreaks, United States, 1982-2002. Emerging Infectious Diseases 11(4): 603-609.

Riley, L.W., R.S. Remis, S.D. Helgerson, H.B. McGee, J.G. Wells, B.R. Davis, R.J. Hebert, E.S. Olcott, L.M. Johnson, N.T. Hargrett, P.A. Blake, and M.L. Cohen. 1983. Hemorrhagic Colitis Associated with a Rare *Escherichia coli* Serotype. New England Journal of Medicine. 308:681-685.

Sharma, R., K. Stanford, M. Louie, K. Munns, S.J. John, Y. Zhang, V. Gannon, L. Chui, R. Read, E. Topp, and T. McAllister. 2008. *Escherichia coli* O157:H7 Lineages in Healthy Beef and Dairy Cattle and Clinical Human Cases in Alberta, Canada. Journal of Food Protection. 72(3): 601-607.

Xu, J.G, B.K. Cheng, H.Q. Jing. 1999. *Escherichia coli* O157:H7 and Shiga-like Toxin Producing *Escherichia coli* in China. World Journal of Gastroenterology. 5(3):191-194.

Yang, Z., J. Kovar, J. Kim, J. Nietfeldt, D. Smith, R. Moxley, M. Olson, P. Fey, and A Benson. 2004. Identification of Common Subpopulations of Non-Sorbitol-Fermenting B-Glucuronidase-Negative *Escherichia coli* O157:H7 from Bovine Production Environments and Human Clinical Samples. Applied and Environmental Microbiology 70(11): 6846-6854.

Zhang, Y., C. Laing, M. Steele, K. Ziebell, R. Johnson, A. Benson, E. Taboada, and V. Gannon. 2007. Genome Evolution in Major *Escherichia coli* O157:H7 Lineages. BMC Genomics 8:121.

VITA

Annette Hartzell

Education

Bachelor of Science, 05/2010
Penn State University Department of Food Science, Schreyer Honors College

- Major: Food Science
- Minor: Biochemistry and Molecular Biology
- Research advisor: Edward Dudley
- Honors Thesis: Genomic Classification of Clinical *Escherichia coli* O157:H7 Strains in the State of Pennsylvania

Professional and Research Experience

Undergraduate Researcher in Food Micro / Molecular Biology, 09/2008 to present
Penn State University

- Determined evolutionary subtypes for *E. coli* O157:H7 strains isolated from patients in Pennsylvania
- Presented research to Penn State Food Micro Group, Penn State Undergraduate Research Exhibition, and College of Agricultural Sciences Research Exhibition
- Experience with aseptic technique, PCR, gel electrophoresis

Science U Camp Curriculum Mentor, 07/2009 to 08/2009
Penn State University

- Facilitated laboratory and recreational activities for middle school-aged campers in Food Microbiology, CSI, and Myth Busters Camps

Research, Quality and Innovation Intern, 05/2008 to 08/2008
ConAgra Foods

- Developed frozen dinner concepts and formulations for the Banquet product line
- Assembled products for consumer testing in the pilot plant
- Collected data for a heating instructions study
- Collaborated on interdepartmental cuttings and brainstorming sessions

Undergraduate Researcher in Confectionary Research Lab, 01/2007 to 01/2008
Penn State University

- Researched the effects of temperature, fat blend ratios and the presence of emulsifiers in preventing fat bloom in confectionary products containing solid fats and liquid oils
- Presented research at annual meeting of Center for Food Manufacturing, Penn State Undergraduate Research Exhibition, and College of Agricultural Sciences Research Exhibition
- Gained experience in NMR analysis, experimental design, materials acquisition

Action Potential Science Experience Curriculum Mentor, 06/2007 to 08/2007
Penn State University

- Coordinated hands-on scientific activities for children in food, forensic and biomedical sciences
- Trained and oversaw junior curriculum mentors and team-taught with other curriculum mentors

Activities and Memberships

Penn State Food Science Product Development Team- Vice President and HACCP team leader, 2008 to present
Penn State University

- IFTSA 2009 Product Development Competition Finalist with *Petit Cadeau*, a unique confection consisting of a wine-based coating and sweet, Brie cheese filling
- Developed HACCP plan, bench top and scale-up processing and production plan
- Collaborated on development of product concepts, formulations, processing methods, safety standards, and marketing plans

Penn State Food Science Club- Secretary, April 2009 to present
Penn State University

- Record and distribute meeting minutes, handle communications with donors, volunteers and company representatives
- Collaborate on organization of annual tailgate, fundraising activities, community outreach

Penn State Food Science College Bowl Team, 10/2009 to present
Penn State University

- Compete against students from other Food Science programs in Food Science trivia

Institute of Food Technologists Student Association, Microbiology Division, Keystone Section, 2006 to present
Penn State University

Scholarships and Awards

- Schreyer Honors College scholar
- Gamma Sigma Delta Agricultural Honor Society
- Phi Tau Sigma Food Science Honor Society
- Speizer Undergraduate Research Award
- Institute of Food Technologists Scholarships
- IFT Food Microbiology Division Scholarship
- Penn State College of Agricultural Sciences Scholarships