GENOMIC AND MOLECULAR ANALYSIS OF THE UNIQUE PHENOTYPES
OF AN E. COLI O157:H7 SUPER SHEDDER ISOLATE

A Thesis in

Pathobiology

by

Matthew Raymond Moreau

©2013 Matthew Raymond Moreau

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2013
The thesis of Matthew R. Moreau was reviewed and approved* by the following:

Vivek Kapur
Professor of Veterinary and Biomedical Science
Thesis Advisor

Subhashinie Kariyawasam
Assistant Professor; Veterinary Microbiologist

Edward G. Dudley
Associate Professor of Food Science

Anthony Schmitt
Professor of Veterinary and Biomedical Sciences
Head of the Graduate Program in Pathobiology

*Signatures are on file in the Graduate School
ABSTRACT

Shiga toxin producing Escherichia coli (STEC) are a subtype of pathogenic E. coli, and recently STEC serotype O157:H7 (EHEC O157:H7) have been recognized as a major foodborne pathogen causing symptoms ranging from having simple intestinal discomfort to bloody diarrhea and life threatening hemolytic uremic syndrome. Asymptomatic cattle are colonized with EHEC O157:H7 at the mucosal interface of the recto-anal junction (RAJ), and coincidentally this colonization of the RAJ is critical for the ability of this organism to be shed from the cattle. Super shedding (SS) is a phenomenon that has been reported in some cattle that shed EHEC O157:H7 at a rate greater than 10^4 colony forming units per gram of feces (CFU/g), 100-1000 times more or greater than normal shedders. The critical nature of colonization of the RAJ and shedding of EHEC O157:H7 indicated that this interaction might have been changed at a cellular level. The development of a unique cattle RAJ cell culture model revealed that O157:H7 employ a LEE-independent mechanism of attachment to the other RAJ cell type, recto-anal squamous epithelial (RSE) cells. SS isolates show a strong aggregative and adherence phenotype on RAJ cells, as opposed to two other O157:H7 strains, 86-24 and EDL933, which is also independent of the LEE operon. Other data show that these SS isolates have a higher affinity of binding leafy greens, such as lettuce and spinach, which is important because contaminated leafy greens serve as another major route of transmission of this pathogen. Since biofilms have been implicated in colonization and bacterial survival in a variety of environments, we investigated if SS17 could form biofilms, and the data indicate that SS17 has an enhanced ability to form biofilms.
compared to non-pathogenic *E. coli* strain K12 as well as EDL933 (a strain not believed to be an SS isolate). Recently, a representative SS isolate, SS17, was fully sequenced, and its genome was analyzed for changes in the genome (deletions and/or acquisition of novel genes) as well as single nucleotide polymorphisms (SNPs), focusing on virulence and adherence related genes. These analyses revealed that SS17 contained 295 virulence related genes, of which 39 had coding non-synonymous SNPs (nsSNPs). Of the 9 genes that were chosen through a rigorous selection process, none had a complete conservation of the same nsSNP in all 11 other SS isolates but were used to generate a phylogenetic analysis of the representative SS isolates. None of the genes chosen were shown to be individually essential for the change in adherence phenotype to the RSE cells seen in SS17; however, *ompA*, *eivA*, *yfaL*, and *wzzB* gene deletions in EDL933 were able to switch the phenotype in EDL933 from moderate aggregative to strong aggregative attachment seen in SS17. These data indicate that they may play a role in this phenotype but in an inhibitory manner and that a protein(s) they interact with may be involved directly. In SS17 it has now been shown that *yfaL* is critically important for the production of biofilm in context of leafy green extract; as is *cah* and *ompA*; but it is unclear if the nsSNPs exhibited by the two former genes have to do with the enhanced biofilm formation exhibited by SS17. All three show importance to the enhanced biofilm formation of SS17, and *yfaL* and *ompA* have for the first time been shown to be critical to biofilm formation in a strain of EHEC O157:H7; and all 3 important for SS17s ability to produce biofilms induced by factors released by damaged leaves.
TABLE OF CONTENTS

LIST OF FIGURES ..................................................................................................... vii
LIST OF TABLES ....................................................................................................... ix
ACKNOWLEDGEMENTS ......................................................................................... x

Chapter 1  Introduction ............................................................................................ 1

Chapter 2  Background and Significance ................................................................ 6
   2.1 Commensal and Pathogenic E. coli ................................................................. 6
   2.2 STEC and E. coli O157:H7 ........................................................... 7
      2.2.1 Mechanisms of Virulence and Virulence Factors ....... 8
      2.2.2 Other Virulence-Associated Genes of EHEC O157:H7.................. 11
   2.3 STEC Colonization of Cattle and Life cycle .............................................. 12
      2.3.1 Colonization of the Recto-Anal Junction ........................................ 13
   2.4 Super Shedding EHEC O157:H7 .............................................................. 14
      2.4.1 Prevalence and Impact of Super Shedding ................................ 15
      2.4.2 Factors Associated with Super Shedding .............................. 16
      2.4.3 Sequence Characterization of a Representative SS Isolate-SS17 .... 20
   2.5 SS17 Displays LEE-Ind. Strong, Aggregative Adherence on RSE Cells...... 21
   2.6 SS17 Virulence-Related Genes and Relevant nsSNPs .............................. 24
      2.6.1 Toxins ......................................................................................... 24
      2.6.2 Adherence Genes ........................................................................ 27
      2.6.3 Virulence Associated Genes ....................................................... 30
   2.7 Different Mechanisms of Adherence of EHEC O157:H7 ......................... 32

Chapter 3  Materials and Methods ....................................................................... 36
   3.1 Bacterial Strains and Plasmids ................................................................. 36
   3.2 Gene Selection Process ............................................................................ 39
   3.3 Recombineering to Generate Knockouts in SS17 ....................................... 40
      3.3.1 Generating Kanamycin-Resistant Cassettes for Recombineering .... 41
      3.3.2 Making Electrocompetent Cells and Induction of pKD119. .......... 43
      3.3.3 Transformation, Electroporation and Screening .......................... 43
   3.4 RSE Cell Adherence Assay .................................................................... 45
   3.5 Lettuce/Spinach Adherence Assay ........................................................... 46
   3.6 Biofilm Assays ....................................................................................... 47
   3.7 Analysis of nsSNPs Across the SS Isolates ............................................... 48
   3.8 Growth Curves and Doubling Times ....................................................... 49

Chapter 4  Results and Discussion ..................................................................... 51
4.1 Strong Aggregative Phenotype on RSE is Shared Among SS-Isolates .......... 51
4.2 SS17 Displays Enhanced binding to Leafy Greens................................. 53
4.3 SS17 Displays Enhanced Biofilm Production in Damaged Spinach.......... 54
4.4 Rationale and Background of the Genes of Interest.............................. 56
  4.4.1 Genes Involved in Adherence and Biofilm formation ...................... 56
  4.4.2 Genes Involved in Virulence and Adherence .................................. 60
  4.4.3 Genes Involved in Virulence and Survival/Dual Function................... 62
4.5 Recombineering and Generation of Knockouts...................................... 64
4.6 Comparative Genomic Analysis on Genes of Interest Across SS Isolates.... 65
4.7 Characterization of the SS17 Mutants’ Phenotypes ............................... 73
  4.7.1 Role of Genes of Interest in RSE Adherence Assay ......................... 73
  4.7.2 Biofilm Production Induced by Damaged Leafy Greens in SS17
Mutants..................................................................................................... 78

Chapter 5 Conclusions and Future Directions ............................................. 86

Appendix A: Virulence-Related Genes in SS17............................................. 103
  A-1: Virulence Genes in the SS17 Genome .............................................. 103
  A-2: Adherence Related Genes in SS17................................................. 108
  A-3: Virulence Associated Genes in SS17.............................................. 113

Appendix B: Primer Tables ........................................................................... 116
  Appendix B-1: Recombineering Primers.................................................... 116
  Appendix B-2: Screening Primers ............................................................. 117
  Appendix B-3: nsSNP Sequencing Primers .............................................. 118

Appendix C: Genetic Analyses of SS strains............................................... 119
  Appendix C-1: PFGE Analysis of SS Isolates............................................ 119

References................................................................................................. 120
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>LEE-Mediated Pedestal Formation with Human Epithelia</td>
<td>10</td>
</tr>
<tr>
<td>2-2</td>
<td>Makeup and colonization of EHEC O157:H7 at the RAJ</td>
<td>13</td>
</tr>
<tr>
<td>2-3</td>
<td>Factors Contributing to Super Shedding</td>
<td>17</td>
</tr>
<tr>
<td>2-4</td>
<td>The Potential Role of QS in Colonization of O157:H7 at the RAJ</td>
<td>18</td>
</tr>
<tr>
<td>2-5</td>
<td>RSE Adherence Phenotype of SS17 Compared to 2 Reference Strains</td>
<td>22</td>
</tr>
<tr>
<td>2-6</td>
<td>RSE Adherence Phenotype is LEE-Independent</td>
<td>23</td>
</tr>
<tr>
<td>2-7</td>
<td>The General Lifecycle of a Biofilm</td>
<td>34</td>
</tr>
<tr>
<td>3-1</td>
<td>Prioritization of Genes Targeted for Knockout</td>
<td>40</td>
</tr>
<tr>
<td>3-2</td>
<td>Schematic of generating KanR Cassette</td>
<td>41</td>
</tr>
<tr>
<td>3-3</td>
<td>Schematic of Primer annealing for Colony PCR Screen</td>
<td>44</td>
</tr>
<tr>
<td>4-1</td>
<td>Shared RSE Adherence Phenotype Amongst SS Isolates</td>
<td>52</td>
</tr>
<tr>
<td>4-2</td>
<td>SS17 and Other SS Isolates Show Enhanced Adherence to Baby Spinach</td>
<td>54</td>
</tr>
<tr>
<td>4-3</td>
<td>SS17 Displays Enhanced Biofilm Formation from Leafy Green Extract</td>
<td>55</td>
</tr>
<tr>
<td>4-4</td>
<td>Results of Screening for ΔGOI strains</td>
<td>65</td>
</tr>
<tr>
<td>4-5</td>
<td>Phylogenetic Analysis of SS isolates by nsSNPs in Genes of Interest</td>
<td>72</td>
</tr>
<tr>
<td>4-6</td>
<td>RSE Adherence of the SS17 Mutants</td>
<td>76</td>
</tr>
<tr>
<td>4-7</td>
<td>RSE Adherence of the EDL933 Mutants</td>
<td>77</td>
</tr>
<tr>
<td>4-8</td>
<td>SS17ΔompA, Δcah, ΔyfaL Show Decrease in Biofilm Formation</td>
<td>82</td>
</tr>
<tr>
<td>4-9</td>
<td>SS17 Deletion Mutants Show no Difference in Growth in LB</td>
<td>84</td>
</tr>
<tr>
<td>4-10</td>
<td>Doubling Times of SS17, Mutants, EDL933, and MG1655</td>
<td>85</td>
</tr>
</tbody>
</table>
Figure C-1: PFGE Analysis and Clustering of SS Isolates

119
LIST OF TABLES

Table 2-1: Toxins with nsSNPs ................................................................. 27
Table 2-2: Adherence Related Genes with nsSNPs ..................................... 29
Table 2-3: Virulence Associated Genes with nsSNPs .................................... 32
Table 3-1: Parent Strains and Isolates Used in This Study ......................... 38
Table 3-2: Plasmids Used in This Study ..................................................... 38
Table 3-3: Mutants Generated for this Study ............................................. 39
Table 4-1: Conservation of Mutations in Genes Selected for Knockout ....... 70
Table A-1: Toxin Genes ............................................................................ 103
Table A-2: Adherence Related Genes ....................................................... 108
Table A-3: Virulence Associated Genes .................................................... 113
Table B-1: Recombineering Primers ........................................................ 116
Table B-2: KanR Screening Primers ......................................................... 117
Table B-3: nsSNP Sequencing Primers ...................................................... 118
ACKNOWLEDGEMENTS

I would like to thank Dr. Vivek Kapur, my advisor and mentor. He believed in me and helped guide me through this tough endeavor and helped me grow as a person and as a scientist. His patience and wisdom helped sculpt this project and me to be ready to move on to the next step. I would also like to thank my committee members. I thank Dr. Subhashinie Kariyawasam for her guidance and her advice on this project. She also helped provide me with the plasmid we used to create the gene knockouts. To Dr. Ed Dudley whose insightful conversations and editing on the thesis coupled with words of wisdom helped me to be a better scientist and a better communicator of my science. I would like to thank those who helped me with ideas for the project and the work in the lab that made it all possible. A special thanks to my lab mates, Rebecca Cote, Robab Katani, and Lingling Li who helped in many ways with this project such as providing me with the sequence and annotation of SS17, helping with experiments and troubleshooting protocols. I would like to also thank with great appreciation Dr. Maria Bandl, Dr. Indira Kudva and Dr. Michelle Carter all of the USDA for their significant contributions to this work and without whom the work could not have been completed. I would like to thank Dr. Michael Mwangi and his group for providing us with the protocols for the doubling times as well as help with identifying genetic changes in SS17. Finally I would like to thank my friends and family for all of their support through all of these years and all of
the ups and downs I have come across. A special thanks to Laura Goodfield, Yury Ivanov, Sarah Muse, and Jihye Park for their insightful conversations, edits on my work, and preparation of different presentations on this work.
Chapter 1

Introduction

Shiga-toxin producing *E. coli* (STEC) is just one of 6 major pathotypes that have evolved in *E. coli* but is responsible for approximately 265,000 infections in the US alone every year [1,2,3]. One serotype, O157:H7 (which is an Enterohemorrhagic *E. coli* or EHEC) is responsible for almost 40% of those infections. The symptoms of these infections can range from simple abdominal discomfort and watery diarrhea to renal damage and the potentially fatal symptoms of hemolytic uremic syndrome (HUS) [2,3,4]. The major routes of EHEC O157:H7 infections are through the consumption of leafy greens that have been contaminated (such as spinach and lettuce), contaminated undercooked beef and contact with infected animal hides (via animal to person) [2,5,6].

With all of these different mechanisms of transmission and the significant threat to human health EHEC O157:H7 poses, it would be most beneficial to find an efficient way to stymie this cycle. However, despite many efforts, the procedures and protocols that exist attempting to prevent this organism from spreading have not worked well, yielding many multi-state outbreaks, which can result in significant illness and even death.

EHEC O157:H7’s primary reservoir is a part of terminal rectum of asymptomatic cattle at a unique region of the gastrointestinal (GI) tract known as the recto-anal junction (RAJ) [7,8]. The RAJ is a section of the GI tract that is the junction of two cell types: the stratified epithelia (RSE) and follicle-associated epithelial cells (FAE) [7]. Colonization of this site is important for the shedding of O157:H7 into the environment. Normally the
shed rate of this pathogen from cows is approximately 10-100 colony forming units per gram (CFU/g) [8,9].

Recently, some cattle have been identified as ‘super shedders’ where the *E. coli* O157:H7 have been found to be shed at greater than $10^4$ CFU/g [8,10,11,12]. The only factor that has been identified to date as being important for this phenotype is the aforementioned colonization at the RAJ [9,10]. Given the threat that this organism poses, this significant increase in shedding bacteria leads to increases of the EHEC O157:H7 in the environment and thus increases the propensity to cause disease [13]. It is believed that three critical factors are involved in this phenotype, the environmental factors of the host and microbe, the host factors, and the microbial factors [8]. Environmental factors for the host include the diet and living conditions; for the microbe this would include facets such as the microenvironment it colonizes. Host factors include the host immune response and even microbiota as they make up a significant portion of the host. Microbial factors include genes involved in shedding and colonization, for example. Aside from the RAJ colonization, no other factor(s) have been successfully identified as being critical for this phenotype.

Because of the importance of the colonization at the RAJ and its association with transmission and shedding EHEC O157:H7, we hypothesized there would be a change not only *in vivo* but also in the cellular/molecular interaction of the bacteria with the cells of the RAJ that would be shared amongst the SS isolates. Previous data presented by Kudva and colleagues as well as Naylor and colleagues indicated that both SSE (RSE) and FAE cells could facilitate binding of EHEC O157:H7 respectively, and the interaction with the RSE cells was through a LEE-independent mechanism [14,15].
Recently, work from this lab and collaborators confirmed that SS17 (a representative SS isolate), did in fact have a distinctly different adherence pattern to RSE cells which was also independent of LEE, leading to our assumption that this change in interaction may be important for the SS phenotype.

Many representative SS strains were PFGE (Pulse-Field Gel Electrophoresis) and LSPA (Lineage specific polymorphism assay) typed to determine the extent of genetic heterogeneity or homogeneity between the SS isolates and previously described EHEC O157:H7 strains. The LSPA and PFGE assays group genetically-alike strains based on their evolution and acquirement of changes in the genome within certain genes and restriction sites. For example, SS17 was shown to be related to other EHEC O157 strains which were involved in outbreaks (Lineage I/II outbreak strains), whereas SS12 is lineage 1 outbreak strains, and SS77 is genetically similar to cattle isolates via LSPA. However, these all cluster distinctly in the PFGE analysis of these SS isolates (Figure C-1). Because outbreaks happen to occur through leafy green contamination (as previously mentioned), we hypothesized that these SS isolates would have good adherence on leafy greens.

