Title of Thesis: WHOLE GENOME SEQUENCING ANALYSIS ON SHIGA TOXIN- PRODUCING ESCHERICHIA COLI O157:H7 FROM CATTLE FED WITH DIFFERENT DIETARY PROTEIN CONCENTRATIONS

Xun Yang, Master of Science, 2017

Thesis Directed By: Professor, Jianghong Meng, NFSC

Escherichia coli serotype O157:H7 was first recognized in 1982 as a human pathogen associated with outbreaks of bloody diarrhea in the United States and is now considered a major cause of foodborne infections because of its high hospitalization rate. Cattle is the major reservoir of *E. coli* O157:H7. Cattle harbor *E. coli* O157:H7 in the hindgut and shed the organisms in the feces, which serves as a source of contamination of food and water. It is hypothesized that dietary ingredients that reach the hindgut are likely to affect colonization and fecal shedding of STEC. Increased flow of dietary ingredients (starch, fiber, protein, and lipid) are likely to alter ecology of the hindgut, resulting in altered pH and fermentation products, which could have a positive or negative impact on *E. coli* O157:H7. The objectives of this study are to investigate *E. coli* O157:H7 populations in fecal shedding of cattle. The cattle in this study were fed with diets with
different levels of ruminally-degradable and –undegradable protein. A total of 286 E. coli O157:H7 isolates were recovered from feces of 576 crossbred calves at the Clayton Livestock Research Center in Clayton, New Mexico. The organisms were sequenced using Illumina Miseq system. De novo assembly of raw reads was performed using SPAdes and SNPs analysis of the isolates was conducted using kSNP3. Virulence factor Database (VFDB), created by the MOH key laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology, was used as a reference for BLAST. The results indicated that increased flow of undegradable protein may increase the shedding of E. coli O157. However, the effect of ractopamine was still unknown. Three clades were identified among the E. coli O157:H7 isolates, including clades 6, 7, and 8, most of which belong to clade 8 (205 of 286). 49588 SNPs were found according to kSNP3. 19043 SNPs were identified as core SNPs. The phylogenetic analysis showed that the E. coli O157 isolates which collected from neighboring Pens were more closely to each other.
WHOLE GENOME SEQUENCING ANALYSIS ON SHIGA TOXIN-PRODUCING *E. COLI* O157:H7 SHEDDING IN FINISHING CATTLE WITH DIFFERENT DIETARY PROTEIN CONCENTRATIONS

by

Xun Yang

Thesis submitted to the Faculty of the Graduate School of the University of Maryland, College Park, in partial fulfillment of the requirements for the degree of Master of Science 2017

Advisory Committee:
Professor Jianghong Meng, Chair
Associate Professor Seong-Ho Lee
Research Microbiologist Ruth Timme
Acknowledgement

I would first like to thank my thesis advisor Dr. Jianghong Meng of the Department of Nutrition and Food Science at University of Maryland for providing me this extraordinary opportunity to study and enlightening me the first glance of research in food safety. This accomplishment would not have been possible without his guidance. I also want to express my gratitude to Dr. Seong-He Lee, and Dr. Ruth Timme of their time and support through serving on my committee.

I would also like to thank Dr. T.G. Nagaraja of the Department of Diagnostic Medicine/Pathobiology at Kansas State University for making this collaborative project happen in the first place.

I would also like to acknowledge my previous and current fellow labmates: Dr. Jay N. Worley, Dr. Xin Gao, Dai Kuang, Colette Le Bienvenu, Dr. Guojie Cao, Dr. Qiongqiong Yan, for all the help and suggestions about my study and research during the past two years. In particular, I am grateful to Dr. Jay N. Worley and Dr. Xin Gao for giving me advice and guidance on data analysis.

Finally, I must express my very profound gratitude to my parents for providing me with both physical and spiritual support and continuous encouragement throughout my years of study at University of Maryland.
# Table of Contents

Acknowledgement ........................................................................................................................................ ii

Table of Contents ........................................................................................................................................ iii

List of Tables .................................................................................................................................................... v

List of Figures .................................................................................................................................................... vi

CHAPTER 1 Introduction ................................................................................................................................. 1

1. Overview of Shiga toxin-producing *Escherichia coli* (STEC) ................................................................. 1

   1.1 Pathogenicity and virulence factors of STEC .................................................................................... 2

   1.2 Different serotypes of STEC and *Escherichia coli* O157:H7 (*E. coli* O157:H7) ....................... 4

   1.3 Reservoirs of STEC ......................................................................................................................... 7

   1.4 Dietary effects on *E. coli* O157:H7 shedding in cattle .................................................................. 8

2. Next-generation sequencing technology ................................................................................................. 10

3. Clade identification of *E. coli* O157:H7 .............................................................................................. 12

   Study Objectives: .................................................................................................................................... 12

CHAPTER 2 Materials and methods ........................................................................................................... 13

1. Animals and Facilities ............................................................................................................................... 13

2. Experimental Design and Treatments ..................................................................................................... 13

3. Fecal sample collection and processing .................................................................................................. 15

4. Culture Method of Detection of *E. coli* O157:H7 ............................................................................. 16

5. PCR Method of Detection of non-O157 Serogroups ........................................................................... 17

6. DNA extraction and whole genome sequencing ..................................................................................... 17

7. Data analysis ............................................................................................................................................ 17

   7.1 Genomic analysis ............................................................................................................................... 17

   7.2 Clade identification ............................................................................................................................ 18

   7.3 Statistical analysis .............................................................................................................................. 18

CHAPTER 3 Results ......................................................................................................................................... 19

1. Clade information though Pens .............................................................................................................. 19

2. Clade profiles with collection dates ........................................................................................................ 22

3. Virulence genes ....................................................................................................................................... 27

4. Phylogenetic relationships ..................................................................................................................... 29

CHAPTER 4 Discussion .................................................................................................................................. 34
1. Clade lineages of *E. coli* isolates.................................................................34
2. Virulence genes ............................................................................................34
3. Re-sequence ..................................................................................................35

CHAPTER 5 Conclusion ......................................................................................38

References ........................................................................................................39
List of Tables

Table 1.1 selected outbreaks of *E. coli* O157:H7 from 2006 to 2017 ................................................. 6
Table 2.1 Components of each diet in this study .................................................................................. 15
Table 2.2 Treatments given date for each block ................................................................................ 15
Table 3.1 Clade information for each diet ............................................................................................ 19
Table 3.2 stx gene profiles .................................................................................................................... 27
Table 4.1 Shiga toxin genes profile of isolates in this study and outbreak strains ......................... 35
List of Figures

