COMPARATIVE ANALYSIS REVEALS A HYPER-VIRULENT
*ESCHERICHIA COLI* O157:H7 STRAIN ISOLATED FROM A SUPER-SHEDDER

By

LIN TENG

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2016
To my dear parents, Lihang Wang and Yongyu Teng, and my dear advisor, Dr. KwangCheol Casey Jeong, who is leading me into the world of science
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. KwangCheol Casey Jeong, for his generous support, continuous patience and strict training during my entire M.S. program. His training and direction help me culture my scientific thinking and evoke my interests to explore the unknown. Thanks to Dr. Soohyoun Ahn for her priceless encouragement and directions in my thesis. I appreciate Dr. Nicolas Dilorenzo’s kind help for the arrangement of sample collections at North Florida Research and Education Center. Thanks to Dr. Dongjin Park for working together with me on this project and giving me lots of detailed directions.

Special thanks to Raies Mir, Zhengxin Ma, Dr. Alejandro Garrido-Maestu, Dr. Sarah Markland, Minyoung Kang, Choonghee Lee Amber Ginn and Shinyoung Lee. Raies has been taking care of me since the first day I joined the lab. With his help and directions, I was able to quickly adjust to a new life at the University of Florida. Thanks to Zhengxin for all the detailed directions she gave me for the experiments, course work and daily life throughout my M.S. career. I appreciate the kind and priceless help and encouragements from Alejandro in the toughest period during my M.S. program. Sarah not only helped me with my experiments but also taught me to be optimistic to overcome the problems I met in the experiments. Minyoung always supported me with the kind help in the experiments whenever I needed. Choonghee taught me how to conduct Western blot and maintained the cell lines for my experiments. Amber taught me phylogenetic analysis and never got tired of my endless questions. I would like to acknowledge Shinyoung for her efforts to optimize the condition of biofilm assay. In addition, I appreciate the help from Cheng Ye and Guan Yang. Even though they are not my lab members, they always cheered me up and motivated me to move forward.
I would like to express my deepest and greatest gratitude to my parents, Lihang Wang and Yongyu Teng for their unconditional support throughout the years. I would also like to thank my girlfriend, Lanwen Zhou, for her support and encouragement during the years.
TABLE OF CONTENTS

ACKNOWLEDGMENTS.................................................................................................................. 4

LIST OF TABLES.......................................................................................................................... 8

LIST OF FIGURES.......................................................................................................................... 9

LIST OF ABBREVIATIONS........................................................................................................... 10

ABSTRACT.................................................................................................................................. 11

CHAPTER

1 INTRODUCTION ....................................................................................................................... 13

2 LITERATURE REVIEW ............................................................................................................. 15

   Shiga Toxin-Producing Escherichia coli O157:H7 ................................................................. 15
   Virulence Factors and Virulence Related Genes ................................................................. 15
      Toxins ................................................................................................................................. 16
      Adhesion .............................................................................................................................. 16
      Secretion Systems ............................................................................................................. 17
      Effectors .............................................................................................................................. 17
   Reservoir and Super-Shedder ............................................................................................... 18
   Role of Super-Shedder ........................................................................................................... 19
   Factors Contributing to Supper-Shedding .......................................................................... 20
      Microbial Factors in Super-Shedding ............................................................................ 21
      Animal Factors Contributing to Super-Shedding ............................................................ 25
   Environmental Factors Contributing to Super-Shedding ..................................................... 26
   Conclusion .............................................................................................................................. 26

3 COMPARATIVE ANALYSIS REVEALS A HYPER-VIRULENT Escherichia coli O157:H7 STRAIN ISOLATED FROM A SUPER-SHEDDER ................................................................. 28

   Background .............................................................................................................................. 28
   Materials and Methods ......................................................................................................... 32
      Bacterial Strains .................................................................................................................. 32
      Genomic DNA Extraction ................................................................................................. 32
      Stx2 Subtyping .................................................................................................................... 32
      Phage Induction ................................................................................................................... 33
      Western Blot Analysis of Stx2 Expression ....................................................................... 33
      Adherence to HEp-2 Cells ................................................................................................. 34
      Static Biofilm Formation Assay ....................................................................................... 35
      Genome Sequencing and Assembly ................................................................................. 36
      Annotation and Comparative Genomics ............................................................................ 36
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Genome statistics of KCJ1266 and reference O157 strains</td>
<td>52</td>
</tr>
<tr>
<td>3-2</td>
<td>Function-based gene comparison</td>
<td>53</td>
</tr>
<tr>
<td>3-3</td>
<td>Genome encoded in the 19 phage regions of KCJ1266</td>
<td>54</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3-1</td>
<td>Shiga-toxin II detection.</td>
<td>56</td>
</tr>
<tr>
<td>3-2</td>
<td>Adhesion of <em>E. coli</em> O157:H7 isolates to Hep-2 cells.</td>
<td>57</td>
</tr>
<tr>
<td>3-3</td>
<td>Biofilm assay of KCJ1266. EDL933 was used as a positive control and ΔfimH mutant of EDL933 was used as a negative control.</td>
<td>58</td>
</tr>
<tr>
<td>3-4</td>
<td>IS629 profile of KCJ1266, SS17, EC4115 and EDL933.</td>
<td>59</td>
</tr>
<tr>
<td>3-5</td>
<td>Genome map of KCJ1266.</td>
<td>60</td>
</tr>
<tr>
<td>3-6</td>
<td>Mauve alignments of KCJ1266 with the three reference genome.</td>
<td>61</td>
</tr>
<tr>
<td>3-7</td>
<td>Comparison of O157 plasmids of KCJ1266 and other 3 reference strains</td>
<td>62</td>
</tr>
<tr>
<td>3-8</td>
<td>Subsystem category distribution.</td>
<td>63</td>
</tr>
<tr>
<td>3-9</td>
<td>Phage comparison of KCJ1266 with reference strains.</td>
<td>64</td>
</tr>
<tr>
<td>3-10</td>
<td>Virulence related genes in KCJ1266 and reference strains.</td>
<td>65</td>
</tr>
<tr>
<td>3-11</td>
<td>Strains are clustered into three groups which belong to Lineage I (green), Lineage II (red) and Lineage I/II (blue).</td>
<td>66</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>CFU</td>
<td>Colony-forming unit</td>
<td></td>
</tr>
<tr>
<td>GI</td>
<td>Gastrointestinal</td>
<td></td>
</tr>
<tr>
<td>HUS</td>
<td>Hemolytic-uremic syndrome</td>
<td></td>
</tr>
<tr>
<td>HV</td>
<td>Hype-virulence</td>
<td></td>
</tr>
<tr>
<td>IDP</td>
<td>IS629 distribution profile</td>
<td></td>
</tr>
<tr>
<td>LEE</td>
<td>Locus for enterocyte effacement</td>
<td></td>
</tr>
<tr>
<td>LSPA</td>
<td>Lineage specific polymorphism analysis</td>
<td></td>
</tr>
<tr>
<td>nsSNP</td>
<td>Nonsynonymous single-nucleotide polymorphism</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
<td></td>
</tr>
<tr>
<td>PFGE</td>
<td>Pulsed-field gel electrophoresis</td>
<td></td>
</tr>
<tr>
<td>RAJ</td>
<td>Recto-anal junction</td>
<td></td>
</tr>
<tr>
<td>SB</td>
<td>Synteny block</td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>Single-nucleotide polymorphism</td>
<td></td>
</tr>
<tr>
<td>SS</td>
<td>Super-shedder</td>
<td></td>
</tr>
<tr>
<td>sSNP</td>
<td>Synonymous single nucleotide polymorphism</td>
<td></td>
</tr>
<tr>
<td>STEC</td>
<td>Shiga toxin-producing <em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>TBST</td>
<td>20 Tris-Buffered Saline and Tween 20</td>
<td></td>
</tr>
<tr>
<td>tir</td>
<td>translocated intimin receptor</td>
<td></td>
</tr>
<tr>
<td>TTSS</td>
<td>Type three secretion system</td>
<td></td>
</tr>
<tr>
<td>WGS</td>
<td>Whole genome sequencing</td>
<td></td>
</tr>
</tbody>
</table>
Shiga toxin-producing *Escherichia coli* (STEC) O157:H7 is a foodborne pathogen that threatens public health on a global scale. STEC O157:H7 predominantly colonizes the terminal recto-anal junction (RAJ) of cattle, which is the major asymptomatic reservoir of this pathogen. Cattle shedding STEC O157:H7 ≥ 10^4 CFU/g of feces are known as super-shedders and are responsible for within-farm and between-farm transmission of STEC O157:H7. The purpose of this study was to perform genetic characterization of KCJ1266, a strain isolated from a super-shedder steer from a farm in North Florida. PacBio sequencing was employed for whole genome sequencing (WGS) to characterize the genomic features of KCJ1266. A comparative genome analysis of KCJ1266 with reference genomes including SS17 (strain isolated from a super-shedder), EC4115 (strain related to spinach outbreak) and EDL933 (strain related to hamburger outbreak) was conducted. WGS of KCJ1266 revealed that it has a genome of 5,478,683 bp encoding 5,545 open reading frames and a plasmid, pO157, of 95,910 bp. *In silico* analysis revealed that KCJ1266 belongs to *E. coli* lineage I/II and clade 8,
which are related to disease-causing isolates. In addition, Mauve alignment showed that KCJ1266 shares a similar genomic architecture with SS17 and EC4115. Comparative analyses also revealed that KCJ1266 has the same virulence and similar functional genes as SS17 and EC4115. Phylogenetic analysis showed that KCJ1266, SS17 and EC4115 clustered in the same group. Taken together, these results reveal that super-shedding STEC strain KCJ1266 is a hyper-virulent strain similar to that of SS17 and EC4115. This is one of the first studies which has utilized PacBio WGS to characterize a hyper-virulent STEC O157:H7 strain isolated from super-shedding cattle. KCJ1266 can potentially be used as a reference strain for future studies regarding the phenomenon of super-shedding.
CHAPTER 1
INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are foodborne pathogens that threaten public health on a global scale [1]. It causes a variety of symptoms including bloody diarrhea and life-threatening hemolytic uremic syndrome (HUS) in humans through the expression of Shiga-like toxin [2-4]. STEC infections cause annual economic loss of more than $1 billion USD in the US [5]. The human infection caused by STEC is primarily attributed to the consumption of food products contaminated by STEC.