Biofilm formation has been implicated in being important for long-standing infection and survival in a variety of hosts and surfaces for many bacteria. Biofilms are also important for a range of functions of bacterial communities such as signaling, antibiotic resistance and generating a second layer of adhesion to different surfaces. There are many genes important for the formation of biofilms encoded in EHEC O157:H7, as well as many adhesins that are involved in biofilm formation, structure and stability. Our collaborators, Michelle Carter and Maria Brandl recently discovered that
EHEC O157:H7 produces more biofilm when compared to K12 laboratory strain. If the first two hypotheses are correct, we would further hypothesize that SS strains also produce enhanced biofilm, and because of the difference in shedding and potentially interactions (adherence) with the RSE cells and leafy greens, they may produce a different amount of biofilm than a non-SS EHEC O157:H7.

From studies by Arthur and colleagues, it was determined that the SS isolates do not share any gross genetic commonalities and the factors involved in the SS phenotype would probably be on the molecular level (e.g. conserved nsSNPs) [9]. To start to identify the nsSNPs and other genetic changes that could potentially be involved, we fully sequenced and characterized a representative SS isolate: SS17. Analyzing the genome and performing comparative genomics, coupled with the transcriptome and proteome provided by Carter/Brandl and Kudva respectively, for proteins involved in virulence and/or adherence, we will be able to identify candidate genes of interest which can be targeted for deletion and verify their role in the phenotype in both the RSE and leafy green adherence assay [14].

Sequencing candidate genes and nsSNPs across multiple SS strains may reveal either a conservation of factors involved in the phenotype across the SS isolates or by cluster/lineage, thereby helping narrow down the list of candidate genes. We hypothesize that these studies will help us begin to understand the molecular mechanism and factors involved in these interactions, as well as the changes that can occur to these factors to generate these observed changes. The ultimate goal of this project is to identify these genes which are involved in any changes in unique molecular phenotypes associated with SS isolates. The long range goal is striving to develop strategies to prevent SS phenotype
of SS isolates in cattle either by preventing colonization and/or the SS phenotype and transmission.
Chapter 2

Background and Significance

2.1 Commensal and Pathogenic E. coli

*Escherichia coli* are gram negative, rod shaped, facultative anaerobic bacteria that can either be commensal microbiota of organisms, or pathogenic [16]. Commensal *E. coli* colonize the gastrointestinal (GI) tract of many organisms, and it is believed that they can play roles in preventing illness through niche filling in the intestine as well as be involved in generating nutrients that humans require [17]. Pathogenic *E. coli* can cause a wide variety of infections in humans which are both intestinal and extra-intestinal related [1,16]. Meningitis, bladder infections, intestinal discomfort, watery diarrhea, and potentially fatal hemolytic uremic syndrome (HUS) are all symptoms associated with human disease of different types of *E. coli* that can infect and survive in a variety of different organs of host organisms [1,3,18]. Horizontal gene transfer through phage transduction and transformation/conjugation of plasmids not only account for much of the diversity between many of the *E. coli* but also is the driving force behind the acquisition of virulence and survival genes allowing both commensal and pathogenic *E. coli* to survive in many different environments.

There are 6 major pathotypes of *E. coli* that cause intestinal infection in human, and they include enterotoxigenic (ETEC), enteropathogenic (EPEC), enteroaggregative (EAEC), enteroinvasive (EIEC), diffusely adherence (DAEC), and most notably STEC
All of these intestinal infections cause different types of diarrhea and other intestinal issues and are all categorized based on the site of infection and/or the interactions they have with host cells. Comparatively STEC strains, and in particular enterohemorrhagic *Escherichia coli* or EHEC, are particularly pathogenic [1,3,19]. One serotype of EHEC, O157:H7 (EHEC O157:H7) are the cause of significant morbidity in humans and are responsible for multi-state outbreaks related to undercooked beef or presence on leafy greens from farms with infected animals [2].

### 2.2 STEC and *E. coli* O157:H7

Of the approximately 265,000 enterohemorrhagic *E. coli* (EHEC) infections estimated in the U.S. alone each year, the EHEC O157:H7 serotype is responsible for just over 1/3 of these cases; making it the most common STEC [2]. The O157:H7 serogroup is serologically identified by its specific O-antigen and flagellar antigens respectively. Another unique identifier of this bacterium is their inability to ferment sorbitol [8,18,20]. There is a breadth of difference within the O157:H7 populations, all belonging to different lineages (based on the Lineage specific polymorphism LSPA) and phage type among others genetic-linkage markers [8]. These help not only identify the EHEC O157:H7 strain but also group them to relatedness as there are different characteristics of each when observing attributes such as pathogenicity and associations with outbreaks. The common route of infection of humans with EHEC O157:H7 is either through undercooked beef of contaminated meat and animal hides or through leafy greens and other vegetation that can be contaminated from shedding cattle that are infected [2].
STEC gets its name from the major toxin it carries that damages host cells with direct cytotoxicity: Shiga toxin [21]. Shiga toxin was acquired through horizontal gene transfer from a lambdoid bacteriophage, which carried the toxin from a *Shigella dysenteriae* species and integrated it into the STEC chromosome [21]. This toxin, also known as verocytotoxin, is the distinguishing feature between STEC and the other 5 intestinal pathotypes of pathogenic *E. coli*. Shiga toxin (*Stx*) comes in two main gene products; Stx1 (classical *Shigella dysenteriae* toxin) and Stx2 based on their relatedness to the gene originally from *Shigella dysenteriae* [21,22,23].

### 2.2.1 Mechanisms of Virulence and Virulence Factors

In humans, the primary site of colonization and infection (bacteria-cell interaction) is in the large intestine, not at the terminal rectum as in cattle [1,3]. STEC O157:H7 in humans, unlike cattle, can cause severe illness with a range of symptoms. The most common symptoms are intestinal discomfort, watery or bloody diarrhea, vomiting, nausea, and stomach cramps [1,2,3]. This will last for about a week, but even humans, after becoming asymptomatic, can become carriers for this pathogen and shed it for up to several months after the symptoms subside [2]. Of all STEC O157:H7 infections, 10% of all those infected will progress to the most dangerous of the symptoms posed by this pathogen: hemolytic uremic syndrome, or HUS. HUS is more prevalent in children and the elderly, more associated with *stx*2, and is characterized by the destruction of the vasculature of certain organs, most notably the kidneys [1,2,4,5,22].
This leads to a rate of about 50% of people who develop HUS will have permanent renal damage [18].

Shiga toxin is unique because it prevents the use of antibiotic therapy in treating this infection, as DNA damage or activation of the SOS response (caused by the antibiotic or even neutrophil release of superoxides), will cause the phage to begin its lytic cycle [24]. This is an intriguing evolutionary processes because STEC do not have a proper secretion system for Shiga toxin and so the only way to release it is through viral-mediated lysis of the bacteria (and they are late genes so they will be expressed in lytic phage expression [21]. Stx2 is more associated with HUS and hemorrhagic colitis than stx1, but they both can cause significant death to any cells bearing the globotriaosylceramides (Gb3) receptor (including endothelial cells and neutrophils), as they are similar in structure of the toxin and mechanisms of action [1,18,21,24]. Shiga toxin is an A/B toxin with a pentamer B subunit non-covalently bound to the active A subunit, which causes an inhibition of protein synthesis in host cells yielding host cell death and tissue necrosis [1,3,18]. There is also some evidence that Shiga toxin which is internalized through a non-Gb3 mechanism may play a role in down-regulating secretion of immune system signaling such as chemokines [1,25].

Though the Shiga toxin is essential for mediating cytotoxicity of the host cells and necrosis of the associated tissue, it is not the only mechanism deployed to infect humans. The Locus of Enterocyte Effacement (LEE) is a mosaic pathogenicity island which are essential to the intimate attachment of EHEC O157:H7 to human epithelial cells [1,3,26,27]. LEE 1-3 encodes for gene products that make up the Type III Secretion System (T3SS), whereas LEE 4 encodes for the proteins and effectors secreted by the
T3SS into host cells directly. These effectors proteins hijack different processes of the host cells and can polymerize and polarize actin filaments. This mechanism is completed by the association of Translocated Intimin Receptor (Tir) and will express Eae (Intimin) on its surface to bind Tir leading to effective attachment and the formation of the effacement lesion or pedestal formation, seen in Figure 2-1 [3,27].

Another major contributor to the pathogenesis and symptoms observed in people who are infected by EHEC O157:H7 are the dual-system of iron sequestration via the hemolysin toxins and the Chu- family of genes [1,28,29]. There are 4 genes which encode for the hemolysins that are found on the 92kb pO157 plasmid in all EHEC O157:H7 and are involved in a variety of processes such as hemolysis and biofilm formation [18,30,31,32].

This hemolysis and subsequent bleeding in the stool as well as destruction of the glomeruli of the kidneys is what defines this bacteria as enterohemorrhagic. Since red blood cells are full of iron-filled heme molecules, EHEC uses hemolysis to compete with the host for the bio-available iron (an essential nutrient) by secreting iron-chelating siderophores. However many siderophores produced by this pathogen can be sequestered
by human produced lipochalin and be competed against by human proteins such as lactoferrin [18,33]. The Chu- system is the set of genes deployed by EHEC O157:H7 to incorporate heme released by red blood cells destroyed by hemolysin in order to survive in an iron-depleted environment [28,29].

2.2.2 Other Virulence-Associated Genes of EHEC O157:H7

STEC O157:H7 have a number of identified genes that have either been shown to be potentially involved in virulence or implicated in survival in a multitude of hosts. For example, EHEC O157:H7 have been shown to not only encode for a LEE-regulated T3SS, but also a second non-LEE (nle-, and eiv- families of genes) regulated T3SS and its associated effectors to be used in it [34,35,36]. Not much is known about them, but some have been shown to affect immune signaling by corrupting cellular response to infection and play a role in regulation in trans of other effectors of pathogenicity, such as the LEE operon [36,37]. Most EHEC O157:H7 encode for a Type 2 Secretion System (T2SS) and a toxB gene, which are both implicated in adherence [31,38,39]. STEC O157:H7 also have many genes that are used for biofilm formation (such as the wca- locus and pO157 factors), resistance (such as gad operon for acid resistance and ter- for tellerium and phage resistance) and has three quorum sensing systems to respond to different types of canonical type I, II, and III autoinducers (that have a role in regulation of a plethora of genes including those for virulence) [3,40,41,42].
2.3 STEC Colonization of Cattle and Life cycle

The most common reservoir in the environment for EHEC O157:H7 is asymptomatic cattle, which harbor the bacteria in the recto-anal junction of the gastrointestinal tract (GI tract) and shed them into the environment through fecal matter [3,7,8,10]. Even though cattle are considered to be the principle reservoir for EHEC O157:H7; sheep, pigs and deer have also been implicated in harboring this pathogen. Cows are asymptomatic because they lack the endothelial Gb3 (globotriaosylceramide) receptor for Shiga toxin [43]. Once shed, the *E. coli* O157:H7 can then be transmitted to other animals and/or other cattle through infecting plants such as grass or getting into the ground water and infecting all the living things that come into contact with that water (e.g. watering the animals, or water used in irrigation systems to water plants) [19,44]. Humans can also be infected by EHEC O157:H7 in contaminated food and can also contribute to the transmission of this organism via fecal-oral route. All of this leads to a cycle and leads to persistence of the *E. coli* O157:H7 in the environment and leading to high amounts of incidences and infections over time [19].
2.3.1 Colonization of the Recto-Anal Junction

The primary site of colonization in cattle is the recto-anal junction (RAJ), which is primarily associated with the mucosal epithelium seen in Figure 2-2. The RAJ is the junction of two distinct cell types; the stratified squamous epithelia (SSE) and the lymphoid follicle associated columnar epithelia (FAE). This site of colonization is critical for the ability to shed this pathogen into the environment [7]. Studies have shown that the LEE operon, important for colonization of the GI tract of humans, is also important for colonization of sheep and cattle [15]. In an intimin negative strain, challenged sheep couldn’t be colonized by EHEC and defective colonization young calves that were orally challenged [15].

Kudva and colleagues recently established a novel cell line to better understand the interactions between EHEC O157:H7 and cattle RAJ epithelia and in particular, the recto-anal junction stratified squamous epithelia, or RSE [14]. By incubating bacterial cells with the RSE cells and use of fluorescence microscopy, different conditions could be used to start to identify critical molecules or proteins involved in the interaction of
EHEC O157:H7 and the RSE cells. In this study, it was shown that RSE cells and EHEC O157:H7 cells do interact in vitro, and the bacteria will bind to the cells in unique patterns depending on the strain (comparing isolate 86-24 and EDL933). Surprisingly, this study also showed that this interaction was through a LEE-independent mechanism of adherence [14]. LEE-mediated attachment might be for interactions with the FAE cells rather than the RSE cells (given that this is shown in vitro), but the RSE cells can also serve as a site of attachment and colonization in cattle [14,15]. This also gives a cell culture model to study this phenomenon at the molecular level.

2.4 Super Shedding EHEC O157:H7

The average load of EHEC O157:H7 transmitted to the environment through fecal shedding from cattle is approximately 10-100 colony forming units per gram (CFU/g) [9]. Recently, a phenomenon has been observed in cattle where the bacterial load is greater than $10^4$ CFU/g, designating these cattle as ‘super shedders’ (SS). This threshold ($>10^4$) is the point that separates a ‘normal’ shedder from a super shedder [8,9,10,11]. Despite a low incidence in the number of cattle that are identified as super shedders, their impact on the transmission and environmental load of EHEC O157:H7 is extremely high [11,13].

In a study conducted by Terrance Arthur of the USDA, 102 EHEC O157:H7 isolates were collected from these super shedder cows to try to identify distinguishing features which might help identify certain strains as super shedder isolates as well as uncover any common genetic traits that might be responsible for this process. Through
various methods including PFGE typing (Figure C-1), phage-typing (Figure C-1), lineage-specific polymorphism assay (LSPA), and the tir allele test, it was concluded that the isolates did not share a common genotype that could be associated with these super shedder isolates [9].

2.4.1 Prevalence and Impact of Super Shedding

There have been various studies on different farms around the world that have shown that only small populations of cattle are actually designated as ‘super shedders.’ In a study presented by Matthews and colleagues, showed that in Scottish farms 80% of all transmission events on farms surveyed were due to less than 10% of the SS cattle [11]. Similarly, another study showed that a similar percentage of cattle were responsible for 96% of the environmentally exposed EHEC O157:H7 in fecal matter [9]. Arthur and colleagues showed that 95% of feedlot pens that contained at least 1 super shedder had a prevalence rate of over 80% on the hides of their pen mates [8]. In fact, having a super shedder in a pen will increase the environmental load of EHEC O157:H7 as a whole [8,13].

The infectious dose for cattle is approximately 300 cells of EHEC O157:H7, so it is not hard to imagine that if one pen mate is excreting $10^4$ CFU/g (and probably producing many grams a day), that the higher shedding would lead to higher transmission rates to susceptible cattle [45]. One study revealed that genotype background does seem to play a role, and thus not all cattle that are exposed to this pathogen will automatically become colonized by it or allow it to shed [13]. These prevalence rates and impact on
exposure to both the hides of cattle and the environment could potentially be a huge risk to the health of consumers who eat produce or meat sold by these farms [8,9,19]. There is a great need to better understand this phenomenon and how to control it, as it will be an effective measure to ensure the safety of foods produced by these farms [8].

2.4.2 Factors Associated with Super Shedding.

For all that we know about the human and EHEC O157:H7 interactions, there is little we know about the interactions between *E. coli* and its host reservoir, the cattle, at the molecular level. Therefore, the factors that contribute to super shedding are not yet known, but are likely contributed by three main factors: The host, the microbe, and the environment, as seen in Figure 2-3 [8]. The three factors most likely contribute something to this phenotype because of the general nature of it such as the seasonal flow of the SS phenotype.
The first factor is the host. The genotype of a host, coupled with its response to an EHEC O157:H7 infection via its immune system among other things could contribute to this phenotype [10]. It is important to keep in mind that at this stage cattle are considered to be super shedders and the bacteria are super shedder isolates. For example, it was recently found that Brahman cows were more resistant to infection and colonization than were Angus-Black cows. The same study found that steers are less susceptible than bulls [13]. The host immune system and microbiota can also play a significant role on the host attributes to this phenotype, as they will play a role in the colonization and length of infection in that cattle. In fact, it has been found that infection with EHEC O157:H7 induces an IgA-mediated innate immune response in the gut in response to lesions [46]. The cells that are involved in this phenotype from the host standpoint are also important as they will determine the type of colonization at the RAJ (the FAE and RSE cells) [7].