Figure 2.1 Pen assignments and tag colors ......................................................... 14
Figure 3.1 Clade information for each diet .......................................................... 20
Figure 3.2 Clade information of each pens with DIP diet .................................... 21
Figure 3.3 Clade information of each pens with UIP diet .................................... 21
Figure 3.4 Clade information of each pens with CON diet ................................... 22
Figure 3.5 Prevalence of *E. coli* O157:H7 of each diet (n = 96) ....................... 23
Figure 3.6 Prevalence of *E. coli* O157:H7 with and without ractopamine treatment (n=48) ................................................................. 23
Figure 3.7 Prevalence of *E. coli* O157:H7 on each collection day of UIP group with clade information (n=96) ...................................................... 24
Figure 3.8 Relative prevalence of *E. coli* O157:H7 with UIP diet between rh treatment and without rh treatment (n=48) .......................... 25
Figure 3.9 Prevalence of *E. coli* O157:H7 on each collection day of CON group with clade information (n=96) ...................................................... 25
Figure 3.10 Relative prevalence of *E. coli* O157:H7 with CON diet between rh treatment and without rh treatment (n=48) .......................... 25
Figure 3.11 Prevalence of *E. coli* O157:H7 on each collection day of DIP group with clade information (n=96) ...................................................... 26
Figure 3.12 Relative prevalence of *E. coli* O157:H7 with DIP diet between rh treatment and without rh treatment (n=48) .......................... 27
Figure 3.13 Selected virulence factors results ..................................................... 28
Figure 3.14 The phylogenetic tree based on SNP differences ............................ 30
Figure 3.15 SNP-based, collapsed phylogenetic tree ........................................ 31
CHAPTER 1 Introduction

1. Overview of Shiga toxin-producing *Escherichia coli* (STEC)

*Escherichia coli* (also known as *E. coli*) is a gram-negative, facultatively anaerobic, rod-shaped, coliform bacterium of the genus *Escherichia* that is commonly found in the environment, foods, and intestines of people and animals. Most *E. coli* are harmless, in fact, are important component of human intestinal tract. However, some *E. coli* can cause severe illness. They can be transmitted through contaminated food or water, contact with animals, or person-to-person. *E. coli* that cause diarrheal disease are divided into six pathotypes based on their virulence factors: enteropathogenic *E. coli* (EPEC); enterotoxigenic *E. coli* (ETEC); enteroaggregative *E. coli* (EAEC); enteroinvasive *E. coli* (EIEC); enterohemorrhagic *E. coli* (EHEC) (2). The term EHEC is often used interchangeably with “STEC” and is intended to denote the subset of STEC capable of causing bloody diarrhea or HUS (3). However, because different definitions of EHEC have been used in the scientific literature, to minimize confusion, “STEC” has been used by the Center for Disease Control and Prevention (CDC) (3).

The *E. coli* that can make Shiga toxin are called Shiga toxin-producing *Escherichia coli* (STEC). STEC are estimated to cause more than 265,000 illness each year in the United States, with more than 3600 hospitalization and 30 deaths (4). STEC cause a potentially fatal zoonotic foodborne illness whose clinical spectrum includes nonspecific diarrhea, hemorrhagic colitis, and the hemolytic
uremic syndrome (HUS) (2, 5-7). Up to 40% of patients with HUS develop long-term renal dysfunction, and about 3-5% of death patients during the acute phase of the disease (8-11).

1.1 Pathogenicity and virulence factors of STEC

The ability to produce one or more of the Shiga toxin \( (stx) \) family cytotoxins constitutes the main virulence attribute of STEC (12). The presence of Shiga toxin 1 or 2 gene \( (stx1 \text{ or } stx2) \) is responsible for the pathogenesis of STEC. However, there several other factors are also involved. The attaching and effacing (A/E) lesion formation is caused by the type III secretion system (T3SS) encoded by the locus of enterocyte effacement (LEE), which injects small effector proteins into the host cell (13, 14), and thus are grouped as LEE-positive STEC. STEC pathogenesis is a multistep process, and besides the production of stx toxins and the A/E lesion, other factors including different types of toxins and adhesions have been described and found to be involved in virulence (15).

1.1.1 Shiga toxin

Shiga toxin is the main characteristic that defines STEC and is the key virulence factor in STEC causing HUS (16). Several in-depth reviews have been published regarding the mechanism of action in renal cells (positive for cell surface globtriaosylceramide \( \text{Gb3}^+ \)) (17), the intestinal epithelium \( \text{Gb3}^- \) (18), and the vascular endothelium \( \text{Gb3}^+ \) (19) and its contribution to the pathophysiology of HUS (20).

Shiga toxins can be divided into two types, Stx1 and Stx2, with Stx1 having 3 subtypes \( (a, c \text{ and } d) \), and Stx2 having 7 subtypes\( (a \text{ to } g) \) (21). STEC can carry a
single Shiga toxins variant, either stx1 or stx2, both stx1 and stx2, or a combination of stx2 subtypes (16). Both stx1- and stx2-containing STEC can lead to HUS; however, stx2 is more often associated with severe disease (22).

Both Stx1 and Stx2 are encoded on prophages that are integrated into the chromosome. Shiga toxin-carrying phages can become lytic during bacterial stress, and it is believed that Stx1/Stx2 is released from lysed bacterial cells during the lytic cycle of the phage (23). Lee et al. (24) has described the mechanism of Stx delivery and trafficking through endothelial cells. In summary, Stx binds to Gb3 on the surface of endothelial cells (25, 26) and is internalized and trafficked through the retrograde pathway from the Golgi apparatus and endoplasmic reticulum (ER) and eventually to the host cell cytoplasm. However, the mechanism of Stx transport from the intestinal lumen across the epithelium is unknown. Schuller et al. (18) hypothesized that STEC-induced inflammation may provide the toxin an opportunity to breach the epithelial barrier, and there are other possible mechanisms. STEC is able to cross the intestinal epithelium through microfold cells and survive in macrophages, and this may be a way the stx to be released into the blood stream, where it can target other organs (27).

1.1.2 Other virulence factors

The ability to adhere to intestinal epithelial cells is another important factor in STEC pathogenesis. STEC is believed to colonize the mucosa of the large bowel, with a characteristic attaching and effacing (A/E) cytopathology that is mediated by components of an approximate 43.4-kb pathogenicity island (PAI) referred as the locus of enterocyte effacement (LEE) (2). The central region of LEE comprises the (28). Downstream of eae are the esp genes, which encode the secreted proteins EspA, EspB, EspD, and EspF,
while upstream are many genes (*esc* and *sep*) that are thought to encode the type III secretion machinery that facilitates the transport of effector molecules to their active sites (28). Although LEE has been described in the major EHEC/STEC serotypes responsible for a high proportion of HC and HUS cases in several countries, its presence is not a required condition for the occurrence of more serious infections as initially thought, because some LEE-negative strains are also capable of causing outbreaks and sporadic cases of HUS (29, 30).