Cattle are the primary reservoirs and asymptomatic carriers of STEC. The terminal recto-anal junction (RAJ), as well as gastrointestinal (GI) tract, are the principle localization sites of STEC [5]. Among the cattle colonized by STEC, a subset of cattle shed *E. coli* O157:H7 at a level of > 10^4 colony forming unit (CFU) per gram feces and are defined as “super-shedders” [6, 7]. Previous studies showed that super-shedders were responsible for most of the within-farm and between-farm transmission events of STEC [1, 8, 9]. The existence of super-shedders may lead to an increase in the contamination of animal food products, such as beef and milk, and thus threaten public health.

Since the phenomenon of super-shedding was first observed, researchers have been trying to determine the bacterial, host, and environmental factors that contribute to super-shedding [10]. However, the mechanisms of super-shedding are still unknown. In order to determine the potential bacterial factors that contribute to super-shedding, we hypothesized that the super-shedding phenomenon is caused by a group of specific genomic traits of bacterial strains isolated from super-shedding cattle. The objective of
this study was to identify the genes which contribute to the specific phenomenon of super-shedding using comparative genome analysis.
CHAPTER 2
LITERATURE REVIEW

Shiga Toxin-Producing *Escherichia coli* O157:H7

Shiga toxin-producing *Escherichia coli* O157:H7 (STEC) are foodborne pathogens that globally threaten public health [11]. In 1982, *E. coli* O157:H7 was first identified as a foodborne pathogen following an outbreak of hemorrhagic colitis in the United States [1]. In 1993, a multistate outbreak of *E. coli* O157:H7 led to the recognition of *E. coli* O157:H7 as a crucial pathogen in the U.S [12]. Reports of *E. coli* O157:H7 outbreaks from 2003 to 2012 increased when compared to the preceding two decades. *E. coli* outbreaks are historically associated with not only beef products, but also a variety of food sources, including unpasteurized apple juice, spinach and salami [13, 14]. *E. coli* O157:H7 quickly became one of the most recognized foodborne pathogens due to the severity of the disease it caused in humans and the significant media attention seen during outbreaks. STEC causes a variety of symptoms ranging from bloody diarrhea to the development of life-threatening hemolytic-uremic syndrome (HUS). Previous research has shown that STEC infections cause more than $1 billion USD in economic loss per year [5].

**Virulence Factors and Virulence Related Genes**

Bacterial virulence and pathogenesis factors play a significant role in infectious disease etiology. With the development of whole genome sequencing techniques, an increasing number of virulence factors can be identified by whole genome comparison [15]. A recent review concluded that about 394 coding sequences were identified as putative or known virulence factors [15]. Without a good animal model for STEC infection, it is difficult to determine the contribution of putative virulence genes to its
pathogenesis [2], therefore most of the putative virulence genes will not be discussed in this review.

**Toxins**

Shiga toxins (Stxs) are encoded by Shiga-toxin converting prophages in the chromosome of STEC. The diversity of Stxs are caused by the high heterogeneity of Shiga-toxin converting prophages [16]. There are two major types of Shiga toxins, Stx1 and Stx2, which are encoded by two main sets of genes, \( stx1AB \) and \( stx2AB \), respectively. Both of Stx1 and Stx2 consist of an A subunit of 32-kDa and five identical B subunits of 7.7kDa [17]. The main receptor of Shiga toxin is globotriaosylceramide (Gb3) which can be found on human cells, including endothelial cells, epithelial cells and microvascular endothelial cells [18]. Although Stx1 and Stx2 have similar structure, their toxic dosage to different cell types vary. An *in vitro* study showed that Stx2 toxin was reported to be 1000 times more cytotoxic in human renal microvascular endothelial cells when compared with Stx1 toxin [19]. In addition, the Stx2 subtypes encoded by \( stx2a, stx2b, stx2c, stx2d, stx2e \) and \( stx2f \) differ by a few amino acids, resulting in diverse toxicity of Stx2 in humans [20]. Among these Stx2 subtypes, Stx2a, Stx2c and Stx2d are often related to severe human diseases such as hemorrhagic colitis (HC) and HUS [21-24]. A recent study showed that STEC O157:H7 harboring the \( stx2a \) gene can express Shiga toxin at higher levels than those harboring the \( stx2c \) gene [25].

**Adhesion**

The ability of STEC to adhere to human epithelial cells is believed to be a critical step for infection because adherence is the initial step in STEC infection [15]. Fimbriae and curli are two major apparatuses which contribute to adherence ability of STEC. Type I fimbriae are encoded by the *fim* operon, including *fimB, fimE, fimA, fimI, fimC*,
*fimF*, *fimG* and *fimH*. Curli play a significant role in surface attachment and biofilm formation. The formation of curli is controlled by the *cgs* operon [26]. Long polar filament encoded by the *lpfABCC'DE* operon are also involved in adherence [27].

**Secretion Systems**

*E. coli* O157 harbors a highly conserved plasmid known as pO157. There are 13 ORFs (open reading frames) named *etpC* to *eptO* which encode a type two secretion system in pO157 [2]. It has been found that this secretion system can secrete StcE (zinc metalloprotease protein), thus it contributes to the virulence of *E. coli* O157:H7 [15]. Another secretion system in STEC O157:H7 is the type three secretion system (TTSS), a needle complex apparatus, by which bacterial effectors are injected into host cells[28]. This apparatus, made up of more than 20 proteins, is necessary for STEC to colonize to cattle. The structure of the TTSS is encoded by genes, including *eprK, eprJ, eprH, epaQ, epaP, epaO, elvJ, elvl, elvA, elvE, elvG and elvF* [15]. When STEC start colonization, Stx prophages are activated by SOS induction followed by cell lysis of a subgroup of STEC [28]. The Stx produced by Stx prophages bind to its receptor, Gb3, on epithelia cells, resulting in redistribution of nucleolin and other factors to the cell surface [28]. The changed surface structure promotes bacterial attachment by binding to bacterial surface protein intimin (Eae) [28].

**Effectors**

Virulence effectors refer to the proteins that are injected to host cells using TTSS [15]. STEC O157:H7 are able to colonize the intestinal mucosa and cause attaching and effacing (A/E) lesions. The regions encoding proteins related to A/E lesion were designated the locus of enterocyte effacement (LEE) [29]. The *espH, espZ* and *espG* are effector genes encoding in LEE. In addition to effectors encoded by LEE, there are
numerous non-LEE-encoding effectors. These effectors are encoded by genes including \texttt{espY4, nleB1, espW, nleG8-2, espM2, espR4, nleA, nleH1-2, espM1, nleG2-2, nleG6-1, nleG5-1, espK, espX2, espY1, nleG8-1, nleD, and nleH1-1} [15].

**Reservoir and Super-Shedder**

Cattle are the primary asymptomatic reservoir of STEC [30]. However, the shedding amount of \textit{E. coli} O157:H7 varies greatly from animal to animal [8, 31, 32]. The number of \textit{E. coli} O157:H7 shed in feces is also affected by the climate, season and breed of cattle. Previous reports showed that the shedding of \textit{E. coli} O157:H7 was observed as a transient event and the typical duration of \textit{E. coli} O157:H7 shedding lasts about 1 month [33]. The term “super-shedder”, was first used by Matthews \textit{et al.} in 2006, which is defined as a subset of cattle that can shed \( \geq 10^4 \) colony-forming unit (CFU) /g feces of \textit{E. coli} O157:H7 [6]. However, the term “super-shedding” are often used to describe the heterogeneous role that the cattle play in the transmission of \textit{E. coli} O157:H7. Based on a review written by Chase-Topping \textit{et al.} (2008) the super-shedder is defined as the individual that can shed an extremely high number of infectious organisms, by which the super-shedder has an increased probability of infecting other individuals [1].

However, a part of the definition remains unclear, specifically what defines the duration of the super-shedding state. In a previous study, the duration of super-shedding lasted up to a month [33]. Since the duration of super-shedding may affect the identification of a super-shedder, a proper sampling method for isolation of super shedding bacteria from animals is extremely important. To get an accurate account of shedding information of \textit{E. coli} O157:H7, Munns \textit{et al.} (2015) modified the sampling method and had each animal sampled twice a day during a 6-day period [10]. The
results of that study revealed that super-shedding is a short-term phenomenon and that shedding amount may not be consistent even between two samplings within the same day [10]. In this case, it is not easy to identify the low-shedder and super-shedder, since their status may change within one day.

**Role of Super-Shedder**

Super-shedders are of significant importance because they are considered to be responsible for the within-farm and between-farm transmission of *E. coli* O157:H7, as well as transmission of *E. coli* O157:H7 into the environment [1]. It has been observed that *E. coli* O157:H7 are heterogeneously shed by cattle within the same farm [6]. In a previous study, more than 96% of total *E. coli* O157:H7 were shed by about 9% of the total 761 cattle tested [34]. In a mathematical modelling study, Matthews *et al.* (2006) [6] showed that 20% of the most infectious individuals of cattle were responsible for about 80% of the infections that occurred within the same farm. Stephens *et al.* (2009) [9] reported that super-shedders were responsible for 47% of within-farm shedding of *E. coli* O157:H7, while this ratio increased to more than 90% in another study [1]. A study involving Scottish farms showed evidence of between-farm transmission of *E. coli* O157:H7, where 15 of 105 pulsed-field gel electrophoresis (PFGE) patterns were detected in more than one farm [35]. Jeon *et al.* (2013) traced the presence of *E. coli* O157:H7 within a herd of cattle after they were transported to a commercial feedlot, finding that the PFGE pattern of *E. coli* O157:H7 isolated in the feedlot was identical to that of commercial farm isolates [8]. This result indicated that *E. coli* O157:H7 could be transferred between farms.