The second factor is the environment. The environment is a two part factor being comprised of the environment of the host (the cattle) and the environment of the bacteria. The host environment consists of management practices, feed, as well as outside
environmental conditions. This makes sense since there seems to be a seasonal flow to the super shedder phenotype [9]. The microbial environment is made up of the nutrients it has access to, the regional microbiota of the cattle at both the site of colonization and upstream on the GI tract. Host microbiota play a critical role especially for the microbial environment [3, 47, 48]. With the observation of quorum sensing amongst bacteria, it is well known that bacteria can signal and alter the transcriptional program of themselves or even other bacteria. It has been hypothesized that this is one of the main contributing factors of the colonization at the RAJ; a high autoinducer I (AI-1) signal at the early gut from the colonized microbiota and a weakening of that signal down through the lower gut into the RAJ. AI-1 works through the SdiA, quorum sensing regulator, and is known to upregulate expression of the flagellar genes for swarming and motility and turning down genes like those in the \textit{LEE} operon for intimate adherence [40, 49]. The \textit{Gad-} operon responsible for survival in acidic environments is upregulated by the \textit{SdiA} signaling pathway, as seen in Figure 2-4. Signals from other bacteria as well as those from a chemotaxis standpoint all influence the expression of genes from EHEC O157:H7 as it colonizes and more than likely sheds from cattle [3, 50].

The last factors that may play a role in this phenotype are the microbial factors, which are probably the factor to study and most easily

\textbf{Figure 2-4: The Potential Role of QS in Colonization of O157:H7 at the RAJ}

![Figure 2-4: Acyl-homoserine lactones produced by microbiota of the gut will upregulate the Gad- operon allowing for the survival through the upper intestine and as the concentration of the AHLs decreases as the EHEC swim toward the terminal rectum, the Fli- genes were be turned down, LEE turned up, Gad- turned down, allowing for attachment and colonization at the RAJ. Figure is from Nguyen, Y and Vanessa Sperandio 2012.}
manipulate. The factors involved in this phenotype could be any number of genes, but more than likely that have to do with colonization, adherence, and maybe possibly virulence associated genes (such as biofilm formation). Due to the lack of experimental evidence on the molecular interactions between EHEC O157:H7 and cattle, it is very difficult to anticipate what factors may be involved in the super shedding phenotype. As previously mentioned, it is known that LEE is important for the colonization of the cattle, but is not essential as orally challenged calves deleted for different LEE genes are still able to be colonized and shed this pathogen. Other studies have shown factors such as stx2 expression, lineage and origin all can play a significant role in the ability to colonize cattle [20]. This suggests that there are microbial factors involved in at least the colonization of the RAJ, but that there are other factors involved that are independent of LEE that allow for attachment and colonization in the cattle that are still unknown. At a molecular level, the microbes that are shed will have to be detached or loosely associated with the RAJ cells to be shed, and thus microbial factors probably play an important role.

Despite this lack of knowledge on the microbial factors involved, we do have some clues as to what may be involved. The recent study that established a good cell culture model of the RAJ (RSE adherence assay previously discussed), Kudva and colleagues also published a top-down proteome of the EHEC O157:H7 EDL933 [14]. This proteome along with a sequenced SS isolate (SS17) provided the basis for which identification of microbial factors involved could be established. The RSE model will allow for the analysis of bacterial attachment to the RSE cells and thus give observations at the molecular interaction of EHEC O157:H7 (including SS isolates) and the RSE cells of the RAJ. Together with the fact that the RAJ colonization is critical for the ability of
this pathogen to be shed from the cattle host, will help identify the microbial factors involved in the SS phenotype in addition to the factors involved with its adherence [10,14].

### 2.4.3 Sequence Characterization of a Representative SS Isolate-SS17

Recently, one of the representative super shedder isolates SS17 was sequenced. SS17 has similar genome make up as other reference genomes, with similar size (5.5MB), number of coding regions (5442), GC (50.5%) content, and large plasmid pO157. SS17 also contains a smaller plasmid names pSS17 all shown in. Previous analysis from Terry Arthur along with the sequence of SS17 revealed that SS17 contains many of the prominent virulence factors of EHEC O157:H7 such as eae, tir, espB, stx2 and stx2c, and is negative for stx1 in the chromosome [9]. Chromosome had both LEE and non-LEE operons and family of genes. In the pO157 large plasmid it contains virulence factors toxB and hlyA-C.

Further phylogenetic analysis was performed on SS17 which helped show its relationships to other EHEC O157:H7. Lineage specific polymorphism assay (LSPA) classified SS17 as lineage I/II which is linked to many of the spinach outbreaks as well as clade 8 strains. These strains have been shown previously to have a stronger adherence to MAC-T cells (approximately 2-fold more), as well as increased expression of many of the virulence factors on both the chromosome and pO157. All things considered, there is not too much difference between SS17 and many other strains whom SS status has yet to
be determined, including isolates that are not of the same lineage or clade (such as EDL933).

2.5 SS17 Displays LEE-Ind. Strong, Aggregative Adherence on RSE Cells
Using the protocol provided by Kudva and colleagues, SS17 was analyzed for its adherence phenotype to RSE cells in the RSE adherence assay. SS17 showed a distinct aggregative and strong adherence phenotype when compared to EDL933 and 86-24 reference strains. This can be seen very clearly in Figure 2-5. SS17 shows a distinct phenotype deemed ‘strong adherence, strong aggregative’ phenotype, as opposed to the phenotype displayed by the 2 references EDL933 (moderate adherence, aggregative), and 86-24 (Strong adherence, diffuse aggregation). It is clear from this that the super shedder isolate 17 has a unique binding to the RSE cells, and that the increase in adherence and aggregation could play a critical role in the super shedding phenotype in the microbe.

Figure 2-5: RSE Adherence Phenotype of SS17 Compared to 2 Reference Strains:
The adherence phenotype of SS17, EDL933 and 86-24 in A. Bar graph representing the Percent mean of either RSE or HEp-2 cells with the EHEC O157:H7 labeled below. B. Panels showing the fluorescence microscopy of SS17, EDL933, and 86-24 with RSE cells depicting the differences in phenotype.
factor facet of factors contributing to that phenotype.

Previously it was reported that this interaction for reference strain EDL933 was independent of the LEE operon. Figure 2-6 shows that this phenotype (the strong aggregative phenotype), is also LEE-independent. Many adhesins that are encoded by EHEC O157:H7 (including SS17) are important for adherence to host surfaces and others responsible for mediating bacterial-bacterial attachment; but few adhesins have been shown to be critical for both processes. A LEE-independent mechanism of attachment has not been observed in eukaryotic cell-EHEC O157:H7 interactions until Kudva and colleagues showed it as previously mentioned, and this mechanism of attachment is also still unknown. It is possible that this phenotype has two potential factors involved, and adhesin important for binding to the RSE cells and another adhesin responsible for the enhanced aggregation we see; but this could also be the work of one or more adhesins that have cross-

activity in attachment and aggregation. I would expect to see either a change in genetic structure and/or protein structure of one of these adhesins (or their accessory

---

**Figure 2-6: RSE Adherence Phenotype is LEE-Independent**
RSE adherence assay of SS17 with and without LEE-antisera is shown. Panel A shows the fluorescence microscopy of the RSE adherence assay to confirm the retention of the unique phenotype observed in Figure 2-5. To ensure the LEE-antisera is functioning, it was tested in Panel B. As it shows (by mean percent of cells with <10< cells), also indicating no significant difference in the quantitative binding of SS17 in the presence of LEE antisera with RSE cells, but significant reduction of adherence with HEp-2 Cells.
proteins), or to see a change in regulation of the adhesins by a change in gene(s) that are involved in an adhesin’s regulation.

2.6 SS17 Virulence-Related Genes and Relevant nsSNPs

In the genome of SS17, 295 virulence related genes were identified in the chromosome and the pO157 plasmid, each containing 276 and 19 virulence genes respectively (See Appendix Table A1-A3) that were also found in previously described O157:H7 E. coli such as EDL933 and TW14359. These genes can be split up into three distinct groups: toxin related proteins, adherence related proteins, and virulence associated proteins. These groups have within them 110, 104, and 81 genes respectively, many of which serve multiple functions in the various virulence-related processes E. coli O157:H7 undergo while invading a host. These groups and some examples of these genes are described below. In all, there are 12 nsSNPs in 12 genes that are toxins or related to toxins. There are 26 nsSNPs in 19 adherence related genes, and 8 nsSNPs in 8 genes that are in genes considered to be associated with virulence.

2.6.1 Toxins.

Toxins are proteins and molecules that either damage host cells or corrupt normal cellular activities and functions, and SS17’s genome reveals many of the critical toxins as compared to the reference strains. Twelve of the 110 toxins and toxin related genes (Appendix A-1.1 ) that had been identified in the genome of SS17 bore nsSNPs (Table 2-
This list includes LEE and non-LEE regulated effectors as well as endotoxin and exotoxins. Because the mechanism of binding to RSE cells is LEE-independent and EDL933 does have the same general mechanism (LEE-independent), these probably don’t play much of a role in this process or explain the difference in the phenotype [26].

Non-LEE regulated type 3 secretion system (ETT2) is poorly studied (but does have links to mitigating immune responses of the host), and a couple of its genes have nsSNPs. The ETT2 and its effectors make for an interesting case for the ones that their functions are still unknown such as the ETT2 itself (such as eivA and eivF), and the esp-family of genes with differences [34,36]. These genes could potentially play a role in the SS phenotype (or at least colonization) because they disrupt the immune signaling which might allow them to get to the RAJ and colonize. The hemolysin genes may have a consequence with the nsSNPs but it probably does not have much to do with the adherence phenotypes observed on the leafy greens and the RSE cells. Any of these genes could play a significant role in vivo and could contribute the super shedding phenotype as a whole at the macro level and just not as far as the molecular attachment and colonization are concerned [20].

The LEE (Locus of Enterocyte Effacement) operon has been a well noted feature of EHEC O157:H7s that is critical for their ability to cause infection in human hosts [1]. This locus contains genes which are essential for the intimate cell-cell interaction O157 has with eukaryotic cells through a series of cellular effectors which effect host cell homeostasis and proteins which mediate the expression, biogenesis and delivery (via Type III secretion system (T3SS)) of these effector proteins [27]. SS17 has 47 LEE encoded genes similar to that of previously described O157. Also in concordance with
reference strains, SS17 also encodes 35 non-LEE encoded effector proteins and their constituents (e.g. non-LEE encoded T3SS). The few non-LEE factors that have been studied have shown to have a role in mediating host responses in a similar manner to the LEE operon in that they effect normal host cellular function, but the majority of which still have roles that are undefined [25,26].

Examples of secreted toxins which are cytotoxic in nature are the Shiga toxin (\textit{stx}) and hemolysin (\textit{hly}-) SS17 is positive for \textit{stx}2 and \textit{stx}2\textit{c} but not \textit{stx}1. Similar to references, SS17 has 2 copies of \textit{stx}2, one located in a similar location to previously described EHEC O157:H7 strains (\textit{stx}2\textit{a} A and B subunits), and the other being \textit{stx}2\textit{c} also with both subunits. As previously mentioned, \textit{stx}2 is more associated with the development of HUS, and \textit{stx} when being taken up can effect immune signaling, and thus could be involved in the \textit{in vivo} phenotype [25]. This could be supported further by the fact that \textit{stx}2\textit{A}1 has an nsSNP which might make it react differently inside host cell or be more efficient at being taken up through pinocytosis to affect immune signaling.

The hemolysins that are encoded by EHEC O157:H7 are important for survival in the host, as it gives the bacteria the ability to gain access to an essential nutrient, iron by freeing heme from hemoglobin which can be taken up and utilized. This is even more important when it is considered that many of the iron acquisition genes encoded by for EHEC O157:H7 (and coincidentally SS17) when compared to other pathotypes of \textit{E. coli} are either lacking or are subject to inactivation by the host (such as enterobactin being sequestered by lipochalin) [18]. Two of these (\textit{hly}A and \textit{hly}C) have nsSNPs which could potentially affect their function; and thus might make them involved (\textit{hly}A has some
evidence being involved in adherence) [20]. The genes that are important for the synthesis of the endotoxin, such as the \textit{lpx}- family of genes, are also in both.

<table>
<thead>
<tr>
<th>Table 2-1: Toxins with nsSNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gene</strong></td>
</tr>
<tr>
<td>\textit{eivA}</td>
</tr>
<tr>
<td>\textit{eivF}</td>
</tr>
<tr>
<td>\textit{escE}</td>
</tr>
<tr>
<td>\textit{escR}</td>
</tr>
<tr>
<td>\textit{escT}</td>
</tr>
<tr>
<td>\textit{espR1}</td>
</tr>
<tr>
<td>\textit{espR4}</td>
</tr>
<tr>
<td>\textit{espX6}</td>
</tr>
<tr>
<td>\textit{Stx2A1}</td>
</tr>
<tr>
<td>\textit{lpxB}</td>
</tr>
<tr>
<td>\textit{nleH1}</td>
</tr>
<tr>
<td>17-0524</td>
</tr>
<tr>
<td>\textit{hlyA}</td>
</tr>
<tr>
<td>\textit{hlyC}</td>
</tr>
</tbody>
</table>

Table 2-3: Toxins and toxin related genes that bear nsSNPs listed with their gene, the amino acid change associated with the nsSNP and protein encoded for by the gene and (where possible) its role in pathogenesis. These are nsSNPs compared to any of the reference genomes EDL933, Sakai, EC4115, and TW14359.

2.6.2 Adherence Genes.

Many adherence genes previously identified as important for the virulence of O157:H7. An analysis of these adherence-related genes revealed 27 nsSNPs in 19 genes of the 104 genes (Appendix A-1.2) in Table 2-2. Because SS17 has such a unique difference in its adherence pattern compared to EDL933 the differences in the adherence-related genes could potentially be the most interesting to look at as it relates to its interactions with RSE and plant cells. Fimbrial and non-fimbrial adhesins, as well as regulators, chaperones and accessory proteins are all the different types of adherence
related proteins found in Table 2-2. Any of these adhesins or other adherence proteins could be definitively responsible for the differences in adherence to RSE and plant cells seen in SS17 compared to the reference strain EDL933. Both monomer protein adhesins (e.g. autotransporter adhesins) as well as various fimbrial-like adhesins have been implicated in being important in a variety of different adherences to different cell types and surfaces.

Fimbrial-like adhesins have been implicated as important for attachment of bacterial cells to many surfaces [51]. These filamentous adhesins are made up of a series of chaperones, usher proteins, and subunits making up the fimbrial complex. Similar to the references, *fim*-, *sfm*-, and *csg*- operons, among others, have all been identified in SS17 [51]. The *csg* operon is not only important for adhesion to different surfaces (including bacteria to bacteria), but is also an important set of genes for biofilm formation, which could also affect various aspects of virulence[52]. *csgG* of this group has an nsSNP and could be interesting in that it is the secretion and stability apparatus for CsgA/B major subunit, and could affect such processes [53].

Non-fimbrial like adhesins have also been implicated as important in adherence of O157, and many previously identified in references have been found in SS17. For example, *eae* (intimin), *tir* (Translocated intimin receptor), and *tccP* (Tir-cytoskeleton coupling protein) are all genes with products that are essential to form the attachment and effacing (AE) lesion on epithelial cells, have all been identified in SS17 [27]. The ability to adapt to the change in conditions in the gut where resources are scarce is incredibly important. Two other non-Fimbrial like adhesins that were previously described in increasing the efficiency of adherence to epithelial cells (one of which has an nsSNP),
yeeJ and yfaL (nsSNP), have also been identified in SS17 [54]. yfaL makes a strong case in that it has 3 separated nsSNPs and has been shown to be important for adherence but also its expression can induce a biofilm state [55].

SS17 also carries the iron regulated adhesin iha gene whose gene product is not only implicated as being important to adherence to epithelial cells; but also serves as an enterochelin transporter [56]. This can help the O157 strain adhere to colonize and employs another mechanism by which it can compete for the essential element. The cah gene is an interesting gene that encodes an Antigen-43 homologue autotransporter which has a truncation mutation. This truncation affects both the C-terminal end of cah and the N-terminal end of the putative neighboring protein. Cah has been shown to participate in aggregative phenotypes and so it could plausibly play a role in this phenotype [57,58].

Many of the usher and chaperone proteins involved with these adhesins are found in both SS17. Important global transcriptional regulators for many types of adhesins related to virulence, such as crl and hha. Crl is a regulator of csgD which regulates the second curlin operon; and oscillating regulation of Ferric Uptake Regulator Fur [59]. Hha is a positive regulator of flagellar and curlin gene expression, but is conversely an important repressor of the hemolysin genes. These do not have nsSNPs in and of themselves, but some of the genes they control do, which could be interesting potentially in a role of regulation of mutant proteins.

<table>
<thead>
<tr>
<th>Gene</th>
<th>AA Change</th>
<th>Gene Product/Role in virulence</th>
</tr>
</thead>
<tbody>
<tr>
<td>cah</td>
<td>G832W, L834A, E835R, T836Opal</td>
<td>Calcium-Binding Antigen 43 Hom. Adhesin</td>
</tr>
</tbody>
</table>
2.6.3 Virulence Associated Genes.