Another important virulence factor of STEC is enterohemorrhagic *E. coli* (EHEC) hemolysin. In STEC, *ehx* was found on pO157 and pO113 plasmids. The EHEC hemolysin (EHEC-*hly*A or *ehx*) is a pore-forming toxin that lyses sheep erythrocytes (31, 32). Several studies have shown some insight on the contribution of Ehx to disease though the role of it in virulence has been unclear. These toxins are associated as cargo with outer membrane vesicles (OMVs), prolonging its activity (33). Ehx was found to be cytotoxic to endothelial cells and may contribute to the development of HUS (34). In addition, Ehx was shown to be inactivated by another STEC virulence factor, EspP (35).

1.2 Different serotypes of STEC and *Escherichia coli* O157:H7 (*E. coli* O157:H7)

Despite over 400 STEC serotypes identified, only a subset of them have been correlated to illness in humans (36). The STEC can be divided into two groups: *E. coli* O157:H7 and non-O157.

It was first recognized in 1982 as a human pathogen associated with outbreaks of bloody diarrhea in the United States and is now considered a major cause of foodborne infections. The 1993 Jack in the Box outbreak of O157:H7 which resulted in 700 people infected, 171 hospitalized, and 4 dead prompted major changes in the US food safety system. In the ten years that followed, there were approximately thirty outbreaks
recorded in the United States (37). This number is likely misleading, however, because *E. coli* O157:H7 infections did not become a reportable disease in any state until 1987, when Washington became the first state to mandate its reporting to public health authorities (38, 39). Consequently, an outbreak would not be detected if it was not large enough to prompt investigation (38, 40). The Centers for Disease Control and Prevention (CDC) has estimated that *E. coli* O157:H7 infections cause 73,000 illness, 2200 hospitalizations, and 60 deaths annually in the United States (41). Although the outbreak surveillance data from CDC reported that *E. coli* O157:H7 infections are decreasing after the peak in 1999, large outbreaks and sporadic cases continue to occur in United States according to CDC reported investigations(41). *E. coli* O157:H7 infections are less common. However, the high hospitalization rate (46%, approximately 2138 cases) (4) and multistate infections make this *E. coli* serotype an important foodborne pathogen in the United States. Furthermore, table 1.1 shows multistate foodborne outbreak investigations involving *E. coli* O157:H7 from 2006 to 2017 according to CDC reports of selected *E. coli* O157:H7 outbreaks investigations. The annual cost of illness due to *E. coli* O157:H7 infections was 405 million dollars, including lost productivity, medical care, and premature deaths (42).

The most common serogroups of non-O157 STEC reported to cause foodborne illness in the United States are O26, O111, O103, O121, O45, and O145, which refers to the “Big 6” described in a study of non-O157 STEC human illness in the United States between 1983 and 2002 (43). In parts of Latin America, non-O157 STEC are more commonly implicated in HUS than is serotype O157:H7, and account for up to
40% of cases of HUS in Europe and for up to 25% of cases in the United States (28).

In a study of diarrheal outbreaks conducted in four Brazilian states, some uncommon serogroups including O1, O24 and O77 among others were detected, but they were all associated with acute diarrhea. It is interesting to note that the majority of patients from whom STEC was isolated were female (57%), and that patients’ ages ranged from 8 months to 80 years, with most being less than five years old (54%) (44). Likewise, these serotypes of STEC are common in many parts of the world, but the overall prevalence differs geographically (45).

Table 1.1 selected outbreaks of *E. coli* O157:H7 from 2006 to 2017

<table>
<thead>
<tr>
<th>Outbreak</th>
<th>Year</th>
<th>Case</th>
<th>Hospitalizations</th>
<th>Deaths</th>
<th>States</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.M. Healthy SoyNut Butter</td>
<td>2017</td>
<td>29</td>
<td>12</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>Beef Products</td>
<td>2016</td>
<td>11</td>
<td>7</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Rotisserie Chicken Salad</td>
<td>2015</td>
<td>19</td>
<td>5</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Ground Beef</td>
<td>2014</td>
<td>12</td>
<td>7</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Ready-to-Eat Salads</td>
<td>2013</td>
<td>33</td>
<td>7</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Spinach and Spring Mix</td>
<td>2012</td>
<td>33</td>
<td>13</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Romaine Lettuce</td>
<td>2011</td>
<td>49</td>
<td>33</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>Lebanon Bologna</td>
<td>2011</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Hazelnuts</td>
<td>2011</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Cheese</td>
<td>2010</td>
<td>38</td>
<td>15</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Beef</td>
<td>2010</td>
<td>21</td>
<td>9</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Beef</td>
<td>2009</td>
<td>26</td>
<td>19</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Beef</td>
<td>2009</td>
<td>23</td>
<td>12</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>
1.3 Reservoirs of STEC

Microbes live throughout the gastrointestinal tract of mammals, including ubiquitous and adaptable *E. coli*. Ruminants such as cattle, both meat and dairy, are widely known to be major reservoirs for pathogenic STEC, and exposure to their fecal matter represents an important source of human illness (46). A survey of ground beef in the United States showed that approximately 24% of samples tested were positive for *stx* genes by PCR, with only a small proportion of isolates being potentially pathogenic to humans (47). STEC O157:H7 has been isolated from other animals and insects, which include but are not limited to swine, sheep, deer, wild boars, rabbits, birds, dogs, rodents, and insects (48). Certain STEC serotypes, such as O157:NM, and O104:H4, have rarely been or have not been reported as being isolated from animals (49, 50). During the 2011 European outbreak of 4,321 cases, fenugreek sprouts were identified as the most likely source of infection (51). The seeds were imported as a lot in late 2009 from Egypt, and it is still unknown if the point of contamination occurred at the site where seeds were produced, during transportation, or at the importer (51). Furthermore,
it was determined that cattle were not a reservoir for the 2011 European outbreak of STEC O104:H4 (52). Finally, human-to-human transmission did occur during the STEC O104:H4 infection in Poland, France, Germany, and the Netherlands during the outbreak (53-57).

As the most important foodborne pathogen affecting human health worldwide, *E. coli* O157:H7 is most commonly found in cows, although chickens, deer, sheep, and pigs have also been known to carry it. A 2003 study on the prevalence of *E. coli* O157:H7 in livestock at 29 county and three large agricultural fairs in the United States found that *E. coli* O157:H7 was isolated from 13.8% of beef cattle, 5.9% of sheep, and 2.8% of goats (58). Over 7% of pest fly pools also tested positive for *E. coli* O157:H7 (58). Among these animals, cattle is the major reservoir of STEC. Cattle harbor STEC in the hindgut and shed the organisms in the feces, which serves as a source of contamination of food and water.