Since super shedder *E. coli* O157:H7 strains may express high level of Shiga toxin, they, with their high shedding load, are a potential threat to public health. A study
performed by Chase-Topping *et al*. (2007) showed that the most common phage type in Scotland was PT 21/28 (41%) [36]. This was found to be correlated with another study in which 58% of *E. coli* O157:H7 human infections were caused by PT 21/28 *E. coli* O157:H7 [37]. There is more evidence indicating the importance for the control and surveillance of super-shedding animals, and the types of bacteria that are being shed in high concentrations. Cote *et al*. (2015) found that an *E. coli* O157:H7 strain isolated from a super-shedding animal in the U.S. Midwest possessed a hyper-virulent characteristic having two copies of the *stx2a* gene, allowing the bacteria to produce high level of Shiga toxin 2 without induction [26].

**Factors Contributing to Supper-Shedding**

Cattle live in various environments, interact with each other and are exposed to different environmental pressures. There may be three major factors, including animal factors, pathogen factors and environmental factors that cause the super-shedding phenomenon [1, 10]. The animal factors include the status of the animal’s immune system, susceptibility to intestinal colonization, animal to animal transmission, stress response and animal breed [8, 10]. The environmental factors include season, animal housing hygiene, and diet. The bacterial factors include the ability of *E. coli* O157:H7 to survive in the gastrointestinal tract, competiveness of super-shedder *E. coli* O157:H7 vs. commensal *E. coli*, potential of *E. coli* O157:H7 to proliferate in rectal contents and the ability to form biofilms. To date, the most studied among these factors are bacterial factors, while experiments regarding host and environmental factors are not well studied.
**Microbial Factors in Super-Shedding**

Previous studies reported that the *E. coli* O157:H7 isolated from super-shedders in Scotland shared a primary phage type, PT21/28, suggesting that PT21/28 phenotype may be a marker of super-shedding strains [1, 36]. However, another study reported that no PT21/28 were identified, while the most common phage type affecting cattle was PT4 [38]. Another study [39] examined the phage type in cattle in Canada, finding that PT14a and PT8 were the most popular among all five of the identified phage types. Although certain primary phage types are common in certain farms or regions, there is no one specific phage that is commonly found on a global scale. These results showed that phage type may not be a good marker for super-shedding strains.

PFGE is believed to be the “gold standard” for epidemiological studies involving outbreaks [40]. In some of the previous studies, PFGE profiles showed that the strains isolated from super-shedders and non-super-shedders have similar PFGE patterns. Jeon *et al.* (2013) reported that the isolates of super-shedders, including medium-shedders and low-shedders in North Florida, have the same PFGE patterns, indicating that these strains are derived from the same source and have the same subtype [8]. Another study performed by Munns *et al.* (2015) showed a similar result that isolates of *E. coli* O157:H7 from super-shedders and low-shedders contained three similar PFGE patterns [10]. Among all these isolates, 97.6% (123/126) of them had the same PFGE pattern. These results may reveal that strains with specific characteristics (PFGE patterns) may have better adaptation abilities than other strains to survive and proliferate within the host. In contrast, some other studies also showed that the super-shedders and low-shedders shared different PFGE patterns. According to a study by Stanford *et al.* (2012) PFGE patterns of *E. coli* O157:H7 isolates from super-shedders
and low-shedders of two commercial farms were varied [41]. Another study reported by Dodd et al. (2013) confirmed the theory that super-shedders do not necessarily share the same PFGE subtype. Arthur et al. (2013) identified 102 stains isolated from cattle of super-shedders, finding 52 distinct PFGE subtypes [38]. However, there were no specific PFGE subtypes found to be responsible for super-shedding. All these results suggest that there is not a specific PFGE pattern of E. coli O157:H7 that is related to super-shedding.

Octamer-based genome scanning methods could discriminate E. coli O157:H7 according to the size of gene markers. By using octamer-based genome scanning and microarray-based comparative genomic hybridization, E. coli O157:H7 can be grouped into three distinct lineages [42, 43]. These three lineages, lineage I, lineage II and lineage I/II, are usually isolated from different sources. In most of the cases, lineage I strains are composed of human clinical isolates. Lineage II isolates usually originate from bovine sources. The lineage I/II isolates are related to strains that can cause human disease, including foodborne outbreaks [44]. Within these three lineages, lineage I and I/II are more likely to cause human disease. Studies showed that a potential hyper-virulent strain isolated from a super-shedding animal belonged to lineage I/II according to a whole genome analysis [26]. An isolate from a super-shedding animal in the study conducted by Jeon et al. (2013) [8] was analyzed by whole genome sequencing, and the result showed it also belonged to lineage I/II. According to Munns et al. (2015) 99.2% and 0.8% of all isolates from super-shedders and low-shedders, respectively, were from lineage I and lineage I/II [45]. However, a broader study showed that 30, 35, and 37 out of 102 strains isolated from super-shedders were identified as
lineage I, II and I/II isolates, respectively [38]. These results indicate that super-shedders isolates do not belong to specific lineages.

The translocated intimin receptor (tir) single-nucleotide polymorphism (SNP) is another typing method used to differentiate *E. coli* O157:H7 [46]. The translocated intimin receptor, an effector of the type three secretion system, was previously used as a marker of virulence because of its significant role in adhering to intestinal epithelial cells. In this typing method, a nucleotide change (either A or T) at position 255 in the *tir* gene was used to predict the virulence of an *E. coli* O157:H7 strain, with hyper-virulent isolates having an increased chance to possess the T allele at this position [46]. Among all the super-shedder strains used in a study by Arthur *et al.* (2013) 71% harbored the T allele, while 29% of them harbor the A allele [38]. Munns *et al.* (2015) reported that in 10 strains isolated from cattle of super-shedder and low-shedder isolates, 9 strains (90%) displayed the T allele at the 255 position of *tir* gene [45].

The Shiga toxin-encoding bacteriophage insertion (SBI) assay is typically used to discriminate the *E. coli* O157:H7 shed from super- and low-shedders. Among a variety of Shiga toxin-encoding genes, *stx*1, *stx*2a and *stx*2c are the most frequently studied in STEC isolated from cattle. Since Shiga toxin-encoding genes may insert within or adjacent to several distinct chromosomal loci including *yehV, wrbA, argW* and *sbcB*. Six individual polymerase chain reactions (PCR) were previously employed to distinguish those insertion sites [47]. It was found in a study by Arthur *et al.* that super-shedding isolates have a higher chance of harboring either *stx*2a or *stx*2c than carrying both genes [38]. Within 12 groups of super-shedding isolates, 55.9% of isolates were identified as SBI genotype 1, 2 and 3 [38].
Sequencing techniques, including whole genome sequencing, are becoming increasingly popular to analyze the characteristics of bacteria. Historically, whole genome sequences of the *E. coli* O157:H7 strain isolated from the Sakai outbreak in Japan, was compared with that of the benign laboratory strain K-12. The result of this whole genome comparison has shown that the Sakai strain has 20 specific tRNAs and 131 potential virulence factors [48]. The whole genome sequence of another *E. coli* O157:H7 outbreak strain, EDL933, was also compared with that of K-12 strain, with 1387 specific genes identified in EDL933 [49]. Since whole genome sequencing of isolates could provide researchers with a variety of pertinent genetic information, researchers have begun to use this powerful tool in order to determine the differences between super-shedding and low-shedding strains at the genomic level. The whole genome sequences of 4 low-shedding isolates and 6 super-shedding isolates were sequenced in a study by Munns *et al.*(2015) [45] The results showed that, based on the analysis of single nucleotide polymorphism, there were no significant genetic differences between super- and low-shedding strains. Another study including the whole genome sequencing analysis of a super-shedding strain was conducted by Cote *et al.* (2015) [26]. Since phage regions of *E. coli* O157:H7 contain a variety of genes inserted by phages, it is possible that the genes which cause the super-shedding phenomenon are located in phage regions. Cote *et al.* (2015) compared the phage regions between EDL933, Sakai and EC4115 [26], but no specific genes were found in these phage regions. Researchers concluded that no specific gene in phage region was responsible for super-shedding characteristics. It is possible that super-shedding strains harbor specific adherence factors which give them ability to be shed in large quantities from the
intestinal tract. To examine this hypothesis, Cote et al. (2015) identified all the adherence factors in a super-shedding strain, SS17 [26]. Through comparison of adherence factors with those of reference strains, researchers found that no specific adherence factors could be associated with super-shedding. SNPs are also a potential factor that could be related to super-shedding bacteria. Recently, a new clustering method, IS629 profile, was proposed. It discriminates E. coli O157:H7 according to the presence of insertion sequences in chromosome. This new method may give researchers more clues for discovering the nature of super-shedding.

Animal Factors Contributing to Super-Shedding

A previous study showed that the prevalence of E. coli O157:H7 strains isolated from super-, medium- and low-shedders could be affected by genetic and physiological factors of animals [8]. Jeon et al. (2013) reported that Brahman cattle are more resistant to E. coli O157:H7 [8]. Some cattle probably have resistance to E. coli O157:H7, because they cannot be infected by inoculation with different doses of E. coli O157:H7, as was shown in previous studies [8, 50].