Genes that are considered virulence associated are genes that are involved in regulating or playing a role in the pathogenesis of EHEC O157:H7 but don’t play a direct role in virulence (so as to not be involved directly with host cytotoxicity and damage).

An analysis of the SS17 genome revealed 8 nsSNPs in 8 (Table 2-3) of the 81 virulence associated genes (Appendix A-1.3). Many of these genes have been found in SS17 in concert with what has been previously found in the reference strains. A few examples of the wide variety of genes are used by O157 strains previously described that have also been identified in SS17 are genes involved in the ability to: resist antimicrobial agents, receive and respond to signals from various sources, migrate to sites of nutrients and take...
them up, and protect against host defenses and increase adherence to the site of colonization.

For example, the ter- and sap- operons are together dedicated to protecting the cell against phage, colicins, and telereum, and antimicrobial peptides (human derived). The mot- and fli- operons control the expression and function of the flagella which have implications in both motility and adherence. Quorum sensing (QS) is an extremely important communicative process by which bacteria can talk to one another, but also can also communicate with the environment. Three QS systems (sdiA, luxS, qseE/F, genes important for the production and/or uptake of autoinducers 1-3 respectively) are found in both SS17 and references. The best example of how this is involved in virulence is QseE/F, which responds to AI-3, can also respond to human nor/epinephrine by activating the LEE operon and initiate the AE phenotype and the subsequent virulence factors [40].

Because of the vast connectivity extracellular signaling and various phenotypes in many bacteria, these would be interesting targets; but the only SNP is in a regulatory region of luxS; and so was not chosen, but could be later on. Quorum sensing is tightly linked to the development of biofilms, and as no surprise the major genes involved in biofilm and capsule biogenesis (wca- and pga- operons) [55]. The nsSNPs found in the wca- operon genes could be of interest toward our biofilm phenotype, but were not chosen yet to be knocked out. O157 E. coli employ a series of nutrient acquisition genes to take advantage of the virulent effects elicited on the host, like freeing heme via hemolysin. The chu- operon is expressed in response to the low intracellular levels of
iron and will compete for the free heme/hemin groups released by the destroyed cells, in concert with their array of other iron sequestering proteins [28].

<table>
<thead>
<tr>
<th>Gene</th>
<th>AA Change</th>
<th>Gene Product/Role in virulence</th>
</tr>
</thead>
<tbody>
<tr>
<td>chuA</td>
<td>G259E, I577N</td>
<td>TonB-Dependent Heme Transporter</td>
</tr>
<tr>
<td>chuS</td>
<td>K11T</td>
<td>Heme Oxygenase</td>
</tr>
<tr>
<td>phoB</td>
<td>Q179L</td>
<td>Phosphorus Signaling regulator</td>
</tr>
<tr>
<td>terD</td>
<td>M155R</td>
<td>Tellurite and Phage Resistance</td>
</tr>
<tr>
<td>terF</td>
<td>S103Opal</td>
<td>Tellurite and Phage Resistance</td>
</tr>
<tr>
<td>wcaJ</td>
<td>D216D</td>
<td>Colanic Acid and Biofilm Synthesis</td>
</tr>
<tr>
<td>wcaL</td>
<td>R73S</td>
<td>Colanic Acid and Biofilm Synthesis</td>
</tr>
<tr>
<td>wcaM</td>
<td>L29F</td>
<td>Colanic Acid and Biofilm Synthesis</td>
</tr>
</tbody>
</table>

2.7 Different Mechanisms of Adherence of EHEC O157:H7

EHEC O157:H7 strains encode a plethora of different adhesins and proteins that aid in the adhesion process. This great diversity in types of adhesins help EHEC O157:H7 attach to a variety of different surfaces and interact with many different types of cells [58,60,61,62]. This makes sense in that EHEC O157:H7 has many different environments in which it is found and can colonize such as cattle, sheep, goats, humans and leafy greens. There are two major types of adhesins found in EHEC O157:H7: Fimbral and non-fimbrial like adhesins. Fimbral and fimbrial-like adhesins are a group of adhesins that are generally comprised of 2 major subunits expressed from the outer membrane into the environment [51]. This is the portion that will interact with cellular receptors of the host and in some instances with abiotic surfaces. Most fimbral like adhesins are multimers that are attached by a secretion apparatus or a complex found in the outer membrane and in some instances spanning the periplasmic space. Many of
them have family-specific chaperones, secretion proteins, and other related proteins that help in the folding and stability of the fimbriae [51]. EHEC O157:H7 encode for different types of fimbrial-like adhesins such as Type I Fimbriae, common pilus forming adhesins, and curlin fibers [51,52].

Non-fimbrial adhesins are adhesins which are single proteins (not generally multimers or complexes) which act in and of themselves as adhesins. Many of these also will have specific host cellular targets based on the different motifs or domains the proteins have. For example, adhesins that contain an RGD domain in their extracellular expressed adhesin will selectively bind to integrins and selectins of eukaryotic cells [63,64]. One type of non-fimbrial adhesin that is used in a variety of adhesion properties of EHEC O157:H7 are autotransporter adhesins [57,64,65,66]. These proteins are characterized by a β-Barrel domain, a leader domain that plays a role in transport of the precursor in the passenger domain, and the passenger domain that is passed through the membrane via the β-barrel and participates in the actual adhesion [64,67,68].

As critical as adhesins are to attachment, many bacteria including EHEC O157:H7 employ a series of genes to produce biofilms for long standing colonization, intra-biofilm signaling and protection against outside insults [17,55,69,70]. There are many adhesins involved in biofilm formation as well as several loci (e.g. the wca- locus) which encode genes that help synthesize the polymers that make up the biofilm [55,71]. Biofilms have been implicated as being a crucial component to establishing colonization on both abiotic and biotic (reactive) surfaces and is just as important to communication such as chemotaxis and quorum sensing [69,72]. Bacterial biofilms have a life cycle of their own as they do have a carrying capacity for shared nutrients and space (Figure 2-7).
Figure 2-7: The General Lifecycle of a Biofilm
The lifecycle of a biofilm of most bacteria follow this pattern where motile bacteria will land on a surface where a genetic switch occurs and the bacteria become sessile and promote biofilm formation. Once the biofilm has reached a critical threshold, the bacteria nearest the biofilm border will become motile again and find new location to lay foundation of a new biofilm. Figure from: http://prometheus.mse.uiuc.edu/glossary/biofilms/lifecycle.png

Sessile bacteria that have attached to a surface will start to secrete the polymers required for a biofilm. These bacteria will divide and the new bacteria will also produce more biofilm. The process augments until a breach point where the newest bacteria will break through the biofilm and swim to a new destination [17,55].

All of these processes are in line step with one another and are critical for the pathogenesis of EHEC O157:H7. Without attachment, there can be no colonization, without colonization there can be no establishment for infection or carriage (per example in a ruminant). There are many genes involved in both processes and many genes that overlap the two. For example, cah and csg- operon are genes which are implicated as being important to EHEC O157:H7 for binding to a variety of surfaces, but also important for aggregation and adherence of bacteria to each other and the formation of the biofilm structure [73,74].

The SS phenotype is a huge threat to human health given the propensity of this organism to cause disease. All of these different facets of the lifecycle of EHEC
O157:H7 are all ways in which they can infect human hosts. Understanding how they mediate these different aspects of their transmission and spread is a way to understand how to mitigate the disease. Interactions with the cattle (in the RAJ), and how it can be attached and survive on produce are two areas not well understood but could go a great way to better understanding how this organism lives in different environments and the factors involved. By understanding these molecular mechanisms it could potentially lead to prevention of the outbreaks this pathogen causes.
Chapter 3
Materials and Methods

3.1 Bacterial Strains and Plasmids

Terrence Author of the USDA-ARS Meat Animal Research Center obtained and characterized 102 super-shedder isolates of *Escherichia coli* O157:H7 in a survey of approximately 1500 feedlot cattle and 3500 cattle from slaughter plants. Isolates were obtained in the summer months over a two year period in Midwestern States. Samples were placed into 4mL of tryptic soy broth (TSB; Difco, Becton Dickenson, Sparks, MD) and shipped to the laboratory on ice. Samples were enumerated for *E. coli* O157:H7 by plating onto CHROMAgar O157 (Becton Dickenson, Sparks, MD), and colony forming units (CFU) per swab (approximately per gram) were calculated to identify super shedders [9]. Animals, whose counts were greater than $10^4$ CFU/swab (equating to approximately a gram of fecal matter) were classified as super shedders and isolates were further characterized [9]. Up to twenty colonies were picked for PCR to confirm that each *E. coli* isolate harbored genes for the O157 antigen, H7 flagella, γ-intimin, and at least one of the Shiga toxin genes. Isolates were also further characterized by phage typing and *XbaI* PFGE as previously described, and represented in Figure C-1 [9]. Of the 102 SS isolates, isolate 17 (SS17) was chosen as it was a good, broadly representative super shedder isolate that is phage type 4 and had $3.1 \times 10^6$ CFU/swab.
The knockout strains (Table 3-3) that were created were chosen by comparing the genome of SS17 with those of other well characterized \textit{E. coli} O157:H7 isolates (Table 3-1). Though their super shedder status remains unclear, these reference strains will be considered until otherwise shown. In order to perform comparative alignments of the nsSNPs and GOIs, 11 other SS isolates were chosen to be sequenced and scanned for these nsSNPs identified in SS17 (Table 3-1).
Table 3-1: Parent Strains and Isolates Used in This Study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Isolation Sources</th>
<th>Genotype*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MG1655 (K12)</td>
<td></td>
<td>K12</td>
<td></td>
</tr>
<tr>
<td>EDL933</td>
<td>Ground Beef</td>
<td>EDL933</td>
<td>[75]</td>
</tr>
<tr>
<td>Sakai</td>
<td>Human</td>
<td>Sakai</td>
<td>[76]</td>
</tr>
<tr>
<td>TW14359</td>
<td>Human</td>
<td>TW14359</td>
<td>[77]</td>
</tr>
<tr>
<td>EC4115</td>
<td>Human</td>
<td>EC4115</td>
<td>[78]</td>
</tr>
<tr>
<td>86-24</td>
<td>Human</td>
<td>86-24</td>
<td>[79]</td>
</tr>
<tr>
<td>SS1</td>
<td>Bovine Feces</td>
<td>SS1</td>
<td>[9]</td>
</tr>
<tr>
<td>SS7</td>
<td>Bovine Feces</td>
<td>SS7</td>
<td>[9]</td>
</tr>
<tr>
<td>SS12</td>
<td>Bovine Feces</td>
<td>SS12</td>
<td>[9]</td>
</tr>
<tr>
<td>SS17</td>
<td>Bovine Feces</td>
<td>SS17</td>
<td>[9]</td>
</tr>
<tr>
<td>SS42</td>
<td>Bovine Feces</td>
<td>SS42</td>
<td>[9]</td>
</tr>
<tr>
<td>SS52</td>
<td>Bovine Feces</td>
<td>SS52</td>
<td>[9]</td>
</tr>
<tr>
<td>SS67</td>
<td>Bovine Feces</td>
<td>SS67</td>
<td>[9]</td>
</tr>
<tr>
<td>SS77</td>
<td>Bovine Feces</td>
<td>SS7</td>
<td>[9]</td>
</tr>
<tr>
<td>SS67</td>
<td>Bovine Feces</td>
<td>SS67</td>
<td>[9]</td>
</tr>
<tr>
<td>SS131</td>
<td>Bovine Feces</td>
<td>SS131</td>
<td>[9]</td>
</tr>
<tr>
<td>RM11326</td>
<td>Environmental</td>
<td>RM11326</td>
<td></td>
</tr>
<tr>
<td>RM11333</td>
<td>Environmental</td>
<td>RM11333</td>
<td></td>
</tr>
</tbody>
</table>

*If Genotype has been determined, please refer to NCBI or Reference given.

Table 3-2: Plasmids Used in This Study

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Isolation Sources</th>
<th>Genotype*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC177</td>
<td>NEB</td>
<td>kanR, ampR, p15A</td>
<td>NEB</td>
</tr>
</tbody>
</table>

Plasmid information can be found at [http://cgsc.biology.yale.edu/Site.php?ID=80304](http://cgsc.biology.yale.edu/Site.php?ID=80304)
Table 3-3: Mutants Generated for this Study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS17Δiha</td>
<td>SS17Δiha::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔompA</td>
<td>SS17ΔompA::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔcsgG</td>
<td>SS17ΔcsgG::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔwzzB</td>
<td>SS17ΔwzzB::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17Δcah</td>
<td>SS17Δcah::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17Δeae</td>
<td>SS17Δeae::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔeaeH</td>
<td>SS17ΔeaeH::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔeivA</td>
<td>SS17ΔeivA::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>SS17ΔyfaL</td>
<td>SS17ΔyfaL::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933Δiha</td>
<td>EDL933Δiha::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔompA</td>
<td>EDL933ΔompA::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔcsgG</td>
<td>EDL933ΔcsgG::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔwzzB</td>
<td>EDL933ΔwzzB::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933Δcah</td>
<td>EDL933Δcah::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔeaeH</td>
<td>EDL933ΔeaeH::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔeivA</td>
<td>EDL933ΔeivA::kanR</td>
<td>This Study</td>
</tr>
<tr>
<td>EDL933ΔyfaL</td>
<td>EDL933ΔyfaL::kanR</td>
<td>This Study</td>
</tr>
</tbody>
</table>

*Genotype of every mutant is the same as the parent aside from the genetic alteration listed.

3.2 Gene Selection Process

To generate a list of potential genes of interest (GOIs), three main attributes were looked at. The first attribute was genes in the genome that were listed as potential or known adhesins (both fimbrial and non-fimbrial). This was done by NCBI BLASTp identification of proteins and subsequent literature searches to determine the type of adhesin and data to support their role as an adhesin or adhesion-related protein. The second factor was the presence of a non-synonymous SNP (nsSNP) which resulted in either a missense or nonsense mutation in the coding sequence. Non-coding SNPs were considered and added, but the primary focus of this work was on genes with nsSNPs in the coding region of the gene. nsSNPs were discovered by a combination of MAUVE
Genome Alignment tool and manual confirmation via NCBI BLASTp from the sequence loaded into the Artemis program (especially for nsSNPs that caused missense mutations in coding regions) [80,81]. The last criteria looked at was the expression of any of the aforementioned genes in either the top-down proteome analysis provided by Kudva and colleagues or in the transcriptional profile on leafy greens (Unpublished) [14]. Meeting all three of these criteria listed the gene as a top priority and any gene matching two or more of these would be in the next tier of priority to be knocked out. This system of selection is represented by Figure 3-1.

3.3 Recombineering to Generate Knockouts in SS17

A combined recombineering protocol was used to generate the cassettes and knockouts of gene targets in SS17. Parts of protocols from Datta and colleagues and Yu et al. were used to develop the general parameters for performing the recombineering [82,83]. Recombineering is the technique in which helper proteins which act in
homologous recombination are used for allelic exchange of a gene of interest with a cassette bearing a selectable marker. In brief, KanR was amplified from pACYC with primers that amplified the KanR gene with arms of homology to flanking regions of genes of interests to generate KanR cassettes to knock out these genes. Cassettes were introduced into electrocompetent SS17 previously electroporated with and carrying the pKD119 plasmid (Yale *E. coli* Research Stock Center, New Haven, CT.), and subsequently electroporated, re-struck to generate isogenic colonies, and colony PCR was performed to identify mutants bearing the KanR insert. KanR inserts and gene deletions were confirmed using PCR based sequencing (Pennsylvania State University Sequencing Core, University Park, PA).

### 3.3.1 Generating Kanamycin-Resistant Cassettes for Recombineering

70mer primers were used to generate all of the kanamycin resistance cassettes that were used in the recombineering protocol. These 70mers contained 20 nucleotides (nt) for amplification of the kanamycin-resistance gene (KanR) from a pACYC vector and the remaining 50 were designed as

![Figure 3-2: Schematic of generating KanR Cassette](image)

This figure shows the two orientations of the KanR gene can take based on the gene of interest and the primer design for the 70mers.
arms of homology. The 20nt portion was designed based on the orientation of the gene of interest (either on the forward strand or complimentary strand). The 50nt portion was designed by viewing the sequences flanking the gene of interest, and without disrupting flanking genes, selecting 50nt and linking it to the previously designed 20mer. These sequences were found from the annotated sequenced genome of SS17 loaded onto Artemis genome editor, copying the sequence around the gene (5’ and 3’) into ApE plasmid editor.