1.4 Dietary effects on *E. coli* O157:H7 shedding in cattle

Ground beef is the most frequently implicated source of *E. coli* O157:H7 outbreaks, and bovine-derived products are linked to approximately 75% of *E. coli* O157:H7 outbreaks (59, 60). Cattle are a major reservoir of *E. coli* O157:H7 and repeated hemorrhagic colitis outbreaks linked to consumption of ground beef, animal contact, manure amendment, or cattle manure contaminated runoff has firmly established the connection between cattle and *E. coli* O157:H7 epidemiologically, and in public perception (61-63).

Fecal shedding of STEC is influenced by several factors and diet is one of the major factors. It is hypothesized that dietary ingredients that reach the hindgut are
likely to affect colonization and fecal shedding of STEC. Increased flow of dietary ingredients (starch, fiber, protein, and lipid) are likely to alter ecology of the hindgut, resulting in altered pH and fermentation products, which could have a positive or negative impact on STEC. Increased flow of protein into the hindgut resulting in increased ammonia concentration could increase the gut pH and possibly favor colonization of STEC. Therefore, it would be of interest to determine fecal shedding of STEC in cattle fed diets supplemented with different levels of ruminally-degradable or –undegradable protein.

Ractopamine hydrochloride is a beta-adrenergic agonist that is used as a feed additive to improve growth performance and carcass quality (64). Beta-adrenergic agonists are physiologically and pharmacologically similar to natural catecholamines, such as epinephrine and norepinephrine (65). Several studies have reported on the effects of natural catecholamines on growth of bacteria. Norepinephrine and dopamine induced growth of *E. coli* O157:H7 and *Salmonella enterica* has been reported (66). Catecholamines have also been shown to induce expressions of virulence genes, such as Shiga toxins (67). Increased bacterial growth has been attributed to the increased iron availability to the bacteria due to interaction of catecholamines with host-iron binding protein (68). Also, catecholamine might induce stimulation of bacterial growth and virulence gene expression of *E. coli* O157:H7 via quorum sensing mechanisms (69-71). In view of the effects of natural catecholamines on bacteria, several studies have been conducted to assess the effects of their synthetic counterparts, beta-adrenergic agonists on fecal shedding of *E. coli* O157:H7. Edrington et al. (2006b) showed increased cecal population and fecal shedding of *E. coli* O157:H7, following oral inoculation of the
pathogen in sheep administered with ractopamine hydrochloride. On the contrary, there was decreased fecal shedding of *Salmonella* in experimentally inoculated pigs administered with ractopamine hydrochloride (72). Edrington et al., 2006 (73) have also reported decreased fecal shedding of *E. coli* O157:H7, but increased shedding of *Salmonella* in cattle administered with ractopamine hydrochloride. Also, there are reports of no effects of ractopamine on fecal shedding of *E. coli* O157:H7 in cattle (74, 75). Walker and Drouillard (76) reported that ractopamine feeding decreases proteolysis due to its direct effect on ruminal bacteria that break down amino acids, which resulted in decreased ammonia concentration. Although natural catecholamines are known to increase bacterial growth and gene expressions, there are contradictory reports regarding the effects of ractopamine on fecal shedding of *E. coli* O157:H7. There has been no study reported on the effects of ractopamine on fecal shedding of non-O157 STEC. Therefore, further research is warranted to investigate the effect of ractopamine on fecal shedding of STEC.

2. Next-generation sequencing technology

Next-generation sequencing (NGS), also called massive parallel sequencing, was developed in the last decade and allows simultaneous sequencing of millions of DNA fragments without previous sequence knowledge (77). The Sanger DNA sequencing method, also named chain terminator sequencing, is a technique for DNA sequencing based on the selective incorporation of chain-terminating dideoxynucleotides (ddNTPs) by DNA polymerase during in vitro DNA replication, was developed by Frederick Sanger and colleagues in 1997. This method, which was later automated and underwent slight modifications, was the sequencing gold standard until late
2000’s (78).

Due to the high costs and intensive work required, traditional sequencing was only performed on specific DNA regions and for specific samples (77). Compared to the traditional Sanger capillary electrophoresis sequencing method (79, 80), which is considered a first-generation sequencing technology, NGS technologies provide higher throughput data with lower cost and enable population-scale genome research. NGS technologies have three major improvements compared to first-generation sequencing: first, NGS methods do not require a bacterial cloning procedure and prepare libraries for sequencing in a cell free system; second, NGS technologies process millions of sequencing reactions in parallel and at the same time; third, detection of bases is performed cyclically and in parallel (81).

Current NGS technology is sorted in two major types, i.e., short- and long-read sequencing (77). Short-read sequencing is mainly performed by Illumina® protocols and machines and is described as cheap “sequencing by synthesis” (SBS) of reads shorter than 300bp (82). The ion semiconductor method (Ion Torrent®) is another cheap short-read sequencer (83). The long-read sequencing is performed mainly by PacBio® or Roche®, is a costly “single molecule real-time” (SMRT) technology of reads longer than 2.5 Kb (84, 85). The Oxford Nanopore Technologies® MinION, using single stranded pore technology, actually allows to sequence very long molecules (>10 Kb) (86) and at a relative low cost but with a relatively higher error rate compared with other sequencers. The advances of NGS technology has led to decreasing cost and growing speed of Whole Genome Sequencing (WGS) (87). The WGS data can be used to identify the phylogeny, virulence factors, SNPs of
microorganisms efficiently to help us better understand the organism and detect the outbreak associated with foodborne pathogen.

3. Clade identification of *E. coli* O157:H7

Several recent outbreaks involving *E. coli* O157:H7 contamination of fresh produce were associated with more severe disease, as defined by higher hemolytic uremic syndrome and hospitalization frequencies, suggesting that increased virulence has evolved. To test this hypothesis, Manning et al. has developed a system that detects SNPs in 96 loci and applied it to more than 500 *E. coli* O157:H7 clinical strains. As a result, the analyses in their study identified 39 SNP genotypes that differ at 20% of SNP loci and separated *E. coli* O157:H7 into nine distinct clades. The study showed the SNP differences between strains associated with outbreaks and which lineage might have acquired critical factors that contribute to more severe disease. In addition, the ability to detect genotype O157 strains belonging to such lineages, identify more virulent O157 subtypes of this useful, rapid, and inexpensive molecular test proposed in this study will have significant impact on both disease diagnosis and treatment guidelines, and also, would be useful for clinical laboratories to identify patients with an increased likelihood of developing HUS.