The microbiome of the GI tract of cattle were examined, since microbial community could potentially reflect host health. High variation in bacterial community composition was observed among feces of individual animals in recent study [51, 52]. By sequencing 16S rRNA genes of bacteria in feces from 11 super-shedders and 11 non-shedders, Xu et al. (2014) found variation of bacterial composition between two sets of animals, with super-shedders harboring a richer microbiome than non-shedders [53]. Some studies showed that changes in the microbiome may lead to a boost in immune response and exclusion of enteric pathogens [54]. As a result, super-shedding strains could be able to accumulate to higher numbers.
Environmental Factors Contributing to Super-Shedding

As previously mentioned, the environmental factors include season, diet, and animal housing hygiene. Shedding of STEC, especially *E. coli* O157:H7, varies significantly in different seasons, with a peak in summer [55-57]. Diet composition is another factor that significantly affects super-shedding. Callaway *et al.* (2011) found a decrease of *E. coli* O157:H7 population in ruminants after the animals were fed orange peel product [58]. In another study, Fox *et al.* (2007) observed that fecal shedding of *E. coli* O157:H7 increased after feeding distillers grains to cattle and claimed that the change of volatile fatty acids (VFA) concentration was responsible for this increase [59]. When Munns *et al.* (2015) made a change of feed for cattle, the super-shedders stop shedding *E. coli* O157:H7 at a high level [45].

**Conclusion**

Among all the factors that might cause super-shedding, bacterial factors are considered to be one of the major ones causing the super-shedding phenomenon in beef cattle. Although bacterial factors were among the most studied, the exact bacterial factors causing this phenomenon have not been clearly defined. We have also been unable to identify the common characteristics between super-shedding isolates. There are some possible reasons that could explain this. First, the super-shedding phenomenon is a transient phenomenon, which means that the status of super-shedders and low-shedders could change in a short period of time. As a result, it may not be easy to distinguish between super-shedders and low-shedders. In this case, the strains isolated from “super-shedder” and “low-shedder” may not be good representatives regarding this phenotype. Second, whole genome sequencing is thought to be a powerful technology to study bacterial factors attributed to super-
shedding. However, by using Illumina sequencing or other sequencing methods, many
gaps will be generated, which means some genetic information present in gaps are
missing. Without accurate sequencing data, the gene traits that cause super-shedding
will be more difficult. A more accurate sequencing technology, such as PacBio
sequencing, should be employed in order to study these bacterial factors. Third,
researchers tend to seek out a certain gene or SNP which is responsible for this
phenomenon. However, it is likely that this phenomenon is modulated by a cluster of
genes, which make this work more difficult.
CHAPTER 3
COMPARATIVE ANALYSIS REVEALS A HYPER-VIRULENT *Escherichia coli* O157:H7 STRAIN ISOLATED FROM A SUPER-SHEDDER

**Background**

Shiga toxin-producing *Escherichia coli* (STEC) are foodborne pathogens which threaten public health on a global scale [10]. It causes a variety of symptoms including bloody diarrhea and life-threatening hemolytic-uremic syndrome (HUS) [2-4]. STEC infections cause economic losses of more than $1 billion USD per year, including direct and indirect costs [5]. The primary reason that people get infected is through the consumption of food products contaminated by STEC. Since no effective treatment method is available for patients infected by STEC, contamination of cattle by STEC must be controlled at the pre-harvest level. Cattle are primary reservoirs of STEC so the prevention of colonization of STEC in cattle has become an important approach for preventing STEC infections in humans [30].

The terminal recto-anal junction (RAJ) as well as gastrointestinal (GI) tract, are the principle localization site by STEC in cattle [60]. Among the cattle colonized by STEC, a subset of cattle have been found to shed > 10^4 CFU of *E. coli* O157 /g feces and are defined as “super-shedders” [6]. The existence of super-shedders raises the chances of contamination of animal food products such as animal hides, beef and milk, and thus threatens public health. Jeon *et al.*(2013) [8] demonstrated that cattle were able to transmit super-shedding strains between farms. Previous reports showed that super-shedding animals within feedlot pens may be responsible for more than 90% of shedding of *E. coli* O157:H7 [1, 9]. In another study, Matthew *et al.* (2006) [6] suggested that 80% of bacterial transmission was caused by 20% of infectious individuals and reduction in shedding by the cattle will reduce human infections.
Three potential factors causing the phenomenon of super-shedding are bacterial, animal host, and environmental factors. To understand characteristics of super-shedding *E. coli* O157 strains, phenotypic and genotypic difference among super-shedder strains have been noted in previous studies, including phage type [1, 45], lineage-specific polymorphisms, *tir* polymorphisms [61], and the clade of *E. coli* O157 [62] and the difference in the presence of *stx*2a and *stx*2c [10]. Previous studies showed that most of the STEC O157 isolated from super-shedders in Scotland had a phage type of 21/28 (PT21/28) [1, 36]. The super-shedder isolates with phage types of PT4 and PT14a were predominantly identified in two studies conducted in the United State and Canada [10, 38]. Based on a lineage-specific polymorphism assay *E. coli* O157 isolates can be classified into three lineages: lineage I, lineage II and lineage I/II [42, 63]. A study showed that super-shedder isolates are distributed among all of the three lineages [38]. Another molecular marker, *tir* polymorphism, used to differentiate *E. coli* O157:H7 strains was based on a nonsynonymous base change (A or T) at position 255 in the *tir* gene [62]. Bono *et al.* (2007) [46] found that the *tir* 255 T allele was more frequently found in human isolates, suggesting the isolates with a T allele at the 255 position in *tir* may be more virulent to humans. Other studies showed that super-shedder isolates harbored both T and A alleles at the 255 position of *tir* and do not have an obvious preference [10, 38]. Clade analysis can discriminate *E. coli* O157:H7 isolates into 9 different clades based on SNPs in 96 loci [62]. Among all the 9 clades, isolates in clade 8 are believed to be hyper-virulent because they have higher chances to cause severe patient symptoms. Although the phenotypic and genetic methods have been conducted, the specific bacterial factors causing super-shedding is still unknown.
Recently, a whole genome analysis revealed that a super-shedder strain SS17 isolated from the Recto-anal junction (RAJ) of cattle in U.S. Midwestern states has a distinctive genome compared with foodborne outbreak strains, Sakai and EDL933, and spinach associated outbreak strains, TW14359 and EC4115 [26]. This study revealed that SS17 has a unique plasmid pSS17, a large number of non-synonymous SNPs (nsSNP) and other polymorphisms compared with the reference strains. In addition, the strong adherence aggregative phenotype associated with SS17 was shown to be unassociated with the locus for enterocyte effacement (LEE) [26].

Previously super-shedder strains were isolated from cattle shedding O157 at a level of more than $10^5$ CFU/g feces which is ten times the minimum load defining a super-shedder [1, 8]. These super-shedder strains shared the same PFGE profile and genotype of $eae^+ , stx1^- , stx2^+ \text{ and } hlyA^+$ [8]. Further studies showed that all these super-shedder strains can express Shiga-toxin. With this information, we hypothesize that the phenomenon of super-shedding is caused by a group of specific genomic traits of bacterial strains isolated from super-shedding cattle. The representative strain, KCJ1266, isolated from super-shedding cattle was sequenced using PacBio whole genome sequencing method, which could generate longer reads and circular bacterial whole genome sequence than Illumina sequencing, followed by whole genome comparative analysis.

This study shows that KCJ1266 can adhere to human HEp-2 cell as good as human pathogen EDL933, has similar biofilm formation ability of EDL933, belongs to $E. \text{coli}$ lineage I/II and clade 8, and shares a similar genomic architecture and virulence factors with SS17 and EC4115. Taken together, these results reveal that super-
shEDDING STEC STRAIN KCJ1266 is a hyper-virulent strain similar to that of SS17 and EC4115. This is one of the first studies which utilizes PacBio WGS to characterize a hyper-virulent STEC O157 strain isolated from super-shedding cattle. KCJ1266 can potentially be used as a reference strain for future studies regarding the phenomenon of super-shedding.
Materials and Methods

Bacterial Strains

KCJ1266 is a strain isolated from a super-shedding cattle located in North Florida in 2012. EDL933 (ATCC48935) is a laboratory strain and was linked to hamburger-borne outbreak in 1982 in the U.S. The laboratory strains, Δeae (KCJ696) and Δtir (KCJ698) and ΔfimH (KCJ867), are deletion mutant strains of EDL933.

Genomic DNA Extraction

Genomic DNA of KCJ1266 was isolated by the use of QIAGEN DNA mini Kit (cat. no. 51304). Briefly, 1 mL overnight culture of KCJ1266 was centrifuged for 10 min at 5000 x g (7500 rpm). Following the manufacturer’s protocol, the bacterial pellet was resuspended in buffer ATL and treated with 20 μL proteinase K at 56°C for 30 mins. The released DNAs were eluted with 100 μL of ddH₂O.

Stx2 Subtyping

Polymerase chain reaction (PCR) was performed to determine the stx2 subtypes (i.e. stx2a and stx2c gene) carried by KCJ1266. Two E. coli O157:H7 strains, EDL933 carrying stx2a and stx1 and FRIK2455 carrying stx2c, were used as controls. Primers used for detecting stx2a (stx2a-F: 5’-CTTTTCGCCACCAACAAAGTTATGT-3’ and stx2a-R: 5’-CACAGTCCCCAGTATCGCT-3’) and stx2c (stx2c-F: 5’-TACTGTGCTGTTACTGGGC-3’ and stx2c-R: 5’-ACAGTGCCCATATCGCC-3’) were designed according to previous study [44]. Each 25 μL PCR reaction mix contains 2.5 mL of 10X buffer, 0.5 mL of dNTP, 0.2 mL of Taq polymerase, and primers. The PCR reaction condition was set as 94°C for 5 min; 94°C for 30 sec, 54°C for 30 sec, 72°C for 60 sec for 30 cycles and a final extension time of 10 min at 72°C. Amplified PCR
product was analyzed in 1% agarose gel electrophoresis and visualized by ethidium bromide staining.