Sequences were selected and combined with the sequences preset for the KanR gene and the whole 70mer was analyzed for homodimers, heterodimers, and secondary structure by the OligoAnalyzer from Integrated DNA Technologies (IDT). Figure 3-2 shows this in better detail. pACYC177 is the vector which houses the KanR gene and it is made linear by an XmnI digest for 1 hour, which cuts the pACYC177 in 2 places, separating the KanR from the AmpR genes. After primers were received, they were diluted to a working stock of 5µM and used in a PCR reaction with the XmnI digested plasmid to generate the KanR construct. After PCR was completed, some of the sample was run out of a 1% agarose gel with ethidium bromide (EtBr) and was analyzed for a size of 1086bp (~1.1Kb) which would be indicative of the size of the KanR gene and the 70mers. Once confirmed this is then DpnI digested overnight to ensure maximum digestion of any parental plasmid remaining and then purified via filter restriction. Nanodrop analysis was performed to calculate the concentration of each cassette.
3.3.2 Making Electrocompetent Cells and Induction of pKD119.

SS17 pKD119 cells were made electrocompetent generally by the protocol described by Yu and colleagues with a few exceptions because we did not use the pSIM plasmids [83]. Because of the difference in plasmid, the cells were grown at 30° C to an O.D. ~0.4-0.6 in LB plus 7µg/mL tetracycline (Tet7) and 1mM arabinose, described by one of our colleagues (Kariyawasam laboratory). Electrocompetent cells were finally resuspended (after washing and spinning at 4°C), in 2mL 10% glycerol and either continued on for transformation or stored at -80° C. After electroporation, for the last step of induction, cells were recovered in SOC (Super Optimal broth with Catabolite repression) media also containing 1mM arabinose.

3.3.3 Transformation, Electroporation and Screening

Competent SS17 pKD119 were incubated on ice for ~5 minutes with ~200ng of DNA and electroporated at 1.8mV. After electroporation, cells were suspended in 500µl of SOC media containing 1mM arabinose and recovered at 37° C for 1 hour. After cells were recovered, all 500µl was plated onto LB + 50µg/ml kanamycin (Kan50) or 30µg/ml if the gene is not well expressed under LB conditions, and incubated overnight at 37° C. The next day, 4-8 colonies were picked and re-struck (4-way streak technique) onto LB + Kan50 to isolate isogenic colonies to screen for KanR insert. We determined this double selection process significantly increased our efficiency on identifying mutants that were correct. These re-streaks were also incubated in the same manner over night.
Prior to selecting colonies for colony PCR, 4 primers were designed for quick identification of potential mutants. Two internal KanR primers along with 2 primers of flanking sequences were designed based on sequence and orientation of the Kan cassette (because we are driving KanR from the resident promoter as to not introduce secondary effects of the constitutive KanR promoter from pACYC, KanR needs to be in the same orientation as the GOI). The primers were listed as GOI-A (forward primer), GOI-B (reverse complement primer), KanC and KanD (2 for the forward strand Kan and 2 for the complement strand Kan). From these primers two PCR reactions were performed with the following pairs: A&B, A&C, and the B&D would have been run should the A&C primer pair fail (which didn’t happen). These mutants that had presence of KanR insert and correct A&B size variant were sent for sequencing at the Pennsylvania State University Sequencing Core with the A&B product being the one sent (to check on flanking sequences and if KanR landed correctly and right orientation). Figure 3-3 shows

![Figure 3-3: Schematic of Primer annealing for Colony PCR Screen](image)

This schematic shows the orientation of the primers and where they land during colony PCR and the sequences they are designed from. This primer pairing holds true for both orientations of KanR.
the schematic of where the primers are designed against and their partners for both the GOI and the KanR cassette.

### 3.4 RSE Cell Adherence Assay

The RSE cell adherence assays were conducted by Indira Kudva at the USDA-ARS National Animal Disease Center (NADC) and performed in two biological replicates and 4 technical replicates. SS17 and control O157:H7 strains, EDL933 and 86-24, were cultured overnight in Dulbecco Modified Eagle Medium-Low Glucose (DMEM; Gibco/Invitrogen Corporation, Grand Island, NY) at 37°C without aeration, pelleted and resuspended in sterile saline as previously described [14]. RSE cells were suspended in 1 ml DMEM–No Glucose (DMEM-NG) ± 2.5% D + Mannose to a final concentration of 10⁵ cells/ml. For each strain, bacteria were mixed with RSE cells to achieve a final bacterial to cell ratio of 10:1. The mixture was incubated at 37°C with shaking for 4 hrs. The mixture was then pelleted, washed thoroughly, and reconstituted to 100μl in ddH₂O. Drops of the suspension (2μl) were placed on Polysine (Thermo Scientific Pierce, Rockford, IL) slides and dried overnight under direct light to quench non-specific fluorescence, before fixing in cold 95% ethanol for 10 minutes.

The slides were stained with 1% toluidine blue or with DAPI and fluorescence-tagged antibodies specific to the O157 antigen and cytokeratins of the RSE cells. Adherence patterns on RSE cells were recorded as aggregative, diffuse, or non-adherent [14]. The percent of RSE cells with bacteria adhering to them was determined. Strongly positive was recorded when more than 50% of RSE cells had >10 bacteria attached, as
moderately positive when 50% or less of the RSE cells contained 1–10 adherent bacteria or as non-adherent when less than 50% of the RSE cells had only 1–5 adherent bacteria.

In order to determine the involvement of LEE-encoded proteins, rabbit antisera generated against EspA, EspB, Tir, and intimin (stock from NADC), were pooled and tested at a 1:50 dilution. Assays were preformed as described above with the antisera added to the resuspended bacterial pellets and incubated at 37°C for 30 min before mixing with the RSE cell suspension [14]. Human HEp2 cells were also used in place of RSE cells for comparative purposes.

3.5 Lettuce/Spinach Adherence Assay

This assay is courtesy of Dr. Maria Brandl and Dr. Michelle Carter of the USDA. Whole baby spinach leaves were immersed upside down into a suspension of 1x10^7 CFU/ml of EHEC O157:H7 (either SS or nonSS) potassium phosphate buffer taking care not to immerse the cut petiole. The leaves were incubated in the suspension, and then rinsed sequentially in triplicate. Two 1-cm diameter discs were cut out per leaf, and each disc was homogenized individually in a mortar and pestle in 2 ml buffer. The resulting homogenates were dilution-plated onto CT-SMAC agar for plate counts. Means were log-transformed and compared by ANOVA with Tukey's Multiple Comparison.
3.6 Biofilm Assays

A single colony was picked from a LBHS (LB half-salt) plate added to 1.5mL of LBHS and incubated overnight at 28° C, 150 rpm to stationary phase of growth: LBHS for SS17 WT, EDL933 WT, LBHS+ 30 µg/ml Kan for ΔeaeH, LBHS+50µg/ml Kan for all other mutants. 1 mL of culture was washed twice and spun down at 8000g for 3 min. The supernatant was removed and resuspended in 1mL sterile potassium phosphate buffer (10mM, pH 7.0) (KP), and was again spun down at 8000g for 3 min and supernatant was removed and resuspended in 1mL sterile KP buffer. OD600 was measured and adjusted to 1.0 with KP buffer resulting in suspension of 10⁹ cells/ml. This was then diluted in 3 times in a 1:10 manner, to have cell counts of 10⁵-⁷ cells in KP buffer (Mixed well by vortexing).

The final concentration of cells from the overnight was 10⁵ cells/ml of which 30µL of cells was added to 3ml of spinach lysate, mixed by vortexing, and aliquoted in 1ml increments into a glass tube then incubated for 24 hours at 28°C without shaking. Spinach lysate was prepared according to Carter et al 2012 and used undiluted [84]. Two replicate tubes were prepared per colony per strain at a final concentration of 10³ cells/ml. The control for each time point was 2 tubes of uninoculated spinach lysate with only KP buffer for every time point. Sampling occurred at the following time points: first at 16 hours, the second at 23 hours, and final time point was at 40 hours. 24 h incubation time was decided upon since the biofilm measurements did not change after 24 h when normalized to the control. After incubation, the planktonic cells were removed and tubes were rinsed twice by gently adding 1 mL sterile distilled water to the tube and draining.
Biofilm rings remaining on the glass were stained with 0.1% crystal violet (CV) for 30 min at room temperature. The CV was removed and tubes were rinsed twice with sterile distilled water. The tubes were inverted and dried at room temperature overnight.

The crystal violet bound to the biofilm was solubilized with 33% acetic acid at room temperature with shaking at 180rpm and constant vortex for 30 min, by which time the stained ring has completely dissolved in the acetic acid solution. 100μl of the resulting solution was dispensed into each of 2 replicate wells of a 96-well plate and the absorbance was measured at 570nm using a microplate reader (SpectraMax 340, Molecular Devices). Data normalized by subtracting the mean OD570 of the two replicate control tubes from the mean OD570 of the 2 replicate wells for each biological replicate. Data analyzed with Prism Version 5.02 (GraphPad).

3.7 Analysis of nsSNPs Across the SS Isolates

In order to determine if any of the genes of interest had a higher correlation to being involved in the change in adherence phenotypes for both the plant and RSE models, we had to determine the presence of nsSNPs through various super shedder isolates in those particular genes. In brief, glycerol stocks from the original stabs of 12 super shedder isolates (SS17 (to confirm observed nsSNP), SS1, SS7, SS12, SS27, SS52, SS67, SS77, SS131, RM11333and RM11326) were struck out onto LB agar and isogenic colonies were picked, grown overnight in LB, and DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen, Valencia CA) and eluted in ddH2O. From the DNA prep 2μl was added to a 20μl PCR reaction with the newly constructed S1 and S2
primers that annealed to sequences approximately 250bp flanking the nsSNP(s). Two genes (*iha* and *ompA*) were sequenced through primer walking strategies as no nsSNP was previously identified.

PCR was performed, products were run out on a 1% agarose gel (1/10 volume) and the remaining product was cleaned by filter exclusion and sent for sequencing in both the forward and reverse direction to have confirmation of the nsSNP. Sequences were aligned using the ApE Plasmid Editor and SeqMan (DNASTAR, Madison WI; Lasergene v. 10) programs and analyzed for either the conservation of a particular nsSNP or the same gene with a different nsSNP local to the original nsSNP by alignment to EDL933 and SS17 for references.

Phylogenetic analysis was performed using MEGA [85]. Codons of containing the SNPs in 9 genes were aligned by formation of FASTA format to generate MEGA alignment. The deletion in *cah* was accounted for by a coded change in a codon. The alignment of the 27 nucleotides was then used to generate the phylogenetic through a Bootstrap 1000 maximum nucleotide likelihood of all the SS isolates. Confidence percentages and sequence alignments were added to the phylogenetic tree to indicate the sequence/SNP tested linked to a certain gene and certain isolate.

### 3.8 Growth Curves and Doubling Times

In brief overnight cultures of EDL933, SS17, and SS17Δ*ompA*, *cah*, and *yfaL* were grown in LB at 37°C shaking at 130rpm. Cells were then subcultured into LB in a 1:200 dilution (resulting in an OD600~0.005) based on final OD600 of the overnight
culture and growth was followed to ~0.2 OD600. Three biological replicates were re-
inoculated into 50ml of LB from the starter culture now in log phase in a 1:50 dilution. These cultures were then grown at 37°C and at 150rpm. Growth was monitored every half hour by reading OD600 of each replicate. Growth was followed to the start of the plateau (entry into stationary phase). The results were recorded, averages and standard deviations calculated and the results were plotted. Doubling times for each strain were also determined through the formula \((t_2 - t_1)/\log_2((O_3/O_2)\times50)\) where \(T_2-T_1\) is the time it takes from the re-subculture to reach an OD600 of ~0.2, \(O_3\) is the actual OD600 reading at that time point, \(O_2\) is the OD600 reading from the starter culture around the OD600 of ~0.2. \(O_3\) is multiplied by 50 to account for the 1:50 dilution, and \(\log_2\) because we are looking for doubling time. These standards were developed by Michael Mwangi and applied to this study.
4.1 Strong Aggregative Phenotype on RSE is Shared Among SS-Isolates

Though it is interesting that SS17 shows a LEE-Independent strong aggregative phenotype on RSE cells, it had to be tested if other super shedders shared the same phenotype. If the SS isolates of different lineages, origins, PFGE clusters (etc…) shared a common phenotype, it is very plausible that the factors involved in the change of interaction between SS isolates and RSE cells in the cow RAJ may be conserved and give credence to them as being important for the SS phenotype. The interaction of SS-isolates and FAE cells have not been tested, but since it was previously determined that this interaction is mediated by the LEE-operon, and we know the change is LEE-independent (as is the general interaction of EHEC O157:H7 with RSE cells), that this is most likely not the cell type of interest for this phenotype [15]. For good measure it may be good to look at these cells interaction with SS EHEC O157:H7 and see if the adherence pattern is different compared to EDL933.
Four other super shedders were tested for their adherence on RSE cells (SS17, SS52, SS67, and SS77), and Figure 4-1 shows that this phenomena is not unique to SS17 but to all super shedder strains tested. All of the super shedders share the strong aggregative adherence phenotype on RSE cells, which further suggests that this mechanism and microbial factors involved in the adherence and aggregation on the RSE cells could be critical for the super shedding phenotype. This also made us believe that there would be great power in being able to sequence and compare all the genomes of the super shedder isolates. Alternatively, because SS17 has been sequenced, we can look for shared nsSNPs and other genome alterations and compare them to other SS isolates to look for patterns or total of conservation of the nsSNPs to find genes of interest.

Because of this shared phenotype across PFGE and Lineage clusters (though a lineage I is not present), there is a possibility that the mechanism will be conserved and the changes in the genome compared to the reference genomes of non-super shedder isolates. There is always the possibility that the changes seen (either nsSNPs or novel
gene insertion/deletions) could be relevant for different PFGE clusters, lineages, or even for each super shedder; but considering the phenotype is shared, it is more likely that the nsSNP or novel gene(s) would be conserved (e.g. nsSNP conserved in the same gene across the super shedders), but also could have different nsSNPs in the same gene (or same pathway) that has the same consequence.

4.2 SS17 Displays Enhanced binding to Leafy Greens

The super shedder isolates have shown increased adherence to RSE cells and have been shown to track to lineage I/II (see section 2.4.3) we wished to discover if this enhanced adherence would carry over to plants. We know that it doesn’t carry over to the HEp-2 cells but these cells are not the true human cells which would be bound to by the EHEC O157:H7 and it is well documented that these interactions are mediated by the LEE operon. To test this, a spinach adherence assay was performed on 3 SS isolates (SS17 and two environmentally recovered SS isolates RM11326 and RM333) compared to a reference strain EDL933. As Figure 4-2 shows SS17, RM11326 and RM11333 all show enhanced attachment to the baby spinach when compared to EDL933, which shows that there is a maintained enhanced adherence to multiple biotic surfaces. This is extremely interesting as it shows another distinctive phenotype unique to isolates identified as super shedders. As with the RSE adherence assay data, there could be a difference in the SS isolates that would explain this phenotype.
Figure 4-2: SS17 and Other SS Isolates Show Enhanced Adherence to Baby Spinach

SS17, RM11326, RM11333 are all previously identified SS isolates and EDL933 is the non-SS control (as described by RSE adherence assay). All SS isolates show approximately a 1-log increase in adherence compared to EDL933, indicating another unique feature to the SS phenotype. * Indicates significant difference and error bars are from standard error.

Plant cells are different than RSE cells (plant vs. animal cells), but there is no data to suggest that this is two separate mechanisms of action (or factors), but could be considering. It is also not known what the microscopic pattern of adherence of SS strains are with the spinach leave cells as was determined with the RSE cells. Either way this is extremely interesting in the context of outbreaks and human infection. This is important as it has been shown that a super shedder cow will increase the environmental load and the transmission of EHEC O157:H7, and so the increased load coupled with increased adherence to plants both increase the probability of infecting other hosts (like humans).

4.3 SS17 Displays Enhanced Biofilm Production in Damaged Spinach

SS17 was compared against EDL933 for formation of biofilm production that can be induced by leafy green extract (see Materials and Methods) replicating factors associated with damages on a spinach leaf. As seen on Figure 4-3, leafy green extract
induced biofilm formation in both EDL933 and SS17. SS17 displays enhanced biofilm formation compared to EDL933. This is very interesting in that it maybe another part of the life-cycle adapted by the SS isolates. From previous information it has been determined that SS17 has a unique RSE adherence phenotype, exhibits the SS phenotype in vivo, and shows enhanced adherence to leafy greens. On plants, like many other biotic (or reactive) surfaces, biofilms are incredibly important as a means of survival and enhanced adherence.

Figure 4-3: SS17 Displays Enhanced Biofilm Formation from Leafy Green Extract
SS17 and EDL933 were compared in their ability to form biofilms on glass after growth in leafy green extract. * Indicates significant difference and error bars are indicative of standard error.
4.4 Rationale and Background of the Genes of Interest

To start to narrow down the genes of interest that were good candidates to be knocked out and tested, the genes from both the microarray and proteome and genes from Tables 2-1 to 2-3 were assessed and chosen by the system outlined in the Materials and Methods. Below is the list of genes grouped by their relative functions and a brief rationale and background as to how they were chosen. NsSNPs can result in one of three possible outcomes; the first being no change in function, second being loss of function, and lastly is a change in function-expression. All three possibilities are distinctly possible in any of these cases and a combination of them could potentially be involved in any number of phenotypes now associated with SS isolates. It is on this basis we chose most of the genes or genes that interacted with genes that did not have one.