Study Objectives:

The objectives of our study were to investigate a possible link of dietary protein levels to *E. coli* O157:H7 populations in fecal shedding of cattle by analyzing the WGS data of 311 isolates.
CHAPTER 2 Materials and methods

1. Animals and Facilities.

A total of 576 crossbred calves at the Clayton Livestock Research Center in Clayton, New Mexico were used for this study. Cattle were housed in soil-surfaced pens (40 ft. x 115 ft.) with 36 ft. of bunk space. Water was supplied to each pen with a metered water tank (Cattlemaster 480, Ritchie Inc.). Cattle were fed a diet to gain approximately 2 pounds daily until they reach approximately 850 pounds, at which time the dietary treatments were given to those cattle for this proposed study.

2. Experimental Design and Treatments.

The cattle were weighed individually, given a commercial growth implant, and ranked in a spreadsheet by body weight after they reached approximately 850 pounds. Then calves were separated into 4 blocks based on body weight. Within each block, calves were assigned consecutively to pens so that the average weight and standard deviation of a pen can be similar among pens. As a result the study became to a randomized complete block design consisting of 48 soil-surfaced pens (40 ft. x 115 ft. with 36 ft. bunk line) and 4 blocks (figure 2.1). Pen were served as the experimental unit. Within each block, pens of cattle were randomly assigned to 6 dietary treatments in a 2 x 3 factorial arrangement.
Dietary treatments were initiated once cattle were weighed using a pen scale and adapted to a finishing diet. These dietary treatments were consisted of steam-flaked corn-based finishing diets and shown in table 2.1. The diets were first administered 35 days before harvest. Cattle in blocks 3 & 4 were first administered the diet on August 23rd, 2015, and cattle in block 1 & 2 on August 30th, 2015. For pens receiving ractopamine, it was administered in the diet on
four days (table 2.2).

Table 2.1 Components of each diet in this study

<table>
<thead>
<tr>
<th></th>
<th>Crude protein</th>
<th>Degradable protein</th>
<th>Undegradable protein</th>
<th>ractopamine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Diet 1</td>
<td>14.5%</td>
<td>9.2%</td>
<td>5.3%</td>
<td>0</td>
</tr>
<tr>
<td>Control Diet 2</td>
<td>14.5%</td>
<td>9.2%</td>
<td>5.3%</td>
<td>400 mg</td>
</tr>
<tr>
<td>High Degradable diet 1</td>
<td>19.5%</td>
<td>13.2%</td>
<td>6.5%</td>
<td>0</td>
</tr>
<tr>
<td>High Degradable diet 2</td>
<td>19.5%</td>
<td>13.2%</td>
<td>6.5%</td>
<td>400 mg</td>
</tr>
<tr>
<td>High Undegradable diet 1</td>
<td>19.5%</td>
<td>9.2%</td>
<td>10.2%</td>
<td>0</td>
</tr>
<tr>
<td>High Undegradable diet 2</td>
<td>19.5%</td>
<td>9.2%</td>
<td>10.2%</td>
<td>400 mg</td>
</tr>
</tbody>
</table>

Table 2.2 Treatments given date for each block

<table>
<thead>
<tr>
<th>Block</th>
<th>Dietary treatment given</th>
<th>Rh given</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block 1</td>
<td>8/30/2015</td>
<td>9/9/2015</td>
</tr>
<tr>
<td>Block 2</td>
<td>8/30/2015</td>
<td>9/7/2015</td>
</tr>
<tr>
<td>Block 3</td>
<td>8/23/2015</td>
<td>9/4/2015</td>
</tr>
<tr>
<td>Block 4</td>
<td>8/23/2015</td>
<td>9/2/2015</td>
</tr>
</tbody>
</table>

3. Fecal sample collection and processing.

Freshly-defecated fecal pats on the pen floor were sampled with a plastic spoon (two to three spoonful). Care will be taken while sampling to avoid ground contamination. The spoon with feces were placed into a Whirl-pack bag (Nasco, Ft. Atkinson, WI), and placed in a cooler with ice packs. Six fecal samples were collected from each pen and labeled with pen numbers and sample numbers, thus blinding laboratory personal to treatment assignments. Samples were sent to the Kansas State
University College of Veterinary Medicine Pre-harvest Food Safety Laboratory for processing by FedExe. In the laboratory, approximately 1 g of feces were added to a tube of 9 mL *Escherichia coli* broth (EC; Difco, ThermoFisher). Tube were vortexed to obtain a good fecal suspension. One mL of the suspension were pipetted into a 1.5 mL tube for DNA extraction (pre-enrichment) for PCR assay. The fecal suspension were incubated at 40° C for 6 hours. One mL of the suspension were pipetted into a tube for DNA extraction (post-enrichment) for PCR assay.


An aliquot of the enriched fecal suspension were subjected to immune magnetic separation (IMS) procedure with Abraxis® (Warminster, PA) O157 IMS beads. Twenty-five microliters of the O157 bead suspensions were spread-plated onto sorbitol MacConkey with cefixime and tellurite (CT-SMAC) medium and plates will be incubated at 37° C for 20-24 hours. Up to six sorbitol-negative colonies from CT-SMAC were randomly picked and inoculated individually onto blood agar plates (Remel) and incubated at 37° C for 24 hours. Isolates obtained from the CT-SMAC plate were tested for agglutination for O157 antigen by latex agglutination, and if positive, tested for indole production. Isolates positive for agglutination and indole were tested by a six-plex PCR assay that targets *rfbEO157, fliCH7, eae*, *stx1, stx2*, and *ehxA* genes (88). Confirmation of *E. coli* O157:H7 were based on a gene profile positive for *rfbEO157, eae, fliCH7* and at least one Shiga-toxin gene (either *stx1* or *stx2*). Positive isolates were stored in CryoCare™ beads (Key Scientific Products, Round Rock, TX). All 317 isolates were cultured onto tryptone soy agar (TSA; BD Difco) slants and shipped overnight in cold storage to the University of Maryland.
5. PCR Method of Detection of non-O157 Serogroups

An aliquot of the fecal suspension (before and after enrichment) will be placed in 1.5 ml microcentrifuge tube, boiled for 10 min and centrifuged at 9,300 x g for 5 min. The supernatant will be purified with a GeneClean DNA extraction kit (MP Biomedicals, Solon, OH) and then subjected to an 11-plex PCR (89) targeting O157 and six major non-O157 serogroups (O26, O45, O103, O111, O121, O145) plus four virulence genes (\textit{stx}1, \textit{stx}2, \textit{eae}, and \textit{ehxA}).

6. DNA extraction and whole genome sequencing

DNA extraction of 317 isolates were performed by using the DNeasy Blood and Tissue kit with the protocol DNeasy Blood and Tissue – Bacterial Pellet – Standard on the QIAcube. By following the Illumina Miseq protocol, the DNA of each isolates were modified to appropriate concentration and sequenced on an Illumina Miseq by using Nextera XT DNA Library Preparation Kit and Miseq Reagent Kit V2 500 cycle (Illumina, San Diego, CA).