**Phage Induction**

Cell lysis caused by phage induction was measured by optical density (OD). Overnight culture of bacterial strains was inoculated into LB media. Mitomycin C was added to the cell culture at the final concentration of 0.5 μL/mL when the OD600 of the culture reached between 0.6 and 0.7. EDL933 was used as a positive control, while DH5α were used as a negative control for phage induction. For phage particle preparation, Mitomycin C was added to 25 mL cell culture when OD600 reached 0.7 to make a final concentration of 1 μL/mL. After 18 h incubation, unlysed cells and debris were removed by centrifugation at 4,000 rpm for 20 min at 4°C, and the resultant supernatant was filtered with 0.22 μm-pore-size membrane. Precipitation of phage particles was performed using 0.25 volume of 20% polyethylene glycol 8000 and 10% NaCl. The polyethylene glycol/NaCl solution containing phage particles was incubated at 4°C overnight, followed by centrifugation at 12,000 rpm for 1 h. The pellet was resuspended and stored in 1 mL SM buffer (0.58 g NaCl, 0.2 g MgSO₄·7H₂O, 0.01 g Gelatin per 100 mL 1M Tris-Cl pH7.5). SDS-PAGE was used to check the phage protein profile after an aliquot of the phage suspension was mixed with SDS-loading buffer and boiled for 5 min.

**Western Blot Analysis of Stx2 Expression**

The method used for identifying Stx2 expression was described in the previous paper [64]. Briefly, the exponential phase culture (OD600=0.7) of a O157 strain was treated with Mitomycin C to the final concentration of 1 μg/mL in order to induce the lytic cycle of phages in the O157 strain and concomitant stx gene expression. Next, the
O157 cell culture was incubated for 18 hrs for complete cell lysis. The cell debris was removed by centrifugation (4,000 rpm for 20 min at 4°C) and the cell-free supernatant was collected by using 0.22 μm-pore-size membrane filter. The total proteins in the supernatant were precipitated by adding 1 × volume of 100% trichloroacetic acid (Fisher Scientific BP555-1) to 4 × volume of the cell-free culture. The mixture was incubated for 30 min on ice and centrifuged at 14,000 rpm, 4°C for 30 min. The protein pellet was finally washed using cold acetone. The total proteins were separated by 12% sodium dodecyl sulfate polyacrylamide gel and transferred to a polyvinylidene difluoride membrane (0.45 μm pores, Immobilon-P, Millipore) for Western blot. The membrane was blocked using 5% skim milk in TBST (10 mM Tris-HCL [pH 7.4], 150 mM NaCl, and 0.05% Tween-20) at room temperature for 1 h. Later, the membrane was incubated overnight at 4°C with Monoclonal Verotoxin II-a subunit antibody (Meridian, Life Science Inc.), and washed using TBST at room temperature for 1 h. The membrane was then incubated with HRP conjugated secondary antibody (GE Healthcare) diluted 1:10000 in TBST, washed with TBST, and incubated with a chemiluminescent substrate (ECL Plus, GE Healthcare). Finally, the membrane was exposed to Kodak BioMax film (Carestream Kodak X-Omat LS film, F1274 Sigma).

**Adherence to HEp-2 Cells**

The ability of KCJ1266 to adhere to human epithelial type-2 (HEp-2) cells was evaluated. EDL933 was used as a positive control and eae and tir mutant of EDL933 was used as a negative control. HEp-2 cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) (Coring, 10-017-CV) composed of 10% (vol/vol) heat-inactivated fetal bovine serum at 37°C and 5% CO2. Approximately 10^5 of HEp-2 cells were seeded into a 24-well polystyrene plate as measured using a hemocytometer and
allowed to grow until 90% confluent (approximately 24 h). Bacterial cultures were grown overnight at 37°C in LB broth and washed with sterile phosphate buffered saline (PBS) three times. The final cell pellet was resuspended in PBS to a final concentration of O.D.600=1 (5x10⁸ CFU). An aliquot (20µl) of the resuspended culture was inoculated into 480µL of DMEM. Then 500 µl DMEM containing 10⁷ bacteria was added to each well (MOI 100). After a three-hour incubation at 37°C with 5% CO₂, the bacterial suspension was aspirated and 500 µl of new DMEM was added to each well, and this process was repeated once more. The infected cell monolayer was washed using sterile PBS three times to remove any unattached bacteria. To detach the HEp-2 cells from the plate, 1mL of 0.1% Triton X 100 in PBS buffer was added to each well and allowed to incubate for 5 min. The media was removed from each well and serially diluted in LB broth and plated on TSA plates to enumerate the number of attached bacterial cells. The adherence assay of each strain was conducted in triplicate. Statistical significance between the level of attachments of isolates to the negative and positive controls were accessed using a student’s t-test (α = 0.05).

**Static Biofilm Formation Assay**

The method used in the biofilm formation assay was the same as previously described one [26]. Briefly, EDL933 strain and ΔfimH mutant of EDL933 were used as a positive and a negative control, respectively. Strains were inoculated in LB media at 37°C with overnight shaking. The overnight culture was diluted (1:100) into M9 medium supplemented with glucose (0.8% wt/vol) and minerals (1.16 mM MgSO₄, 2 µM FeCl₃, 8 µM CaCl₂ and 16 µM MnCl₂) and incubated at 37 °C with shaking for 48 h. Then the overnight cultures were diluted (1:100) into the same medium and 200 µL of diluted cell
cultures were pipetted into a 96-well plate (Corning, NY, USA) in 6 replicates. The cell cultures were incubated at 30°C for 48 h, removed the supernatant and washed 3 times using PBS. The plate was dried at 37°C for 20 min until there were no water in the well, followed by staining the biofilm using 0.1% crystal violet (wt/vol) for 2 min. Then the crystal violet was removed and wells were washed using PBS for 3 times to remove the remaining crystal violet. The plate was dried at 37°C for 20 min to remove all the liquid in the wells, followed by adding 200 µl of 80% (vol/vol) ethanol and 20% (vol/vol) acetone to release the stain. The released stain was measured by microplate reader using the absorbance at λ 595 nm.

**Genome Sequencing and Assembly**

To acquire the whole genome sequence of KCJ1266, whole genome PacBio sequencing was performed in this study. Eighty-five thousand reads with a mean subread length of 7660 bases were obtained with 115.7 fold coverage. To obtain contigs of genome and plasmid, *de novo* assembly was conducted by Falcon ([https://github.com/PacificBiosciences/FALCON](https://github.com/PacificBiosciences/FALCON)). RS HGAP Assembly.3 protocol was used with control filtering set as ‘KeepControlReads’.1.xml. Option details were set as shown below: Minimum subread length was 500. Minimum polymerase read quality was set as 0.80. Minimum polymerase read length was 100. At this analysis, the minimum seed read length was adjusted to 2000. Target coverage, overlapper error rate, overlapper min length and overlapper K-Mer were set as 25, 0.06, 40 and 14. Contigs were connected to form a circular DNA by overlapping each end of contigs.

**Annotation and Comparative Genomics**

The genome of KCJ1266 was automatically annotated by RAST (Rapid Annotation using Subsystem Technology). A FASTA file of KCJ1266 was unloaded to
RAST (http://rast.nmpdr.org/). Mauve 2.3.1 (default setting) was employed to align the genome of KCJ1266 and reference strains. The SNPs between KCJ1266 and reference strains was identified using galaxy in HiPergator of the University of Florida.

To identify the position of phage related genes, the PHAge Search Tool (PHAST) (http://phast.wishartlab.com/) was used. After assembly was done and the contig of genome of KCJ1266 was available, FASTA file of the genome was submitted to the PHAST web server. The genome of was circularly presented by CGview [8]. Genes encoding known and putative virulence factors in the genome of KCJ1266 were found using VFDB [65-68] in Patric [69-71]. Multilocus sequence typing (MLST) of KCJ1266 and reference strains was conducted using Center for Genomic of Epidemiology (CGE).

Reference genomes used for comparative analysis were E. coli O157 strain EDL933 [NC_008957] and EC4115 [NC_011353], and super-shedding strain SS17 [NZ_CP008805].

**In silico Analysis of LSPA-6 Profile and Phylogenetic Clade**

Six pairs of primers described in Yang, et al. were used to perform BLAST against whole genome sequence of KCJ1266 using SnapGen View. LSPA-6 profile of KCJ1266 was decided based on the length of a region targeted by each pair of the primers. An Isolate that has a LSPA-6 profile of 111111 was defined as lineage I. A strain possessing 222222 or 222223 was defined as lineage II. Other genotypes were defined as lineage I/II [44].

Two methods were used to identify the clade of KCJ1266. An *In silico* clade 8 rhsA assay, the presence of a SNP (3468C) in the rhsA gene, was used to determine whether the isolate belongs to the Clade 8 [72]. Besides, another *In silico* analysis based on 32 SNPs was also conducted [62].
**In silico Analysis of IS629 Profile**

A total of thirty-two pairs of primers were used to detect IS629-distribution profile [44]. The primers were aligned against KCJ1266 using BLAST in NCBI. EC4115, EDL933 and SS17 were used as control. Positive was assigned if the *in silico* PCR amplicon has the same size as expected, while negative if the *in silico* PCR amplicon has either no IS629 or different size of amplicon [44].

**Function-Based Gene Comparison**

Whole genome sequence of KCJ1266, SS17, EC4115 and EDL933 were submitted to RAST (http://rast.nmpdr.org/). Then subsystem information of each strain was available for function-based analysis. The RAST supply the function of function-based comparison of two strains and output the unique genes that only exist in one of the strain.