4.4.1 Genes Involved in Adherence and Biofilm formation

The gene ompA codes for a β-Barrel outer membrane protein which is found in many serotypes of *Escherichia coli*, including O157:H7 and meets the standards set in category IIb. Although an analysis of the SS17 genome showed that the *ompA* gene without an nsSNP, there are many reasons that this is such an interesting target. First, it is expressed in the proteome published by Kudva et al. which details the protein (including adhesins) deployed by EHEC O157:H7 given the limiting nutrient conditions they were grown in [14]. It has been shown to be important for the formation of a hyperadherent phenotype when overexpressed and diminished binding of WT EHEC O157:H7 by 13.5% to Caco-2 (important because they are colonic cells) and HeLa cells,
as well as it appears to be critical for the binding to alfalfa sprouts [58,61,86]. It has been established (albeit by an unknown mechanism) that $ompA$ mutants have diminished capacity to generate Type I fimbriae based on the inversion switch, which may have further implications on adherence and biofilm formation [86]. This has also been shown to be involved in immune evasion but also to be a main target of the immune system (as a major antigen). The only instance where OmpA has been shown to be important for virulence is it is essential for K1 *E. coli* to be capable of being septic or meningeal infection, as well as being involved in immune evasion despite being a major antigen [86]. The regulator of $ompA$ is $tdcA$ (or any member of the $tdc$- locus), and it does not have an nsSNP either. $tdcA$ is the only known regulator of $ompA$ known at this time (a $tdcA$ mutant overexpresses $ompA$), but it is not known if it is the only regulator [61].

$csgG$ is a gene that encodes an outer membrane protein which is essential for the expression, secretion, and stability of the curlin fibers, and is a category IIa gene [53]. Though it is not in the proteome provided by Kudva and colleagues, two positive regulators of the curlin operons (Ihf and Crl) was expressed [14]. $csgAB$ represent the 2 genes that are produce the major curlin fiber, which are regulated (as are the accessory proteins) by $csgD$. $csgG$ is regulated by $csgD$ gene product (and thusly regulation on $csgD$), and is essential for expression and stability of the curlin apparatus as a whole[87,88]. CsgEFG are the accessory components of the curli [52,53,59].

Curli fibers are fimbrial-like adhesins which are involved in adhesion to eukaryotic cells, to plant cells, and are involved in invasion of certain *E. coli* pathotypes, as well as has been shown to be important for aggregative phenotypes [58]. Though they have not had any direct evidence to be directly involved in pathogenesis, but many
pathogenic *E. coli* do express them *in vivo*. Consequently, curli fibers have been shown to promote adherence to multiple cell lines as well as binding to plants, as well as are critical for biofilm formation. Both O157:H7 and *Salmonella* have increased attachment to cultured mouse intestinal epithelial which correlates nicely to the fact that they are more virulent in the mouse model[52]. Curli production is also important for biofilm secretion and stability [52]. The only information against *csgG* is that it was not identified in the proteome provided by Kudva and colleagues [14].

The *cah* gene (calcium binding-antigen 43 homologue), is another autotransporter adhesin gene identified in *E. coli* O157:H7 which has the typical structure of an AIDA-I like autotransporter with the only exception being that it does not have a pertactin-like domain [74]. This gene is a category IIa gene. It is described as having a transmembrane domain including calcium binding motifs in the passenger domain, and a $\beta$-Barrel domain that includes an RGD motif, classically identified as having $\beta$-integrin ligands on eukaryotic cells [63,74]. *cah* is critical for the formation of biofilms as well as can confer adherence to non-adherent bacteria to both mammalian and plant cell surfaces, and the deletion of *cah* reduced the ability of the bacteria to generate a biofilm and the autoaggregative phenotype, but did not eliminate adherence to Caco-2 and HeLa cells and alfalfa sprouts [58,60,74,89].

Cah expression has been implicated as being involved in an autoaggregative phenotype [60]. All of these phenotypes, as well as Cah itself is confined to being conserved in diarrheagenic *E. coli*, which implicates Cah as a virulence contributing factor, as well as its involvement in biofilm and autoaggregation phenotypes [89]. Cah does not have any evidence in being involved in binding eukaryotic immortalized human
cells (Caco-2 and HeLa cells), but we are interested in binding to RSE cells, and it also may be a conditional regulation that is unknown. This is a truncated protein, and is not expressed in the Kudva proteome [14].

The *yfaL* gene was originally discovered by a combined inactivation study in MG1655 *E. coli* K12 as a gene involved in enhanced adherence when overexpressed and formation of more stable biofilms to abiotic surfaces [54]. A BLAST search through 28 different genomes of *E. coli* of autotransporter (Type V Secretion Apparatus) was used to identify novel autotransporters by comparing to previously published sequences of identified autotransporters. *yfaL* was found to have similarity to the AIDA-I group of adhesins, which are implicated in many binding events in *E. coli* such as binding to eukaryotic and prokaryotic target cells [64]. AIDA-I like proteins have a Pertactin-like domain and an autotransporter-like domain, and YfaL contain both along with the signal peptide commonly seen with these kinds of proteins.

The effectors of YfaL have not been identified, but AIDA-I like autotransporters have been shown to secrete many different things such as invasins, adhesins, toxins, etc…, making this a very interesting target [64,65,68]. For example, Uropathogenic *E. coli* upregulates *yfaL* in vivo along with expression of other virulence factors such as flagellar protein and hemolysin, and all of this in a non-LEE mediated manner [90]. Both the former and the latter indicate that in pathogenic *E. coli* this protein can be virulence-associated and the fact that is expressed in K12 might only be by conservation as it can induce biofilm formation[55]. The fact that is expressed in K12 (though conditional or other gene interactions might be at play) and the fact that it was not found in the
proteome provided by Kudva and colleagues are the only two reasons why this gene would not be a good target to generate a knockout [14,54].

4.4.2 Genes Involved in Virulence and Adherence

\textit{wzzB} is the gene responsible for O-antigen (OAg) small chain length modal distribution, which is involved in regulating the variable portion of the OAg length and distribution there within and is a category I gene [91]. It is conserved in many different enteric organisms (such as \textit{Salmonella}, \textit{Shigella}, and \textit{Vibrio} species) [91,92]. O-antigen, in combination with LPS, have been shown to be important for binding to epithelial cells as well as is upregulated in binding to plants and is expressed in the proteome provided by Kudva \textit{et al} [14,91]. \textit{WzzB} plays a role in the virulence and pathogenesis of different enteric organisms such as EHEC O157:H7 (in a non-LEE mechanism), \textit{Shigella} and \textit{Salmonella}, not only by supporting adherence to host cells, but also playing a role in immune evasion, complement protection, and damaging effects of the endotoxin [93].

Similarly, OAg has been implicated in being important for the colonization of the GI tract of various organisms; most importantly for this project is the observation that a deletion in genes responsible for OAg shows decreased colonization of the bovine rectal mucosa, an important feature of the life cycle of EHEC O157:H7 by a (perosamine synthetase) \textit{Per} knockout [93]. The only reason that this could not be a good target is if the nsSNP in \textit{WzzB} would synthesize an OAg with a different O-antigen modality which might change the O-serotype which has already been determined to be O157 for all of the SS isolates [94].
The *eaeH* gene encodes a putative AIDA-like adhesin identified sequence similarity to the *eae* gene in the sequence of EDL933 *E. coli* O157:H7, and is a category IIa gene [67,95]. In a ClustalW alignment shows 2 regions of good similarity to the *eae* gene. Yong and colleagues have reported, based on a crystal structure of two pathotypes of *E. coli* *eae* gene indicated 4 critical residues which may be essential to identification and binding to Tir [96]. This study yielded (but unconfirmed *in vivo*) 4 critical residues in *eae* were identified, all identified in the surface exposed C-terminus [67]. If the *in silico* data and the structures are true, these 4 residues are: S890, T909, N916, and N927, which corresponds to the EaeH D1284, A1307, L1314, and E1325. These are all observed mutations via the CLUSTALw analysis. It is worth noting that EaeH is considered to be an autotransporter AIDA-I like adhesin as is Eae, however Eae is only 934 amino acids and *eaeH* composed of 1417 amino acids.

When looking at LEE negative strains of different pathotypes of *E. coli* many of them are *eaeH*+ [95]. Based on comparison to, for example, in ExPEC there is one gene; *fdeC* which has 95% identity compared to *eaeH* and it is expressed and can confer attachment to urothelial cells (based on that interaction) [97]. These facts make it seem like it is potentially involved in pathogenesis of EHEC O157:H7; but unfortunately there is little information on this gene other than its sequence (published in NCBI), and its conserved regions when compared to other intimate attachment-related genes to host cells such as Eae and FdeC [97]. This was not recognized as being expressed in the proteome provided by Kudva and colleagues, but that could be (if the information on *fdeC* is correct) is that it could potentially be because the cells were not interacting with the RSE
4.4.3 Genes Involved in Virulence and Survival/Dual Function

Our analysis of SS17’s genome revealed the presence of the *iha* adhesin gene, as in many other EHEC strains, however it did not bear an nsSNP and it is a category IIb gene. *Iha* (*Iron Regulated Gene A Homologue*) is an outer membrane adhesin which is expressed by many pathotypes of *E. coli*, and has been shown to be crucial to the virulence of certain types of *E. coli* such as EHEC and Extra-Intestinal Pathogenic (ExPEC) and is a good target for this study [98]. Considering this gene is involved in adherence and found in the proteome provided by Kudva and colleagues gave it further reason to be added to our gene list [14]. It has been shown to be upregulated in conditions where Iron (III) is limiting, thought to have come from the *Vibrio cholerae irgA* gene [56]. *Iha* can act as a dual-functioning outer-membrane protein as an adhesin and as a TonB-dependent enterobactin receptor [62,98]. It has been shown to confer adherence to Caco-2 and HeLa cells of pathogenic strains of *E. coli* as well as non-pathogenic strains that do not express this protein (i.e. K-12), which also indicates it is specific to pathogenic strains of *E. coli* [56,62].

SS17 shows ns-SNPs in a variety of iron sequestering genes such as *aroB, fhuA, fhuE, chuA*, and *chuS*. This would indicate a possible change of ability to acquire Iron (III) and an upregulation of *iha*. An upregulation in *iha* is seen in a EHEC O157:H7 strain in cattle when compared to human [99]. Because Iha is a siderophore receptor, it
will also play a role in the regulation of adhesins that might be expressed by Iron (III) directly or indirectly. For example, *crl* the genetic factor that controls synthesis of *csgA* through control of *csgD* (the Transcriptional activator of the curlin locus) is controlled by expression of *fur* which is controlled by the amount of bioavailable iron in the cell [59]. One last reason that this is a good target is that it is not LEE-regulated, which is important because Kudva and colleagues have indicated this mechanism of hyperadherence/hyperaggregation is non-LEE mediated [14].

The *eivA* gene is a predicted Non-LEE encoded Type 3 Secretion System (T3SS) channel protein that belongs to the *eiv*-family of genes that bear homology to the *S. typhimurium inv*-family, and is a category IIa [35]. The role of the canonical T3SS that is LEE-encoded has been shown to be crucial to the intimate attachment of *E. coli* O157:H7 to human epithelial cells, such as Hep2 [67]. The second T3SS has been named by many authors whom study it ETT2, for *E. coli* Type Three Secretion System 2 [34,35,100]. This secretion system has been shown to be widely distributed as a complete copy (with all of the *eiv* genes) in the O157 serotype, but incompletely in other serotypes of STEC; such as O119 or O26 (33). Interestingly, this locus has been conserved between intestinally pathogenic *E. coli* (such as EAEC and EHEC), but has undergone degenerative mutation, but is still essential for virulence, in strains like sepsis causing O78 strain 789 (30). This second ETT2 locus has been shown in K1 as being essential for intracellular invasion and survival in the endothelial cells, and its lack of conservation in K12 or other commensal strains would indicate that it has, potentially, a role in the virulence of pathogenic *E. coli* [35,36,100].
The essential nature of the ETT2 deletion in K1 had a similar phenotype with the single knockout of \( eivA \), implicating it as an essential protein in this apparatus. EivA is not found on the proteome provided by Kudva et al. and has no direct information on its ability or knockout to interfere with adherence to eukaryotic cells. Non-LEE (\( nle- \)) have been implicated in a role in modulating host immune responses through NF-\( \kappa \)B; which may play a role in the colonization and survival aspect in the cow [37]. \( eivA \) is not the only gene from this family to have an nsSNP as \( eivF \) has an nsSNP (C to a T at nucleotide position 97) encoding a H163Q change, which is a basic amino acid to polar uncharged amino acid switch. The \( eivF \) gene encodes for (by NCBI BLAST) a thermo-regulatory protein of the ETT2 and has been shown to influence the levels of LEE-encoded genes in a negative fashion. Deletion of \( eivF \) generates an enhanced LEE-mediated adherence pattern [36,100].

4.5 Recombineering and Generation of Knockouts

A critical part of the experimental process in determining the microbial factors involved in the enhanced adherence, aggregation and biofilm formation is to knockout the genes of interest. To ensure the knockouts were generated successfully, all potential clones were screened by the method described in Materials and Methods 3.3.3. We were able to successfully knockout all 9 genes of interest in SS17 and EDL933 (Figures 4-1). Our original protocol using the pSIM5 and pSIM9 plasmids were inefficient at generating knockouts, but pKD119 was able to generate them all, indicating our new protocol we designed for the recombineering was good.
We may consider either using the Wanner method to delete the KanR cassette in order to restore native promoter function and translation of downstream genes that may be affected by this setup [101]. This would ensure that the promoter of KanR is no influencing the phenotypes that we observe in the assays we run on these mutants.

4.6 Comparative Genomic Analysis on Genes of Interest Across SS Isolates

SS17’s genes were analyzed for nsSNPs and these genes of interest were chosen based on our process of selecting genes to analyze with sequencing, RSE and plant adherence outlined in the Materials and Methods. In this section, each gene selected was analyzed over 11 other SS strains to determine the relative distribution of the mutations observed in SS17 and to look for novel mutations in other strains. These 11 SS strains were chosen based to represent the different PFGE clusters presented in Figure C-1.
There are different factors that can influence a PFGE pattern. These include nsSNPs acquired at the cut site and general deletion/insertion of genetic material, and strains under similar evolutionary pressures will be similar in their genetic makeup, and those that aren’t will differ. In light of this, we wanted to see if any of these mutations are conserved across the SS isolates, or if different isolates belonging to different lineages, clusters, or origins could plausibly use different factor(s) to generate these unique phenotypes observed.

The 7 genes selected that contained nsSNPs were sequenced around the location of the nsSNP to confirm in SS17 and look in the other 11 super shedders. *iha* and *ompA* were primer walked over the whole gene to determine any nsSNPs in other isolates belonging to different lineages or PFGE clusters. The nsSNPs are reported in Table 4-1 as they are read in their codons (5’→3’) so reverse strand genes are read in the reverse complement. Genes which have an nsSNP could have a change or could have no change, but if a change is observed in all SS isolates, it would be more likely to be involved. As Table 4-1 shows, none of the genes that were chosen had nsSNPs found in SS17 were conserved across all of the SS isolates, but did seem to be conserved with relative closeness of PFGE patterns and lineages.

Primer walking sequencing of other SS isolates revealed the conservation of the *iha* amino acid sequence, despite 3 strains (SS12, SS131, and RM11333) having a synonymous SNP compared to the rest of the isolates (Table 4-1). The only role a synonymous SNP could play in affecting protein is through codon bias, but at this time it is not known if this is an issue between the CTG and CTA codons for Leucine at nt position 1470. A deletion in *iha* is not detrimental to overall adherence (just adheres ‘less
well’), and so the phenotype is not that drastic (though it was not a clean deletion) [56]. This could indicate an enhancement or redundancy of adherence with a partner that is unknown, but in and of itself is a reason why iha may not be our target. All the information and data considered suggested to us that it would make a good target however, and so we went forward and made the knockout of iha in both EDL933 and SS17. All of the SS isolates had the same sequence identified in EDL933, also indicating that if ompA is involved in any way it would be most likely through a mechanism of regulation and not change in function (Table 4-1).

_csgG_ has a ns-SNP that converts a C to an A at nucleotide position 226, that yields the amino acid change P76T (Table 2-3 and Table 4-1). This was only conserved in SS52 and SS67. Both are Lineage I/II, but so are SS1 and SS7 which do not share this nucleotide and resulting amino acid change. Interestingly, SS17, 52, and 67 all are part of the same branch point from the 3^rd major branch point based on PFGE clusters shown on Figure C-1, and SS1 and SS7 are separated. It is possible that this conservation may play a role in this PFGE-related cluster.