7. Data analysis

7.1 Genomic analysis

\textit{De novo} assembly was performed on Miseq raw reads by using SPAdes Genome Assembler 3.10.1 (Nurk, Bankevich et al., 2013). Quality check of the draft genomes was performed by using QUAST 4.0 (Gurevich et al., 2013). kSNP3 (Gardner, S.N., T. Slezk, and B.G. Hall. 2015.) was used to identify the SNPs of 317 draft genomes. Phylogenetic trees based on the SNP data was created by RAxML. 
\textit{(Need to be modified)} \textit{stx} and other gene identification was done using Python 3.5, Biopython, and BLAST, with parameters specified in the results (90, 91). A previously
published core-genome for all of *E. coli* was previously published and used to extract core-gene sequences using BLAST (92). We used the Virulence Factor Database as a reference to identify virulence genes in the genome assemblies (93). We used Genbank accessions NC_002128 and NZ_CP008807 for pO157Sakai and pSS17 gene references, respectively (94, 95).

7.2 Clade identification

Clade information of isolates were identified with BLAST as the clade structure based on SNPs differences in the study of Manning et al. (2008). A subtyping scheme based on 39 SNPs genotype was developed to distinguish of O157:H7 isolates into nine distinct evolutionary clades (96).

7.3 Statistical analysis

Analysis of variance (ANOVA) test was used to evaluate several groups of data: differences between the prevalence of O157 through three different diets, prevalence of O157 with and without ractopamine treatment, the prevalence of O157 on each isolation day of each diets treatment. A values of p<0.05 was considered significant of each group.
CHAPTER 3 Results

1. Clade information through Pens

The clade information of 286 isolates are shown in table 3.1. All 286 draft genomes were identified in three clades, which are clade 6, clade 7, and clade 8.

Table 3.1 Clade information for each diet

<table>
<thead>
<tr>
<th>Clade</th>
<th>UIP</th>
<th>DIP</th>
<th>CON</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade 6</td>
<td>90</td>
<td>46</td>
<td>69</td>
<td>205</td>
</tr>
<tr>
<td>Clade 7</td>
<td>12</td>
<td>29</td>
<td>8</td>
<td>49</td>
</tr>
<tr>
<td>Clade 8</td>
<td>7</td>
<td>12</td>
<td>13</td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>87</td>
<td>90</td>
<td>286</td>
</tr>
</tbody>
</table>

The distribution of each clade for each diet is shown in figure 3.1. Overall, the majority of *E. coli* O157:H7 isolates (205/286, 71.7%) were identified as clade 6 strains. 49 isolates (49/286, 17.1%) were clade 7 strains. 32 isolates (32/286, 11.2%) were clade 8 strains. The total of 87 isolates in high degradable protein diet (DIP) group, 46 isolates (52.9%) were belong to clade 6 strains, 29 (33.3%) were defined as clade 7 strains, and 12 (13.8%) isolates were clade 8 (Figure 3.2). For high undegradable protein diet (UIP) group, 90 of 109 isolates (82.6%) were clade 6 strains, 12 (11%) isolates were identified as clade 7 strains, and only 7 of 109 (6.4%) isolates were clade 8 strains (Figure 3.3). The total of 90 isolates of control diet (CON) were also identified into the same three clades: 69 (76.7%) clade 6 strains, 8 (8.9%) clade 7 strains, and 13 (14.4%) clade 8 strains (Figure 3.4).

Forty-eight pens (18 for each diet group) were included in this study. For isolates of DIP group, clade 6 strains were found in 10 pens, clade 7 strains were found in 6 pens, and clade 8 strains were occurred in 5 pens. Clade 6 strains were dominated in 8
(100%), 22 (100%), 31 (100%), 37 (100%), 43 (100%), 45 (100%) pens. Clade 7 strains were found to be domination in 11 (66.7%), 15 (100%), 18 (100%), 29 (75%), 41 (100%) pens. Clade 8 strains were found in 5 pens which 2 of 5 pens were dominated by them: 1 (83.3%), 20 (100%) (Figure 2). For UIP group strains, clade 6 strains dominated in most of the pens: 3 (83.3%), 6 (75%), 9 (100%), 12 (100%), 26 (100%), 28 (100%), 33 (100%), 36 (100%), 38 (94.4%), 40 (100%), 46 (100%), and 48 (100%). Clade 7 strains were identified in 4 pens and dominated in 14 (100%), 17 (66.7%) pens. Clade 8 strains were only found in three pens: 3 (16.7%), 21 (62.5%), 23 (50%) (Figure 3.3).

![Figure 3.1 Clade information for each diet](image-url)
Figure 3.2 Clade information of each pens with DIP diet

Figure 3.3 Clade information of each pens with UIP diet
2. Clade profiles with collection dates

Overall fecal prevalence of *E. coli* O157:H7 of each diet was shown in figure 3.5. For UIP diet group, there was a tendency of increasing between the first and the second sample collection (July 30 and August 12), but then decreased to 28.1% and 21.9% at the third and fourth collection day. The prevalence of *E. coli* O157:H7 in control diet group has been decreasing through the four collection day. The prevalence in DIP diet group shown a different pattern which ranged from 6.3% at the first collection day to 43.8% at the last collection day. The last collection date of each block contained ractopamine treatment. The prevalence of *E. coli* O157:H7 with and without ractopamine treatment on last collection day were shown in figure 3.6. The positive rate of *E. coli* O157:H7 for each diet group without ractopamine treatment were 12.5% for UIP, 45.8% for DIP, and 6.3% for CON. With ractopamine treatment, the positive rate for each diet group were 31.3% for UIP, 41.7% for DIP, and 22.9% for CON.
Figure 3.5 Prevalence of *E. coli* O157:H7 of each diet (n = 96)

Figure 3.6 Prevalence of *E. coli* O157:H7 with and without ractopamine treatment (n=48)

After combine the clade information and the prevalence data through times, the results have shown more clear information. The population tendency of *E. coli* O157:H7 of UIP and CON diet group had been declining over months (Figure 3.7,
Figure 3.9). Figure 3.7 shown the prevalence of *E. coli* O157:H7 of UIP diet group with clade information. Although clade 8 strains were collected on the last collection day, none of them were associated with ractopamine treatment (Figure 3.8). The results of CON diet group shown the similar pattern with the results of UIP diet group (Figure 3.9 and figure 3.10).

![Graph 1](image1.png)

**Figure 3.7 Prevalence of *E. coli* O157:H7 on each collection day of UIP group with clade information (n=96)**

![Graph 2](image2.png)
Figure 3.8 Relative prevalence of E. coli O157:H7 with UIP diet between rh treatment and without rh treatment (n=48)

![Bar chart showing relative prevalence of E. coli O157:H7 with UIP diet between rh treatment and without rh treatment.