**Phylogenetic Analysis**

The whole genome sequences of 28 *E. coli* O157 reference strains for building phylogenetic tree were downloaded from Patric (https://www.patricbrc.org/portal/portal/patric/Home). Mauve was used to align all the 29 sequences, including sequences of KCJ1266 and the reference strains, using HiPerGator, a super computer located in University of Florida. The output file was converted to a fasta alignment using an in house perl script, followed by extracting the variable sites within contig position of each strain using MEGA6. The MEGA6 output fasta file containing variable sites was submitted to IQtree webserver to construct a maximum likelihood phylogeny performing 1000 bootstrap iterations for branch support with the Kimura-2 nucleotide substitution model. Figtree (v.1.4.2) was used to edit the phylogenetic tree.
Results

Shiga Toxin Expression of KCJ1266

Super-shedder isolate of O157 strain, KCJ1266, was collected from cattle in North Florida Research and Education Center (NFREC) [8]. Fecal samples from rectal anal junction were collected using sterile cotton swab as previously described [8]. Since all the strains isolated from super-shedder, medium-shedder and low-shedder have stx2 gene and same PFGE pattern, KCJ1266 was chosen as a representative strain for further analysis [8]. PCR, Mitomycin C treatment, SDS-PAGE and Western blot were conducted to subtype stx2 gene and identify expression of Shiga toxin (Figure 3-1). The result shows that KCJ1266 contains both stx2a and stx2c genes and express Stx2 at a similar level of EDL933.

Adherence to Epithelial Cell Line

The adherence ability of STEC to epithelial cells plays a significant role in bacterial colonization and persistence in cattle. Since SS17 has been shown to have high levels of adherence ability to Hep-2 cells [26], we hypothesized that the high adherence ability may be a conserved genetic property in supper-shedder strains. To test the hypothesis, we compared the adherent ability of KCJ1266 to Hep-2 cells with that of control strains including EDL933, Δeae, Δtir and DH5α strains. The data showed that KCJ1266 has similar adherence ability to Hep-2 cells compared with EDL933 (Figure 3-2), while the KCJ1266 shows significant difference (P < 0.05) in adherence to Hep-2 cells compared with Δeae, Δtir and DH5α strains. This result reveals that KCJ1266 have a high adherence ability to the epithelial cells.
Biofilm Formation

A better bacterial biofilm formation ability is related to a better survival ability in the environment. In the biofilm assay, the biofilm formation of EDL933 shows significant higher than its fimH gene mutant (Figure 3-3.). At the same time, the KCJ1266 showed significant higher ability of biofilm formation on polystyrene surface. This result suggests that KCJ1266 may have a strong ability to survive in the environment, contaminate farm products and reinoculate other cattle.

Lineage Determination Based on LSPA-6

Previous studies show that O157 could be classified into three groups, lineage I, lineage II and lineage I/II, which of each is characterized by distinct genome structure, host origin, and virulence potential [73, 74]. Lineage I and I/II strains are usually isolated from clinical and bovine source, while Lineage II isolates usually originate from bovine source [44, 73]. In particular, Lineage I/II include Spinach outbreak strain that displayed high virulence and Stx expression [75]. Previously, the super-shedder strain SS17 was shown to belong to lineage I/II [26]. The lineage of KCJ1266 determined by LSPA-6 [43] showed a profile of (211111) which is the same as that of SS17 (Figure 3-4.). This result reveals that the two independent super shedder strains, SS17 and KCJ1266 belong to lineage I/II, implying their potential epidemiological significance.

Clade of KCJ1266

Manning et al. identified the presence of a hyper-virulence clade (clade 8) among E. coli O157 strains, based on an SNP analysis of 32 loci [62]. Clade 8 strains were characterized by high expression of Stx and high cytotoxicity [62]. In silico analysis shows that KCJ1266 belongs to clade 8 (Figure 3-4). Previously, SS17 was also shown to belongs to clade 8, indicating that KCJ1266 and SS17 are closely related. The
findings that KCJ1266 and SS17 belong to the same hyper-virulence clade and share super-shedding property, together with their distinct geological origins imply that this type of O157 strain characterized by super-shedder (SS) and hyper-virulence (HV) may be more prevalent in the natural habitat than was previously considered.

**IS629 Profile Distribution**

To further investigate the genetic relatedness between KCJ1266, SS17, and the prototypic strain EC4115, we performed *IS629* distribution profile (IDP) analysis, a method that can be used for not only lineage typing but also genetic differentiation within a given lineage [44].

IDP of KCJ1266 was determined by *In silico* analysis, and compared with IDP of SS17 EC4115 (lineage I/II control) and EDL933 (lineage I control). As expected, IDP of KCJ1266, SS17 and EC4115 displayed typical IDP of lineage I/II and were distinctive from lineage I type IDP represented by EDL933 (Figure 3-4). Interestingly, IDPs among the 3 lineage I/II strains were not identical but varied, reflecting their genetic divergence at the sub-strain level. Five IDP loci are polymorphic between KCJ1266 and SS17, all of which locate in prophage regions (Figure 2), indicating prophage polymorphisms are a primary attribute for sub-strain diversification in *E. coli* O157.

**Overview of KCJ1266 Genome**

In order to further detail the genetic traits of KCJ1266 its whole genome sequence was determined using PacBio sequencing technology. The KCJ1266 has a chromosome of 5478683 bp which encodes for 5545 open reading frames, 106 tRNA and 22 rRNA (Fig.3-5). The KCJ1266 genome contains a plasmid, pO157, of 95910 bp. This plasmid contains 117 genes, including genes translate the type V secretion serine protease (*espP*), hemolysin (*hlyABCD*), cytotoxin (*toxB*), catalase peroxidase (*katP*) and
the type III secreted proteins (etp operon). The overall genomic features of KCJ1266 are summarized in comparison to reference strains; SS17, EC4115, and EDL933 in Table 1. The size of KCJ1266 genome is smaller than other two lineage I/II strains, SS17 and EC4115. It contains a 95.91 Kbp plasmid which is homologous to the pO157 plasmid of SS17 and pO157 plasmid of EC4115, but lacks the additional plasmid commonly possessed by SS17 and EC4115. These data indicate that KCJ1266 contains reduced genomic features compared to its close relatives.

**Comparative Genomic Analyses.**

LSPA-6 typing, clade analysis and IDP revealed that all three lineage I/II strains investigated in this study, KCJ1266, SS17, and EC4115, commonly belong to hyper-virulence clade but have been diverged genetically. Further, genetic makeup of KCJ1266 and SS17 seem to be divergent each other despite they shared super-shedding phenotype. To understand the genetic diversity of the lineage I/II strains in the context of HV and SS properties, we conducted a series of comparative genomic analysis of KCJ1266, SS17, and EC4115.

The homology of KCJ1266 and reference strains was showed by the use of progressive Mauve, displaying 7 different synteny blocks (Figure 3-6). Aligned against genome of SS17, EC4115 and EDL933, the genome of KCJ1266 has a higher homology to SS17 and EC4115 than to EDL933. The non-homology region between KCJ1266 and reference strains were identified to be primarily phage regions.

In details, the lengths of the 7 synteny blocks (SB) varied from ~45.4 kb to ~2,276 kb (Fig.3-6). SB1 is a ~1,249 kb sequence containing phage region 1 to 4. SB2, a sequence of ~429 kb, SB3 with a length of ~ 45.4 kb contains part of phage region 8 (Table 3-3). SB4 of ~338.5 kb is inverted in EDL933. This block contains sections of
phage region 8 and 9. SB5 is a ~1,077.8 kb sequence that contains phage regions 10 to 14 and a section of phage region 9. SB6 with a ~62 kb length contains part of phage region 15. In the whole genome of EDL933, this block was moved, inverted and inserted between homologs of SB1 and SB2 of KCJ1266. SB7 contains phage regions of 16 to 19 and part of phage region 15 where a non-homology area was found. According to the analysis, most of the non-homology areas between KCJ1266 and SS17 exist in phage regions, suggesting that KCJ1266 and SS17 share highly conserved core genome region. However, when the genome of KCJ1266 was compared to EC4115 and EDL933, non-homology regions were frequently found in the core genome regions.

The plasmid of KCJ1266 was also compared with plasmids of reference strains. There are 99% query coverage and 99% identity of pO157 plasmids between KCJ1266 and SS17 (Figure 3-7).

**Functional Characterization of the Genome**

KCJ1266 have 5545 coding sequences (CDS). Fifty-five percent (3082) of the total CDS, including 2914 non-hypothetical and 168 hypothetical CDS, were categorized into 27 main subsystem categories, including a total of 599 subsystems categories (Fig. 5). The other 44% (2463) of the total CDS were not included in these 599 subsystem categories (Figure. 3-8).

Pair-wise comparisons of the functional genomes between KCJ1266 and the 3 reference strains were conducted based on the subsystem categories in order to find unique gene entities in each genome (Table 3-2). In comparison between KCJ1266 and SS17, KCJ1266 had three unique genes encoding nitrogen regulation protein NR (I), DNA adenine methyltransferase (phage-associated) and KefF (glutathione-regulated potassium-efflux system ancillary protein), while SS17 have one unique gene encoding
type III secretion protein SsaH. When KCJ1266 and EC4115 were compared with each other, KCJ1266 had three unique genes encoding Nitrogen regulation protein NR (I), DNA adenine methyltransferase and LSU ribosomal protein L32p, while EC4115 has one unique gene encoding SsaH protein. Seventeen unique genes present in genome of KCJ1266, when KCJ1266 were compared with EDL933. Beside the five genes shown in Table 2, 12 phage encoding genes encoding phage tails and proteins were among the KCJ1266 genes absent in the genome of EDL933.

All together, a majority of genes assigned to the subsystem categories in the genome of KCJ1266 are also present in SS17 and EC4115, except a few variable genes listed in Table 3-2. Genetic or physiological impacts of these minor differences, if any, are not known.

**Single Nucleotide Polymorphisms**

By comparing coding sequences of KCJ1266 and the reference strains, SNPs of KCJ1266 were identified. KCJ1266 has 109 and 196 SNPs compared with SS17 and EDL933, respectively.

Of the 109 SNPs between SS17 and KCJ1266, 22 are non-synonymous single nucleotide polymorphisms (nsSNP). ORFs containing the nsSNPs encode putative functions including tetratricopeptide repeat protein, Alpha-ketoglutarate permease, and PTS system glucitol/sorbitol-specific transporter subunit IIA.