As Table 2-3 and Table 4-1 show, _wzzB_ has an n-SNP which results in a C to a T nucleotide change at position 403, resulting in the missense mutation P136S. Where this is not in an indicated critical region (which was only identified in _Salmonella_ and _Shigella_), it lies in between the 3 critical regions identified (67-95, 200-255, and 269-274 amino acids), and would be potentially facing the inner pocket [94]. Sequencing _wzzB_ in the region of the nsSNP revealed the P136S has been confirmed in SS17 and again conserved in SS52. The other 10 SS isolates all had Serine at position 136 shown in Table 4-1.
A gene coding for \textit{cah} has been identified in SS17, and it has a truncation of approximately 850 nucleotides which spans the C-terminus of Cah and the N-terminal end of the downstream neighboring gene. This deletion results in an in-frame shift of 3 new amino acids and a ‘premature’ stop codon 832W, 834A, 835R, and 836Opal all occur in the \( \beta \)-domain as shown in Table 4-1. This truncation observed in SS17 has been confirmed and is unique to SS17. After aligning to all the other SS strains and EDL933, this appears to be a truncation at the C-terminal end of the \textit{Cah} downstream which landed 3 random amino acids with a stop codon into the coding region. This is a very interesting mutation and because it is such a big difference in protein structure of an autotransporter, and if it was the gene of causing the change in phenotype then it would be conserved. This mutation was tested because of its importance in adherence and aggregation on different surfaces (and biofilm formation) and the possibility that the SS strains have use different receptors/adhesins in this phenotype.

Analysis of SS17’s \textit{eaeH} revealed a C-terminal nsSNP resulting in a C to an A nucleotide at position 3902 and thus changing the amino acid T1301K amino acid change (Table 4-1 and 2-3 respectively). This nsSNP and the consequential amino acid change is worth noting in that, based on blocks of alignment, this nsSNP is in the middle of the block of identity aligned to \textit{eae} that contains these 4 critical residues. At the same relative position based on this alignment, the residue in \textit{eae} is R903. Though Arginine and Lysine are basic, positive charged amino acids, the change in R group structure and its relative location could be influential in changing protein folding and ultimately function. The gene \textit{eaeH} was sequenced and was the only gene that had a conservation of the nsSNP across different lineages. The T1031K amino acid change is observed in
SS17 is conserved in SS1, 7, 27, 52, 67, and 77 as shown in Table 4-1. This nsSNP is the most conserved of all the nsSNPs observed in the nine genes of interest. This is very enticing as it is the only conserved nsSNP across lineages and PFGE clusters, but at the same time was not conserved in SS131 or RM11326/11333. As with other genes, there always lies the potential of another nsSNP in the gene or in an accessory gene.

As shown in Table 4-1 \textit{eivA} also has a C:T nsSNP which correlates to the P33S (Table 2-3) that is consistent when compared to all references. Proline changes are interesting because their pentagon-like structure causes them to have unique interactions especially in $\alpha$-Helicies as they can cause a ‘kink’ in the chain. Proline is also a non-polar amino acid which has been substituted for a serine which is a polar uncharged amino acid, either or both of which can effect protein folding in a significant way. \textit{eivA} could be extremely interesting in that much of the non-LEE T3SS and its effectors are poorly understood, but do have involvement in intracellular signaling as well as host immune signaling. The P33S encoding nsSNP was confirmed in SS17 but it was unique to SS17 as seen in Table 4-1.

A review of the sequenced SS17 genome showed a gene that code for \textit{yfaL}. Table 4-1 shows that \textit{yfaL} has 3 SNPs (C to a T, G to an A, and C to a T at positions 67, 112, and 311 respectively) resulting in the amino acid changes S23P/D38N/T104I; one nsSNP in the signal peptide region, and two in the passenger region which are responsible for the binding of autotransporters as the translocated portion of the protein. Though the essential amino acids for this protein have yet to be worked out, these three portions of the protein play 3 very critical roles in its function [64,65,68]. The signal peptide allows for correct localization of the protein, the passenger region in many AIDA-I like proteins
is responsible for being cleaved and re-folded into extracellular effectors, and the β-Barrel transporter is essential for the passage of all of the elements used in the T5SS.

Table 4-1: Conservation of Mutations in Genes Selected for Knockout

<table>
<thead>
<tr>
<th>NT Position</th>
<th>iha&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ompA&lt;sup&gt;b&lt;/sup&gt;</th>
<th>csgG</th>
<th>cah&lt;sup&gt;c&lt;/sup&gt;</th>
<th>wzzB</th>
<th>eaeH</th>
<th>eivA</th>
<th>yfaL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1470</td>
<td>-</td>
<td>-</td>
<td>226</td>
<td>2502</td>
<td>403</td>
<td>3902</td>
<td>97</td>
<td>67</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Position</th>
<th>EDL933</th>
<th>SS1</th>
<th>SS7</th>
<th>SS12</th>
<th>SS17</th>
<th>SS27</th>
<th>SS42</th>
<th>SS52</th>
<th>SS67</th>
<th>SS77</th>
<th>SS131</th>
<th>RM11326</th>
<th>RM11333</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
<td>CTG</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
<td>CCG</td>
</tr>
<tr>
<td></td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
<td>FL</td>
</tr>
<tr>
<td></td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
<td>CCT</td>
</tr>
<tr>
<td></td>
<td>ACA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
<td>CCC</td>
</tr>
<tr>
<td></td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
<td>TCT</td>
</tr>
<tr>
<td></td>
<td>GAT</td>
<td>AAA</td>
<td>AAA</td>
<td>TCC</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
<td>AAA</td>
</tr>
<tr>
<td></td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
<td>ACC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eight genes of interest based on the criteria set forth in the materials and methods were sequenced around previously identified nsSNPs and the conservation of them across the different SS isolates are reported. The blue highlight indicates conservation of the nsSNP with SS17, with the Red letter being the nt change. Those with an orange letter and highlighted green are synonymous SNPs.

a. iha did not have a reported nsSNP in SS17, but the CTA is a synonymous SNP with the CTG found in others.
b. ompA did not previously have a reported nsSNP in SS17, so - indicates sequence of EDL933.
c. cah has a large deletion spanning ~850 nt portion of the C-terminal end of the gene represented by the Del. FL stands for full length for strains that the truncation is not conserved.

In situations where the nsSNP is not carried in the same location in the coding region, the nsSNP could be elsewhere in the gene and this would need to be further analyzed. It is also recognized that if the gene is involved in a pathway or is a chaperone gene, another gene in that pathway/accessory protein could have an nsSNP yielding the same phenotype. For example, eivA has an nsSNP, but if it is found to be not involved in this phenotype (as the nsSNP is not conserved), another gene in the eiv- family (eivF) also has an nsSNP, and could be involved through or individually in the context of eivA.
Finally, it is understood that nsSNPs in SS17 that are conserved in other SS isolates, but yield no change in phenotypes strongly suggest that change to the protein is not consequential individually for the change in phenotype for SS isolates that bear that SNP; but will need to be tested in all strains that bear that nsSNP against their genetic background. If the gene does not bear an nsSNP (compared to EDL933), it most likely does not play a significant role in the change in phenotype observed in SS17 (as EDL933 has its own particular phenotype not shared with SS isolates). The SS17 and EDL933 parents and mutants were also analyzed for biofilm formation based on the original observation that EHEC O157:H7 create extensive biofilms induced by damaged leafy green.

These SNPs were then aligned using the MEGA program to generate a phylogenetic tree of the isolates used in this study with a Maximum nucleotide based likeness analysis (Figure 4-5). Coincidentally, the nsSNPs found in this subset of genes closely clustered these isolates as they are clustered by PFGE pattern (Figure C-1) and by lineage. This gives us a view of the evolutionary divergence of these SS isolates with EDL933 from one another on a small scale. If all 102 isolates were done, the common ancestor might be more apparent, but it is interesting how different tests for genetic similarities group bacteria almost in the same manner, giving support to those methods in what they mean.

Moving forward, the different characteristics exhibited by the various SS isolates may give credence to conservation of mechanism or differences in mechanisms; and this may help in identifying these (by seeing conservation of phenotypes based on clusters). If any of these genes are found to be prudent to any of the aforementioned phenotypes,
Figure 4-5: Phylogenetic Analysis of SS isolates by nsSNPs in Genes of Interest
Phylogenetic analysis of SS isolates based on their nucleotides in genes with SNPs and deletion is shown above. A dendrogram was generated by using Maximum Likelihood method of MEGA 5.0 program with Tamura-Nei model of nucleotide substitution. The evolutionary distances were computed using maximum composite likelihood method. The result of the bootstrap analysis (1000 replicates) are represented as percentages and marked on each branch. There were total 27 sites (9 codons) were included in the final dataset. Each SS isolate was aligned (right) to the particular SNP found in each gene indicated above that column. The +/- for cah is indicative of the truncation observed in SS17.

this might also help in the development of a new set of Multi-virulent Locus SNP Typing (MVLST) different EHEC O157:H7 isolates, linking lineages and phenotypes based on genotypes. For example, looking for these nsSNPs can help identify an SS isolate with high biofilm formation and enhanced attachment, if there is truly separation of different phenotypes from SS17 (which has enhanced phenotypes for all three).
4.7 Characterization of the SS17 Mutants’ Phenotypes

As previously shown, SS17 has many unique phenotypes associated with different steps in the process of colonization of the reservoir, transmission to the environment, and better adherence and survival strategies outside of the cattle on leafy greens. To start to analyze the role of these different processes and the genes that are possibly involved SS17 and all of the mutants created through recombineering were assayed using both the RSE adherence assay and the biofilm formation assay (induced by leafy green extract).

4.7.1 Role of Genes of Interest in RSE Adherence Assay

All of the mutants in SS17 and EDL933 were tested for their adherence on RSE cells. The expected phenotype of a gene involved is that there will be some sort of change in the adherence and/or the adherence pattern of the two different strains vs. the parental strains. When these mutants were tested in the SS17 background, none of them yielded a significant difference in phenotype, all of them exhibiting a strong aggregative phenotype as previously seen. This would indicate that these genes are not essential for the adherence or the adherence pattern unique to these strains with the RSE cells. There were some that showed partial phenotypes as Figure 4-6 shows, but it is not significantly different from the parent. The biggest effects that were seen in SS17 mutants were with the SS17ΔeaeH mutant, which all RSE cells sampled were bound by at least 10 SS17 cells, and SS17ΔompA, which had a modest decrease quantitatively of cells bound to RSE cells (~87%).
The EDL933 mutants are actually the mutants that could potentially shed more light on the matter. As shown in Figure 4-7, three genes that have nsSNPs in SS17 compared to EDL933, and one that does not show a dramatic shift in phenotype observed in EDL933. EDL933ΔompA, (the gene without an nsSNP) has the most significant of these phenotype switches, seeing the mean bacteria per RSE cells of greater than 10 change from 40% to 100%. This was also correlative to the qualitative adherence pattern observed on the RSE cells. EDL933ΔwzzB, EDL933ΔeivA, and EDL933ΔyfaL, all show the same sort of switch but to varying degrees. The EDL933Δcah mutant has a very interesting phenotype as it seems to do the opposite of the previously mentioned other genes; in that it appears to switch almost all of the bacteria to have the diffuse, moderate binding with almost 100% of RSE cells having less than 10 bacteria per cell, but this ratio has been seen before. EDL933Δcah shows a complete opposite phenotype in that it shows greater than 90% of RSE cells have less than 10 bacteria per cell from 40% having less than 10 bacteria per cell in the parent.

Together, all of these data show that OmpA, WzzB, EivA, and YfaL are all critical in maintaining the standard cell interaction with RSE cells in the context of EDL933. It appears that OmpA, despite not having an nsSNP, could be involved in this phenotype but in an inhibitory way or be involved with a protein that interacts that, when its interaction with OmpA is negated, something about its expression and/or function changes. The same could be said about WzzB, EivA, and YfaL. All 3 have nsSNPs in SS17, and if they act in an inhibitory manner in the wild type, the nsSNPs observed could change their function in a way that disallows for that inhibitory process; and when they are deleted, the protein(s) involved haven’t changed and with the deletion you see no
change in SS17, but you do see a change in EDL933 mutants. On the other hand, full
length Cah appears to decrease the ability of EDL933 to adhere to RSE cells indicating it
may be involved in the general adherence mechanism of EDL933 to RSE cells. It is
important to note that there still lies the possibility that any number of these genes (all
being in the outer membrane) could also perturb the membrane such that other adhesins
have access and/or a change in folding because of the changing dynamics in the
membrane. To test if these are involved and truly inhibitory, complement studies
coupled with overexpression of these proteins in the SS17 isolate would be good tests.

The nsSNPs that are conserved in other strains will have to be tested separately
(as the genetic background of the SS isolates could be different, and so can’t other genes
that might play a role). However, based on these results it would not be very surprising if
the same results weren’t yielded in other strains bearing these nsSNPs. Just because the
rest of the genes appear to be not critical for these phenotypes (and the nsSNPs for the
change in phenotype shown in SS17), does not mean they are not involved or can be
ruled out as a part of the mechanism. There could very well be overlap with many of
these genes with other genes that they associate with that bear nsSNPs (either known or
unknown function) that can lead to a complement of the phenotype. The other prospect is
that these phenotypes are complex with a multitude of factors, some bearing nsSNPs and
others not which could also yield this phenotype.
Figure 4-6: RSE Adherence of the SS17 Mutants
SS17 mutants and their phenotypes were assayed with the RSE Adherence Assay. Panel A. Quantitative Analysis of the RSE Adherence Assay of the various EDL933 deletion mutants with means of percentage of RSE cells with either <10< bacteria per cell. Error bars are representative of standard error. Panel B. Single fluorescence microscopy slide of RSE cells with EDL933 and mutants in the following: 1. WT EDL933, 2. Δiha, 3. ΔompA, 4. ΔcsgG, 5. ΔwzzB, 6. Δcah, 7. ΔeaeH, 8. ΔeivA, 9. ΔyfaL, 10. Δeae.
Figure 4-7: RSE Adherence of the EDL933 Mutants

EDL933 mutants and their phenotypes were assayed with the RSE Adherence Assay. Panel A. Quantitative Analysis of the RSE Adherence Assay of the various EDL933 deletion mutants with means of percentage of RSE cells with either <10< bacteria per cell. Error bars are representative of standard error. The red-enclosed bars are those that have a change in phenotype from moderate aggregative to strong aggregative.

Panel B. Single fluorescence microscopy slide of RSE cells with EDL933 and mutants in the following: 1. WT EDL933, 2. Δiha, 3. ΔompA, 4. ΔcsgG, 5. ΔwzzB, 6. Δcach, 7. ΔeaeH, 8. ΔeivA, 9. ΔyfaL.
4.7.2 Biofilm Production Induced by Damaged Leafy Greens in SS17 Mutants

Biofilm formation on a variety of surfaces has been implicated in the survival and various other aspects of the life-cycle of many bacteria, including EHEC O157:H7. As shown in Chapter 4.3, SS17 generates a significantly greater amount of biofilm than its EDL933 counterpart in the biofilm assay induced by leafy green extract. Many of the mutants showed no significant difference in biofilm formation when compared to the parent SS17 strain. These included iha, csgG, wzzB, eaeH, and eivA. For the genes that did not show a difference in phenotype, it is important to remember that this phenotype is induced by spinach lysate and they could play a role in biofilm formation in other conditions or in vivo.

A test of the SS17Δiha showed no deficiency in forming biofilms compared to the parent SS17 as shown in Figure 4-8. This result indicates that Iha probably plays no major role in biofilm formation in the SS17’s biofilm formation induced by a damaged leafy green. This was not surprising as Iha has had no previous data supporting it as being important for biofilm formation. Because Iron is an essential element to maintain biofilm stability, it could have had an effect with its function as an enterobactin receptor. However, there seems to be much functional redundancy in that mechanism to prevent that from showing a phenotype.

Despite much literature on the importance of curli expression on biofilm formation, SS17ΔcsgG resulted in no significant decrease in biofilm formation[52]. This
could mean either \(\text{csgG}\) is not essential for biofilm formation in SS17 in the conditions they were grown in, or perhaps there is a second mechanism that allows for the expression of the curlin fibers despite the deletion of \(\text{csgG}\). The role of \(\text{csgG}\) in biofilm formation could be dedicated to \textit{in vivo} or in a different condition. SS17\(\Delta\text{wzzB}\) also has no deficiency in biofilm formation compared to the parent SS17 strain as shown in Figure 4-8. It would be reasonable to surmise that \(\text{wzzB}\) (and its observed change) does not play a significant role in the enhanced biofilm formation, and probably not involved in biofilm formation of EHEC O157:H7 in general.