Figure 3.9 Prevalence of E. coli O157:H7 on each collection day of CON group with clade information (n=96)

![Bar chart showing prevalence of E. coli O157:H7 on each collection day of CON group with clade information.

Figure 3.10 Relative prevalence of E. coli O157:H7 with CON diet between rh treatment and without rh treatment (n=48)

![Bar chart showing relative prevalence of E. coli O157:H7 with CON diet between rh treatment and without rh treatment.

For DIP diet group, the prevalence of E. coli O157:H7 on each collection day...
and clade information were shown in figure 3.11. The positive rate increased to 43.8% on the last set of collection day. More clade 6 and clade 7 strains associated with rh treatment were collected, but clade 8 strains appeared more without ractopamine treatment (Figure 3.12). However, we still cannot conclude that the effect of ractopamine on fecal shedding _E. coli_ O157:H7.

![Figure 3.11 Prevalence of _E. coli_ O157:H7 on each collection day of DIP group with clade information (n=96)]
Figure 3.12 Relative prevalence of E. coli O157:H7 with DIP diet between rh treatment and without rh treatment (n=48)

3. Virulence genes

We blasted a database of known E. coli effectors were used along with some putative type III effector-like genes, plus other pathogenicity genes, such as hemolysins, against our genomes by using a cutoff of 90% sequence identity and 60% gene length for calling a gene present or absent. We used a process that assigned partial hits to genes if they met the sequence identity threshold, then added those lengths up to calculate the total gene length present. This process is more sensitive in calling genes present, but problems with assembly, such as low quality of the assembled draft genomes, may not be able to distinguish well between multiple alleles. (This paragraph need to be modified)

The blast results showed the same information of shiga-toxin genes with PCR results: all isolates were positive for stx2 gene but no stx1 gene were found in all isolates. Stx2a gene were found in all clade 8 strains (n=32), stx2c gene were found in all clade 6 (n=205) and clade 7 (n=49) strains (table 2).

Table 3.2 stx gene profiles

<table>
<thead>
<tr>
<th></th>
<th>stx1</th>
<th>stx2a</th>
<th>stx2c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade 6 isolates (n=205)</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Clade 7 isolates (n=49)</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Clade 8 isolates (n=32)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>
Among LEE-encoded genes, all isolates were positive for intimin (*eae*), translocated intimin receptor protein (*tir*), type III secretion effectors (*esp*B, *esp*D, *esp*H), but all isolates were negative for *esp*F gene (figure 3.13). For non-LEE-encoded type III secreted effectors, all clade 6 isolates (n=205) were negative for *esp*X6 gene, but clade 7 (n=49) and 8 (n=32) isolates were all positive for *esp*X6 gene. Two *nle* variants, *nle*F and *nle*H2 genes were only presented in all clade 6 (n=205) and clade 8 (n=32) isolates but were not observed in any clade 7 (n=49) isolates.

<table>
<thead>
<tr>
<th></th>
<th>Clade6</th>
<th>Clade7</th>
<th>Clade8</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>eae</em></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>tir</em></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>esp</em>B</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>esp</em>D</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>esp</em>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>esp</em>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>esp</em>X6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>nle</em>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>nle</em>H2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>toxB</em></td>
<td>80</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td><em>stcE</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><em>hly</em></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3.13 Selected virulence factors results

The LEE pathogenicity island chromosomally encodes for the intimin gene (*eae*) for attaching and effacing, translocated intimin receptor gene (*tir*), and Type III secretion system effector genes, all critical components of the Type III secretory system in EHEC and EPEC. These proteins (*esp*B, *esp*A, *esp*D, *tir*) are exported via a type III secretion pathway similar to other specialized protein secretion systems that are shared among a variety of Gram-negative pathogens (97). Whereas an *esp*F deletion mutant does not synthesize or secrete EspF, surprisingly it retains the ability to induce host signaling events, perform A/E activities, and invade host epithelial cells (97).
4. Phylogenetic relationships

Phylogenetic tree was created by analyzing the SNP data of 286 isolates on RAxML with 100 bootstraps and the GTRGAMMA substitution model. The output file (.tree) was annotated using FigTree 1.4.2 (figure 3.14). Overall, the 286 isolates clustered based on the clade information. Outside the tree, there are two datasets: four collection days represented by orange rings and different diets of all strains. The first ring represents the first collection day (July 20\textsuperscript{th}). The second ring indicates the second collection day (August 12\textsuperscript{th}). The third ring represents third collection day (August 30\textsuperscript{th} or September 2\textsuperscript{nd}). The fourth ring represents fourth collection day (September 27\textsuperscript{th} or October 4\textsuperscript{th}). The fifth ring represents the different diets treatment of all strains: UIP is filled with purple, DIP is filled with light green, and CON is filled with yellow.
Figure 3.14 The phylogenetic tree based on SNP differences

The clade in original phylogenetic tree were collapsed with branch length distance to leaves level below 0.02 (figure 3.15). The stx gene information is also shown in figure 8. In 32 clade 8 strains, only one strain (KSU146) was collected on the first isolation day; ten strains were collected on the second isolation day; two strains were collected on the third isolation day; nineteen strains were collected on the fourth isolation day (figure 3.13). Among 49 strains of clade 7, three strains were collected on the first isolation day; 15 strains were collected on the second isolation day; four strains were collected on the third isolation day; 27 strains were collected on fourth isolation
day (figure 3.14). Among 203 strains of clade 6, 59 strains were collected on the first isolation day; 57 strains were collected on the second isolation day; 58 strains were collected on the third isolation day; and 30 strains were collected on the fourth isolation day.