**Virulence and Colonization Factors**

Since KCJ1266 belongs to the hyper-virulence clade 8 and displays supper-shedder property we analyzed genes known to be involved in virulence and colonization of this bacterium. A total of 252 genes related to virulence and colonization were identified in KCJ1266. Of these, 232 genes are encoded by chromosome and 20 are
encoded by pO157 plasmid. According to their putative protein function, these genes are divided into five groups, adhesion, effectors and toxins, secretion systems and other virulence-associated genes.

In chromosome of KCJ1266, there are 85 genes in adhesion group, including fim operon, encoding fimbrial proteins and csg operon encoding curli related proteins. Other genes, such as fliC, efa-1 and ompA, which are not in operons are also found.

The effectors contain LEE-encoded effectors and non-LEE-encoded effectors. The LEE-encoded effectors are EspA, EspB, EspF, EspG, EspH, EspZ, Tir, and mitochondrial associated type III secreted effector protein encoded by map. Non-LEE-encoded protein contains effectors encoded by nleA, nleC and nleD. The stx2 that encodes A and B subunits of Shiga-Like-toxin 2 are found in chromosome of KCJ1266. KCJ1266 contains both stx2a and stx2c genes, which is the same as reference strains, SS17 and EC4115.

The secretion system group is consisted of genes that encode type three secretion system (TTSS). Among these genes, TTSS components are encoded by sepL, escD and escQ. The needle structure of TTSS is encoded by escC.

The other virulence-associated genes contain iron uptake genes, such as chuA, chuS, chuT, chuU, chuW, chuX, chuY and iroB. These genes may play critical roles in iron acquisition that will affect the outcome of the infection significantly in bacteria-host interactions

Virulence factors of reference strains were also analyzed. The result shows that KCJ1266, EC4115 and SS17 have almost the same virulence factors (Figure 3-9). All of
KCJ1266, SS17, EC4115 share similar composition of virulence genes, suggesting that KCJ1266 may have the same virulence as EDL933, EC4115 and SS17.

**Phage Regions in KCJ1266**

A typical *E. coli* O157:H7 genome contains multiple lambdoid prophages which are known to be a major contributor for the genomic diversity among strains. Consistently, the IDP profile and MAUVE analyses in this study revealed that prophage-related polymorphisms are a main contributor for genetic dynamics detected in KCJ1266 and its relative strains. In addition, *stx* genes and a multitude of putative virulence genes are carried by prophages. The integration sites for *stx2a* and *stx2c* in the phage regions are shown to be *argW* and *sbcB*, respectively, in most of lineage I/II and are conserved in KCJ1266. A total number of 19 phage regions were found across the genome of KCJ1266 and major genes encoded in each phage region are summarized in Table 3-3 and Figure 3-10. The completeness of phage regions are described as intact, incomplete and questionable based on the scoring of each phage region, which is decided by the number of phage related coding sequences and the homology of suspected regions to the sequences of phage regions in the database [76]. Most of these genes in phage regions of KCJ1266 were also present in phage regions of SS17.

Previous study showed that Stx2a-encoding phage could be classified into four major and two minor subtypes [25]. *In silico* analysis showed that KCJ1266 belongs to Stx2a phage subtype of ϕStx2a_γ that can produce high level of Shiga toxin, suggesting the hyper-virulence property of KCJ1266.
**Phylogenetic Analysis**

To understand the evolutionary position of KCJ1266, the phylogenetic analysis was performed using whole genome sequences of 28 *E. coli* O157 strains, including strains of human source, bovine source and outbreaks (i.e., Spinach outbreak and Taco bell outbreak). Figure 3-11 shows that these strains were clustered into three major groups which are congruent to the lineage I, lineage II, and lineage I/II. As expected, KCJ1266, SS17, EC4115, other spinach outbreak strains, and Taco bell outbreak strains formed a discrete lineage I/II group while EDL933 and Sakai strains belonged to the lineage I group, and selected stains of bovine source form a lineage II group. Lineage I/II group contained two distinct clusters; cluster 1 (10 spinach outbreak strains) and cluster 2 (2 Taco bell outbreak strains). KCJ1266, SS17, another super shedder SS52, EC4115 and two other spinach outbreak strains (EC4024 and EC4192) were shown to be outgroups without forming a distinct cluster (Figure 3-11). KCJ1266 is closest to EC4192. Eppinger *et al.* previously classified Spinach outbreak strains based on SNP genotypes [73]. In this classification, all three EC4115, EC4024, and EC4192 strains were genotyped as Group B while stains in the inner Spinach outbreak cluster were genotyped as either Group A or Bovine genotype II. All lineage I/II strains included in this study have clade 8 genotype. All together, these data support that KCJ1266 are closely related to the phylogenetically distinctive outgroups of lineage I/II, yet possessing high virulence potentials.

**Discussion**

To discover the genetic traits that contribute to super-shedding, whole genome sequencing and a variety of gene markers have been used in previous studies. The Illumina sequencing method has been identified as one of the most popular methods...
used for whole genome sequencing, however, gaps are generated within whole genome sequence. At the same time, a large amount of Illumina sequencing data is not able to be closed. In the current study PacBio sequencing was employed in order to generate a gap free whole genome sequence to better characterize KCJ1266.

Whole genome sequencing data of KCJ1266 was compared with other reference strains at a genomic level (Table.3-1), proving KCJ1266 a super-shedder-hyper-virulence strain by analyses of phage regions, virulence factors, LSPA-6, IS629 profile, function-based gene, whole genome, SNPs and phylogeny.

In this study we identified several critical features of E. coli O157:H7 KCJ1266 isolated from a super-sheding cow. KCJ1266, SS17 and EDL933 strains belong to clade 8 and have the same LSPA-6, revealing they share the same phenomenon of hyper-virulence. Second, IS629 typing shows the additional genetic variability among these three closely related strains. This relationship and diversity are further detailed by genomic characterization of KCJ1266 and comparative analysis with SS17 and EC4115. In addition, virulence factors of KCJ1266 are highly similar with other reference strains, suggesting that KCJ1266 is a hyper-virulent strain.

It is commonly believed that the phenomenon of super-shedding is caused by three major factors which are bacterial, host and environmental factors [1, 8]. Among all these factors, bacterial factors are studied the most based on phenotypic and genetic characteristics. Manning et al. (2015) showed that clade 8 E. coli O157:H7 strains were more likely to cause hemolytic uremic syndrome in humans after testing more than 500 E. coli O157:H7 strains [62]. Another study showed that E. coli O157:H7 strains of clade 8 isolated from cattle may partially be responsible for the high amount of HUS cases in
Argentina [77]. In this study, KCJ1266 was identified as a clade 8 strain belonging to lineage I/II, suggesting that KCJ1266 may have hyper-virulence characteristics. This result is consistent with the comparison of virulence related genes among KCJ1266 and reference strains. Except one phage related gene, all the virulence related genes in hyper-virulence reference strain (EDL933) were found to be present in KCJ1266. Ogura et al. tested STEC O157:H7 isolates harboring different stx2 genes, finding that strains harboring stx2a or both stx2a and stx2c could express higher levels of Stx2 than those strains harboring only stx2c [25]. This finding is also consistent with the hyper-virulence properties of KCJ1266. LSPA profile of KCJ1266 (211111) is the same as SS17 and EC4115, which is consistent with phylogenetic analysis, suggesting that they are phylogenetically related to each other. Although KCJ1266, SS17 and EC4115 are phylogenetically related to each other, diversity in their IS629 profiles were observed. The difference among inserted sequence based profiles may be representative of the diversity of these three strains.

Its super-shedding and virulence gene characteristics are major factors that identify KCJ1266 as a hyper-virulent strain. Since KCJ1266 was isolated from cattle shedding STEC O157:H7 at a level of > 10^4 CFU/g feces, there is a greater chance for this isolate to contaminate the environment and farm-related products. In addition, the enhanced ability of KCJ1266 to produce biofilms indicates that KCJ1266 has enhanced environmental fitness. The ability to produce biofilms means this strain may survive in animal food products after food processing. In addition, the enhanced biofilm formation ability of KCJ1266 suggests that it may have high affinity in cattle, which means KCJ1266 may have enhanced within-farm and between-farm transmission.
The whole genome sequence of a low-shedder *E. coli* O157:H7 would have been an additional reference to help identify the genetic traits that contribute to super-shedding. However, there is no such whole genome sequence available. The reference strains, SS17, EC4115 and EDL933, were chosen because they have similar genome architecture and good quality. This may explain why no such specific genes responsible for super-shedding were identified. In addition, the whole genome analysis at the genomic level may not reflect the differences between super-shedder strains and low-shedder strains. The existence of a gene in two different isolates does not necessarily mean the same level of transcription and expression of this gene. Even in the same isolate, the transcription and expression level of a gene may be affected or regulated by other stimulus. It should be noted that a comparison of transcriptome and proteomics of super-shedder isolates and low-shedder isolates may help understand the essence of the phenomenon of super-shedding.

Although this research focuses on the bacterial factors related to super-shedding, animal host factors are also related to super-shedding. A previous study showed that the prevalence of *E. coli* O157:H7 strains isolated from super-, medium- and low-shedders could be affected by genetic and physiological factors of animals [8]. Jeon et al. (2013) reported that Brahman cattle are more resistant to *E. coli* O157:H7 [8]. Some cattle probably be resistant to *E. coli* O157:H7, because they cannot be infected by inoculation with different doses of *E. coli* O157:H7 [50, 78]. In addition, the innate and adaptive immune response of the host are believed to be the critical host factors [10]. Although cattle have no receptor for Shiga toxin [79], its innate immune response can be induced by colonization of *E. coli* in the GI tract [10]. In addition, dietary animal
husbandry practices play important roles in super-shedding. Diet composition, feed processing and bactericidal properties of ruminal and volatile fatty acids could impact *E. coli* shedding and surviving.

**Conclusion**

This study characterized a hyper-virulent STEC O157:H7 strain, KCJ1266, isolated from a super-shedder. Although the genes related to super-shedding phenotype are not identified in this study, the identification of the accurate, gap-free, and circular whole genome sequence of KCJ1266 is a cornerstone for the optimization of outbreak source tracking. In addition, KCJ1266 will serve as an ideal reference for the future study of the super-shedding phenomenon.

There are several possible reasons explaining the result that none of the genes related to super-shedding were identified in this experiment. The experimental setting was not perfect for identifying super-shedding phenotype. Instead of comparing stains used in this experiment, it will be better to compare the whole genomes of isolates from super-shedders and low-shedders. In addition, the expression of proteins is not only related to presence of genes, but also regulated by transcription factors. In this case, the transcriptome and proteome of strains isolated from super- and low-shedders should be compared. Furthermore, animal factors and environmental factors are also critical. There is a probability that animal factors are modulating super-shedding phenotype.
Table 3-1. Genome statistics of KCJ1266 and reference O157 strains

<table>
<thead>
<tr>
<th></th>
<th>KCJ1266</th>
<th>SS17</th>
<th>EC4115</th>
<th>EDL933</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Sequence (Mb)</td>
<td>5.48</td>
<td>5.52</td>
<td>5.57</td>
<td>5.53</td>
</tr>
<tr>
<td>G+C ratio (%)</td>
<td>50.5</td>
<td>50.5</td>
<td>50.5</td>
<td>50.4</td>
</tr>
<tr>
<td>Coding DNA Sequence (CDSs)</td>
<td>5274</td>
<td>5430</td>
<td>5608</td>
<td>5587</td>
</tr>
<tr>
<td>No. rRNA</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>No. tRNA</td>
<td>106</td>
<td>107</td>
<td>109</td>
<td>100</td>
</tr>
<tr>
<td>No. Plasmid</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Protein</td>
<td>KCJ1266</td>
<td>SS17</td>
<td>EC4115</td>
<td>EDL933</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------</td>
<td>------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>NR (I)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Enzyme</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>KefF</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SsaH</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>L32p</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>RTX</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>YopD</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>tRNA dihydrouridine synthase A</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Phage Regions</td>
<td>Size (Kbp)</td>
<td>Completeness</td>
<td>Synteny block</td>
<td>Encoded genes</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
<td>26.5</td>
<td>Intact</td>
<td>SB1</td>
<td>Phage DNA invertase; Phage integrase; and phage assembly protein Lysozyme and Attachment invasion locus protein precursor</td>
</tr>
<tr>
<td>2</td>
<td>38.5</td>
<td>Intact</td>
<td>SB1</td>
<td>Attachment invasion locus protein precursor</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>Incomplete</td>
<td>SB1</td>
<td>Attachment invasion locus protein precursor and 3 tRNAs</td>
</tr>
<tr>
<td>4</td>
<td>29.4</td>
<td>Intact</td>
<td>SB1</td>
<td>Attachment invasion locus protein precursor and phage lysis</td>
</tr>
<tr>
<td>5</td>
<td>50.5</td>
<td>Intact</td>
<td>SB2</td>
<td>Integrase and phage assembly proteins</td>
</tr>
<tr>
<td>6</td>
<td>46.5</td>
<td>Intact</td>
<td>SB2</td>
<td>Integrase, phage lysis, attachment invasion locus protein precursor</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>Questionable</td>
<td>SB2</td>
<td>Mobile element proteins</td>
</tr>
<tr>
<td>8</td>
<td>57</td>
<td>Intact</td>
<td>SB2 and SB3</td>
<td>Attachment invasion locus protein precursor, 2 tRNAs, phage lysis, acfC, and integrase</td>
</tr>
<tr>
<td>9</td>
<td>94.8</td>
<td>Intact</td>
<td>SB4 and SB5</td>
<td>Attachment invasion locus protein precursor, phage lysis 5 tRNAs and integrase</td>
</tr>
<tr>
<td>10</td>
<td>24.8</td>
<td>Incomplete</td>
<td>SB5</td>
<td>Integrase and 3 tRNAs</td>
</tr>
<tr>
<td>Phage Regions</td>
<td>Size (Kbp)</td>
<td>Completeness</td>
<td>Synteny block</td>
<td>Encoded genes</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>--------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>11</td>
<td>54</td>
<td>Intact</td>
<td>SB5</td>
<td>Integrate, phage lysin, <em>inv</em> and 5 tRNAs Attachment invasion locus protein precursor, integrate, <em>yeeV</em>, <em>yeeX</em>, <em>stx2A</em>, <em>stx2Bc</em> and 3 tRNAs</td>
</tr>
<tr>
<td>12</td>
<td>61.2</td>
<td>Intact</td>
<td>SB5</td>
<td>Phage assembly proteins and Phage lysin Integrate and attachment invasion locus protein precursor</td>
</tr>
<tr>
<td>13</td>
<td>24</td>
<td>Intact</td>
<td>SB5</td>
<td>Stx2 converting 1717 NC 011895</td>
</tr>
<tr>
<td>14</td>
<td>73.5</td>
<td>Intact</td>
<td>SB5</td>
<td>Stx2 converting 1 NC 003525</td>
</tr>
<tr>
<td>15</td>
<td>77.2</td>
<td>Questionable</td>
<td>SB5, SB6 and SB7</td>
<td>Phage lysin and integrase</td>
</tr>
<tr>
<td>16</td>
<td>23.9</td>
<td>Questionable</td>
<td>SB7</td>
<td>Phage lysin and integrase</td>
</tr>
<tr>
<td>17</td>
<td>13.1</td>
<td>Questionable</td>
<td>SB7</td>
<td>Stx2 converting 1717 NC 011357</td>
</tr>
<tr>
<td>18</td>
<td>20.5</td>
<td>Questionable</td>
<td>SB7</td>
<td>Stx2 converting 1717 NC 011357</td>
</tr>
<tr>
<td>19</td>
<td>21.7</td>
<td>Questionable</td>
<td>SB7</td>
<td>Entero mEp460 NC 019716</td>
</tr>
</tbody>
</table>
Figure 3-1. Shiga-toxin II detection. A. Phage induction was measured. MMC was added to cell culture of EDL933, KCJ1266 and DH5α at the 3rd, 3rd and 4th hour, respectively (Red arrows). B. PCR of stx2a and stx2c. C. SDS-PAGE analysis of phage proteins. D. Western blot.
Figure 3-2. Adhesion of *E. coli* O157:H7 isolates to Hep-2 cells, MOI: 100. The Assay was performed in triplicate.
Figure 3-3. Biofilm assay of KCJ1266. EDL933 was used as a positive control and $\Delta$fimH mutant of EDL933 was used as a negative control.
Figure 3-4. *IS629* profile of KCJ1266, SS17, EC4115 and EDL933. A shaded gray box indicates the presence of the *in silico* amplicon with correct size and locus, while an unshaded white box indicates a negative *in silico* result.

The three reference strains (EDL933, EC4115 and SS17) are compared.
Figure 3-5. Genome map of KCJ1266. Marked characteristics are shown from outside to the center. KCJ1266: CDS on forward strand, tRNA and rRNA in forward strand, tRNA and rRNA in reverse strand, CDS on reverse strand, GC content and GC skew; pO157: CDS on forward strand, tRNA and rRNA in forward strand, tRNA and rRNA in reverse strand, CDS on reverse strand, GC content and GC skew.
Figure 3-6. Mauve alignments of KCJ1266 with the three reference genome. Alignment of KCJ1266 with the three reference strains (SS17, EC4115 and EDL933) reveal 7 blocks of homology each of varying from ~45.4 kb to ~2,276 kb. Each block is a different color with lines connecting corresponding homologous blocks, with the white regions indicating non-homology. (* indicates the block that only exists in SS17 and EDL933)
Figure 3-7. Comparison of O157 plasmids of KCJ1266 and other 3 reference strains
Figure 3-8. Subsystem category distribution. The subsystem contains 56% of total genes. The other 46% of genes don’t belong to any of the subsystem.
Figure 3-9. Phage comparison of KCJ1266 with reference strains. From inner ring: phage regions of EC4115, phage regions of SS17, phage regions of KCJ1266. The same color represent homologous regions.
Figure 3-10. Virulence related genes in KCJ1266 and reference strains. A. Known genes detected in KCJ1266 B. Putative genes, hypothetical genes, mobile genes and phage related genes detected in KCJ1266
Figure 3-11. Strains are clustered into three groups which belong to Lineage I (green), Lineage II (red) and Lineage I/II (blue). Within Lineage I/II, there are three subgroups which are cluster 1, cluster 2 and outgroup. KCJ1266, EC4115 and SS17 belong to outgroup of Lineage I/II.
LIST OF REFERENCES


BIOGRAPHICAL SKETCH

Lin Teng was born in Fuzhou, Fujian, China in 1990. He enrolled at the Huazhong Agricultural University in Wuhan, China, one of the most foremost Agricultural Universities in China in 2008. His B.S. project focused on constructing deletion mutant strains of extraintestinal pathogenic *Escherichia coli* and evaluating the change of toxicity of mutant strains. Lin earned his Bachelor of Agronomy degree at Huazhong Agricultural University in 2012. After graduation from the university, Lin served in Fujian Fengze Group Cooperation, a feed and meat producer, for one year as a research assistant. In 2014, he was accepted by the Department of Animal Science of University of Florida as a M.S. student and joined Dr. KwangCheol Casey Jeong’s lab. During his M.S. program, he got scientific training and participated in some other project, such as the projects of Chitosan nanoparticles and antimicrobial resistance. His long-term plan is to work in the industry in China. His immediate plan is to learn more bioinformatics techniques in the current lab and finish his other ongoing projects.