Figure 3.15 SNP-based, collapsed phylogenetic tree

5. SNPs analysis

The SNPs count tree was performed by using Python. Numbers of SNPs of each isolates and branches of phylogenetic tree were observed. Among isolates of clade 8, KSU146 and KSU316 share 3 unique SNPs. KSU146 was collected on the first
collection day (July 20th) from pen 13 and no treatment was given. KSU316 was collected on the fourth collection day (October 4th) from pen 23 and UIP diet treatment was given. KSU117 and KSU081 shared three unique SNPs. These two isolates were collected from the fourth collection day and same diet treatment, DIP, were given. KSU081 was collected from pen 34, and KSU117 was collected from pen 1. There are two main lineages (7A and 7B) of isolates of clade 7. There are 63 SNPs differences between lineage 7A and lineage 7B. Among isolates of lineage 7A, KSU078 and KSU100 shared 6 unique SNPs. KSU078 was collected on the third collection day (September 2nd) from pen 11 with DIP diet treatment. KSU100 was collected on the fourth collection day (October 4th) from pen 14 with UIP diet treatment. Among isolates of lineage 7B, KSU212 and KSU202 shared 1 unique SNP. KSU212 was collected from the second collection day (August 12th) from pen 19 with CON diet treatment. KSU202 was collected from the second collection day (August 12th) from pen 15 with DIP diet treatment. KSU203 and KSU303 shared 98 unique SNPs. These two isolates were collected from the same pen (pen 16) with same diet treatment (CON). However, KSU203 was collected on the second collection day (August 12th), and KSU303 was collected on the fourth collection day (October 4th). Among the isolates of clade 6, KSU264 and KSU263, shared 248 unique SNPs. They were collected from the same collection day (August 12th) and the same pen (pen 3). Same diet treatment, UIP, were given to pen 3. KSU 314 and KSU305 shared 84 unique SNPs. They were from different pens (pen 17 and 21) but collected from the same day (October 4th) with UIP diet treatment. KSU260 and KSU192 shared 2 unique SNPs. KSU260 was collected on the third collection day (August 30th) from pen 48 with UIP diet treatment. KSU 192
was collected on the second collection day (August 12th) from pen 9 with UIP diet treatment. KSU283 and KSU248 shared 3 unique SNPs. KSU 283 was collected on fourth collection day (September 27th) from pen 31 with DIP diet treatment. KSU 248 was collected on third collection day (August 30th) from pen 40 with UIP diet treatment. KSU265 and KSU 189 shared 2 unique SNPs. KSU265 was collected on the third collection day (September 2nd) from pen 3 with UIP diet treatment. KSU 189 was collected on the second collection day (August 12th) from pen 8 with DIP diet treatment.
CHAPTER 4 Discussion

1. Clade lineages of *E. coli* O157:H7 isolates

   Based on the distribution of clade information of *E. coli* O157:H7 for each diet, differences were observed. More isolates (n=109) were collected from the pens with high undegradable protein diet than isolates collected from high degradable protein (n=87) and control diet (n=90). High percentage of isolates belonging to clade 6 were collected from pens with UIP diet and CON diet than from those with DIP diet. However, clade 7 isolates were the majority (33.33%) among pens with DIP diet. In addition, a lower percentage (6.42%) of clade 8 isolates were observed with UIP diet.

2. Virulence genes

   For the 286 isolates, the distribution of stx genes information was very interesting. Only stx2a gene was found in the 32 isolates of clade 8. Strains of clade 8 lineage have caused two recent and unusually severe outbreaks linked to produce, are associated with HUS, and more frequently carry both stx2 and stx2c genes (Table 4.1). Eppinger, et al. (2011) reported that stx2a and stx2c genes were detected in clade 8 strains associated with the spinach and Taco Bell of outbreaks of *E. coli* O157:H7 infections in 2006. A second set of stx genes was found in lineage I/II clade 8 strains within Stx2c-converting prophage integrated at the sbcB locus (ECH74115_2944) (98). The Stx2c prophage is a unique genomic feature of the spinach and Taco bell outbreak strains and the HUS strain EC508 (99). The Stx2-converting phages target two chromosomal loci (wrbA and argW) and could differ in their regulatory circuits and verotoxin expression levels (100). In addition, in clade 6 (n=205) and clade 7 (n=49) isolates, only stx2c
gene was identified. Unlike the Stx2 prophage, the Stx2c and Stx1 prophages seem to target preferentially the sbcB and yehV loci, respectively. These alterations in phage inventory and dynamics are mediated by unknown prophages and phage combinations as well as by varying integration loci (101), particularly in the stx converting phages (98).

Table 4.1 Shiga toxin genes profile of isolates in this study and outbreak strains

<table>
<thead>
<tr>
<th></th>
<th>stx1</th>
<th>stx2a</th>
<th>stx2c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade 6 isolates (n=205)</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Clade 7 isolates (n=49)</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Clade 8 isolates (n=32)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>EC4045</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4042</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4113</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4076</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4084</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4127</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4191</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TW14359 (=EC1611)</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4001</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EC4205</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

3. Re-sequence

In this study, we performed whole genome sequencing on Illumina Miseq. The Illumina sequencing workflow was based on three simple steps: libraries are prepared from virtually any nucleic acid sample, amplified to produce clonal clusters, and sequenced using massively parallel synthesis (82). The density of clonal clusters has a large impact on sequencing performance in terms of data quality and total output (82). The goal of a sequencing run is to sequence at high enough densities to maximize total data output, while maintaining low enough densities to avoid the negative effects of
overclustering. The quality of results basically reflect on three data: reads passing filter (%PF), Q30 scores (%Q30), and the size of data output. The %PF is an indication of signal purity from each cluster. Overclustered flow cells typically have higher numbers of overlapping clusters, which leads to poor template generation, and then causes a decrease in the %PF metric. The Q30 score is a prediction of the probability of an incorrect base call 1 in 1000 times. The ratio of base intensity to background for each base is decreased due to overload signal intensities. This decrease often results in ambiguity during base calling, and leads to a decrease in data quality (82). The data output is based on the number of samples and the type of organisms.

It was hard to identify the problem in terms of the multiple steps of manual experiment. We got three sequence results with low quality scores. The first run was performed on November 15th, 2016, with low %PF (52.54%), low average%Q30 (64.91%), and low data output (1.75Gbp). We did another two separate libraries for another two runs which still showed low average%Q30 (31.43% and 32.98%). Then we tried third re-sequencing with the results of 89.33%PF and 83.35%Q30. In addition, we excluded the possible errors of manual library preparation and quantification. The reason why we got low quality of several runs is still unknown. We concluded that there might be some problem associated with the flow cells. The flow cell need to be clean-up manually before sequencing. It is possible that the flow cells did not get cleaned thoroughly. Another reason might be the flow cells were damaged during the transporting. For further sequencing, methods such as qPCR can be performed to ensure the accuracy of quantification of library; Bioanalyzer or similar technology can be used for verification of library quality in order to check for library integrity, average
insert size, and the presence of contamination (82).
CHAPTER 5 Conclusion

This study undertook a whole genome comparative analysis of *E. coli* O157:H7 isolates collected from feces of cattle fed with different composition of dietary protein. Our findings indicate that the change of dietary ingredients could affect the colonization and fecal shedding of STEC, and increased flow of undegradable protein may increase the shedding of *E. coli* O157. The pen layout could be a major factor of fecal shedding of STEC. However, it is important to note that other multiple factors could contribute to fecal shedding of *E. coli* O157:H7 in cattle.
References


72. Edrington TS, Callaway TR, Smith DJ, Genovesi KJ, Anderson RC, Nisbet DJ. Effects