The \textit{eaeH} gene was further studied in its role in biofilm formation. SS17\(\Delta\text{eaeH}\) has no data suggesting it is important for biofilm formation and \textit{in vitro} as it shows no deficiency in forming a biofilm (Figure 4-8). This suggests that \textit{eaeH} is not involved in biofilm formation in EHEC O157:H7 and as a result the observed nsSNP for \textit{eaeH} in SS17 is not involved in the enhanced biofilm formation seen in SS17. There is no information on \textit{eivA} or any of the ETT2 genes involved in biofilm formation but due to its involvement in LEE signaling, it was decided that it could be an interesting gene to follow up on in the biofilm formation. As Figure 4-8 shows, SS17\(\Delta\text{eivA}\) does not show any deficiency in forming biofilms which rules it out as being important for the formation of biofilm in SS strains (and EHEC O157:H7), but also that the nsSNP is non-consequential for this change.

As Figure 4-8 shows, SS17 \(\Delta\text{ompA}\) is extremely deficient in forming biofilms. The expression of OmpA has been linked to inducing a biofilm state, but in EHEC O157:H7 has never been shown to be critical for biofilm formation. There is no significant difference between the \(\text{ompA}\) deletion in SS17 and EDL933. The data
coupled with the lack of an observed SNP across all EHEC O157:H7 isolates sequenced, also strongly suggest that it is critical for the general mechanism of damaged leafy green induced biofilm formation in EHEC O157:H7. Given the nature of the importance biofilms play in the lifecycle of many bacteria, this could be extremely interesting to further study in the context of survival and pathogenesis from leafy green contamination. If OmpA is truly essential for biofilm formation in EHEC O157:H7 (general mechanism), it can only be determined after the EDL933 mutants have been run. Despite not having the EDL933 results, it is still clear that OmpA is at the very least critical for the enhanced (and appears to be) and general molecular mechanism for biofilm formation in damaged leafy greens for SS17.

Previous studies have identified that ompA may be involved in biofilm formation through regulation of biofilm required genes (such as fimbriae), but also because it is upregulated in biofilm states [86,102]. In EHEC O157:H7, OmpA has also been shown to be important for a variety of processes. To date it has never been shown in EHEC O157:H7 that OmpA is critical or essential for biofilm formation induced by damaged leafy greens (simulating factors that SS17 or other EHEC O157:H7 would be exposed to on a leafy), and after the EDL933 results have been analyzed this could be the beginning of a very interesting study. To our knowledge this is first study directly showing OmpA as a critical gene in biofilm formation in SS17 and could become the first to show that it is important for any EHEC O157:H7 biofilm formation induced by leafy green extract.

Despite the major truncation to the C-terminal end of Cah, it still functions in its capacity for biofilm formation in SS17, and the deletion could possibly play a role in the enhanced biofilm formation in SS17. Figure 4-8 show that SS17Δcah is deficient in its
ability to generate the enhanced biofilm produced in the SS17 parent. A Pfam search of 
the Cah that is encoded for by SS17 appears to be missing a large portion of an important 
activity domain located in the barrel region of the protein. Despite the loss of this 
domain it appears that it is important for enhanced biofilm formation in SS17 (as it has a 
change and SS17 has enhanced biofilm formation). Whether this loss of enhanced 
biofilm production is based on the mutation observed in Cah or just because it is involved 
in general biofilm production in these conditions has yet to be determined (because the 
EDL933 control mutant has not been run). The reduction in biofilm formation still is at 
the level of EDL933, further reasoning that the deletion in cah is contributing to the 
phenotype observed in SS17.

SS17ΔyfaL shows a deficiency in generating biofilms compared to the parent 
SS17 strain. In order to determine if this phenotype is associated with the nsSNPs in yfaL 
important for the formation of biofilms or just involved in the difference in biofilm 
formation in SS17 we will need to run the EDL933 mutant. The loss in biofilm 
formation in the SS17ΔyfaL mutant is approximately the same ratio and relative level of 
EDL933. This result would suggest that the nsSNPs in yfaL may be playing a role in the 
enhanced biofilm formation in SS17 either in concert with, or separately, other changes 
in the genome (e.g. the deletion in cah). It is clear though that YfaL plays a role in 
biofilm formation, which is interesting in that the current literature on YfaL implicates its 
expression can induce a biofilm state, but no data is currently present implicating as it 
being important for biofilm formation, which this data suggest. The three nsSNPs found 
in yfaL could be playing a role in this phenotype; and given that SS17 does not
completely lose the ability to generate biofilm in this mutant suggests further that it may be involved in the enhanced biofilm phenotype attributed to SS17.

The biofilm formation induced by damaged leafy greens also makes another important point. Healthy leaf tissue on spinach or otherwise might not be as permitting of biofilm formation as damaged leaves. That said, it is completely possible that the genes important for attachment to leafy greens (as in Figure 4-2), might not be the same genes important for biofilm formation or even attachment to damaged leaves. If the autotransporters involved in this biofilm formation are important for biofilm formation induced by the leafy green extract, it is probable that they may not be important for binding healthy greens, because the extract will contain factors associated with the

![Figure 4-8: SS17ΔompA, Δcah, ΔyfaL Show Decrease in Biofilm Formation](image)

All mutants of SS17 were assayed for their ability to form the enhanced biofilm formation exhibited by SS17. The three mutants (ΔompA, cah, and yfaL) which exhibited decreased biofilm formation were tested a second time along with eaeH, and eivA. To normalize the data, the mean OD570 of each mutant and EDL933 was compared in ratio to SS17, and the ratio was reported. * Indicates significant difference and error bars are of the standard error.
damaged leaves because the extract is made by damaging leaves.

The EDL933 control is a critical control for this experiment (as it is for all of the experiments performed). As the considered non-SS isolate, it will help determine the role of genes (in this case) and their nsSNPs (if applicable) in the phenotypes observed in the SS isolates. In this case, it would be expected that EDL933 would also see a proportional loss in biofilm formation with a deletion of a gene that is critical for the general mechanism of biofilm formation in EHEC O157:H7. If the loss is negligible or not proportional it could mean that gene is more important for the enhanced phenotype exhibited by SS17, and not about the general mechanism of biofilm formation and the genetic change could be responsible.

For all the mutants that showed a difference in phenotype (including EDL933 because it is so much lower than SS17 in biofilm production), growth curves were run with cell counting to determine if there was a difference in growth that could explain these results (outside of importance to the phenotype). In Figure 4-9, the growth curve data show that there was no significant difference in growth rate that would indicate a growth difference. This gives more credence that the phenotypes exhibited by the deletion mutants were in fact due to their involvement in the formation of biofilm induced by leafy green extract.
Figure 4-9: SS17 Deletion Mutants Show no Difference in Growth in LB
SS17’s deletion mutants that showed differences in phenotype in the biofilm production assay from spinach extract show no significant differences in growth curves. Cell Counts can be found in the Appendix.
In order to determine further confirm no significant difference in growth of these different strains, doubling times were determined for all isolates used in the growth curve above. The doubling time will aid in ensuring the data respective to the growth curve is correct. As Figure 4-10 shows, the doubling time of SS17 is not significantly different compared to the SS17 mutants. All SS17 mutants and SS17 were slightly slower in doubling time compared to MG1655, but were slightly faster than EDL933. Taken together, there was no significant difference in growth that would explain the difference in biofilm formation, meaning the data collected in that assay were truly caused by the mutations introduced into those strains and not a difference in growth.
Chapter 5

Conclusions and Future Directions

Despite many efforts, EHEC O157:H7 is still a major contributor to outbreaks of human disease resulting in significant morbidity and mortality, despite the number of cases decreasing. Control methods have worked, but have not produced desired results and hence a more robust way to control the environmental load of EHEC O157:H7 is important. The recently observed phenomenon of super shedding could potentially shed light on why prevention measures have failed to reduce the number of incidences with EHEC O157:H7. It is clear that the presence of even just a single super shedding cattle results in significantly higher environmental load of EHEC O157:H7, and by doing so increase the risk to human health and contamination of various goods from farms (such as cattle beef and leafy greens) [9,13].

To better understand the relationship of all EHEC O157:H7 to its reservoirs would help us better develop control methods directed against most of the EHEC O157:H7 that colonize cattle. However, given its prevalence and significant changes in the dynamic in the epidemiology of EHEC O157:H7, it appears that being able to control the SS phenotype by better understanding it at the molecular level also seem just as important, and could be the other or even the main key to preventing the transmission, outbreaks, and illness this organism causes.
The recto-anal junction (RAJ) has been previously identified as the site of colonization of EHEC O157:H7 and has been implicated as being extremely important for the ability for the bacteria to be shed and transmitted from the cow [7,15]. We believe there are 3 major contributing factors to the super shedding phenotype: the host, the environment (host/microbe) and the microbe; with the microbial factors being the easiest to possibly elucidate. By understanding and targeting the microbial factors we can garner a better understanding of how these genes and their products work as well as not have to target any factors related to the cattle (and worrying about side effects).

The RAJ is made up of two cell types, the FAE (or follicle associated columnar epithelia) and the SSE/RSE (or stratified squamous epithelia) [7]. Previous studies have shown that the interaction of EHEC O157:H7 with FAE cells is mediated through the LEE operon, but the interaction with the RSE cells is through a LEE-independent mechanism. We hypothesized that there would be a change in the molecular interaction between the cells of the RAJ and the SS O157:H7 when compared to a non-SS EHEC O157:H7, and found that SS EHEC O157:H7 had a distinctive strong and hyper aggregative phenotype on RSE cells when compared to the EDL933 reference strain not thought to be a SS. This phenotype, as is with the non-SS EHEC O157:H7 strains, is through a LEE-independent mechanism. The result seems to also indicate that it is this interaction (the RSE cells with the SS isolates and not the FAE cells) which may be critical for the SS phenotype observed as colonization and cellular interaction at the RAJ are critical for transmission.

Based on LSPA analysis and genetic characterization of SS17 grouped with other previously identified EHEC O157:H7 strains involved in outbreaks, we hypothesized that
this enhanced adherence may carry over to the SS isolates interactions on leafy greens (a major source of outbreak contamination). We found that these strains did in fact have enhanced adherence to the baby spinach leaves they were tested on. Further analysis on leafy green-environmental production of biofilm formation indicated the SS strain SS17 produced more biofilm than EDL933. These data indicate a mutual relationship between adherence on leafy greens and biofilm formation (with potentially changes of adherence mechanisms at the RAJ); and all together would indicate a series of changes (or perhaps changes of redundant functional nature) that have an overall mechanism important for adherence, colonization and transmission of this organism based on changes to genes involved in these critical processes.

When a change in phenotype is observed in biology there usually is a change in the genotype of that organism to explain the change in phenotype. This can be done either through horizontal gene transfer (the acquisition of genes through transformation, transduction, or conjugation), the loss of genes (degenerate evolution), or through the acquirement of non-synonymous single nucleotide polymorphisms (nsSNPs) which will result in missense mutations in coding regions or changes in nucleotide composition of promoters and promoter elements. Any of these changes can change the structure, function, regulation or any other number of aspects a particular protein is characterized by and/or responsible for.

A full sequencing and annotation of SS17 revealed 295 virulence associated genes (Appendix Tables A1-A3) that have been previously identified in other EHEC O157:H7 strains. Within these genes, we hypothesized, would be nsSNP(s) or novel genes that could potentially shed light on this new phenotype in both the SS phenotype, the unique
cellular adherence and biofilm phenotypes. Our analysis of the genome revealed no
acquisition or substantial loss of coding sequences or novel genes, but did show many
nsSNPs when compared to the 4 previously sequenced reference strains, but the most
comparing to EDL933. Compared to previous references we identified 295 virulence
associated genes, but of particular interest were 12, 19, and 8 genes in toxin, adherence,
and virulence-associated genes respectively which bore nsSNPs and could all be
potentially important for the adherence phenotype observed with the RSE cells and could
also be potentially important for the SS phenotype (either overlapping or separately).

There were other genes which have significant roles in the regulation of virulence
genes that had coding nsSNPs and these could potentially be important down the road.
Examples of this are the mutations in *fepE* and *fhuA* which could impact the acquisition
of iron in non-pathogenic states and could thus impact the expression of *hly*- and *iha*.
Finally there were a few genes which had SNPs in non-coding regions upstream of the
translational start site which could potentially be important via the regulation of genes of
interest that could be involved in this phenotype.

It is very interesting indeed that there appears to be (based on phenotype) a shared
set of changes in the genome amongst the SS isolates but couldn’t be previously
identified through the other characterizations by Arthur and colleagues through gross-
genome analyses (PFGE typing, phage typing, etc…). To see if any of these nsSNPs are
shared, we decided to hunt for the nsSNPs in strains that spanned the PFGE clusters
provided by Arthur *et al.* (Figure C-1), to get a representative look at the nsSNPs that are
conserved and if so how they are conserved. This will be used to help gauge which genes
we have selected have conservation of nsSNPs, but doesn’t rule out secondary mutations
or accessory mutations which might not be picked up on this screen (as previously mentioned). This also may shed some light if there is a common mechanism or through divergent evolution of the isolates, are the genetically close isolates different in there manner of the SS phenotype (and associated in vitro phenotypes).

In addition to SS17, another 11 isolates were chosen representing different lineages and different origins; two factors previously implicated as separating different EHEC O157:H7 ability to adhere to and colonize cattle and humans. These were SS1, SS7, SS12, SS27, SS42, SS52, SS67, SS77, SS131, RM11326, and RM11333. This was done to provide insight into the genes selected in their conservation of nsSNPs or if novel and separately conserved (based on lineage, PFGE, origin, etc…) would show the conservation of a mechanism and associated factors and/or if different SS strains had evidence that they use separate mechanisms and factors to generate these phenotypes.

It has been shown previously that the origin, lineage and other factors contribute significantly to various aspects of the life of EHEC O157:H7 strains, and this makes sense as different environmental pressures and exposures are going to generate different selective pressures and mutations, thus creating the differences seen in EHEC O157:H7 and presumably the SS strains (as shown by the diversity by Arthur and colleagues) [9,20]. We believe that the similarities in the phenotypes would be through a conserved mechanism, but we are aware it is very plausible that these could be different mechanisms. Unfortunately none of the adhesins we chose showed conservation of any nsSNP or genetic change across the different lineages and the different clusters. Using the different nsSNPs to generate a phylogenetic tree showed that, based on this particular set of genes, strains clustered into branches and clades based on lineage; indicating that
the genotypes might be different and leaves the probability that different SS isolates use different factors and or mechanisms in these unique phenotypes.

It appears that all of the genes that were chosen and tested are not essential for the unique adherence of EHEC O157:H7 SS17 isolates to RSE cells. Because we cannot rule out the possibility that these phenotypes observed are complex and/or have redundancy, we cannot say at this juncture if they are involved; only not essential.

Despite seeing no lack of adherence in SS17’s mutants, there were interesting observations made about the mutants in EDL933. Based on these data, it appears that OmpA, YfaL, EivA, and WzzB all play a role in this phenotype; as their deletions allowed for EDL933 to behave similar to the wild-type SS17’s phenotype on RSE cells. It appears that these genes may play a role in maintaining the interaction of RSE cells with EHEC O157:H7 similar to that of a non-SS isolate (standard adherence), as their deletion allows for the strong aggregative phenotype to manifest in EDL933. If this is the case the results of SS17’s phenotypes with the mutants make a little more sense, if you delete an inhibitory gene you will not see loss of the adherence phenotype. Cah on the other hand, resulted in a completely opposite phenotype, resulting in a decreased adherence pattern in EDL933; indicating its role as being important for the adherence of EDL933 to RSE cells.

Interestingly, ompA did not have an nsSNP, but the phenotype switch in EDL933 was the most significant in the ompA mutant. Based on this it would appear that OmpA is involved, and because it doesn’t have an nsSNP, it may be due to a protein it interacts with (and the corresponding gene may have an nsSNP or another change. The other three genes all have nsSNPs, and it is possible that the nsSNPs they have is causing their
activity or regulation to change and thus influencing the phenotype that is observed in SS17. It is possible that there may be a gene or set of genes which may interact with some or maybe all of these genes that shows a genetic change in SS17, resulting in this gene switch that appears in the EDL933 mutants. The deletion in Cah is very interesting as it may give us an initial target for preventing colonization by non-SS isolates, as these can still infect cattle and infect humans and preventing their colonization and spread is also important.

The only caveat is that all the genes that have nsSNPs in this case are not conserved in other SS isolates that have the nsSNP conserved; and so either different SS isolates have different mechanisms or these genes are involved in and another set of changes mediates the same mechanism. Looking at these genes may point to other genes of interest through the genes they interact with and possibly regulate (through interactions). At this juncture we are unsure if the mechanisms at play are shared or not; only that it appears where the nsSNPs and/or regulation in genes that have been considered are not conserved thus far appear not to be involved in these phenotypes. There are still 31 other genes of interest with coding nsSNPs, and these do not even include those with non-coding SNPs.

SS17, its mutants, and EDL933 were grown in spinach extract to show the biofilm induction that would occur in the presence of damaged spinach leaf factors. Many of the genes tested were found not to be critical for the formation of biofilm or the enhanced biofilm of SS17. As was with the RSE assay, we cannot say that the genes that did not show a phenotype are not involved in this phenotype; but they are not critical for it in the context of leafy greens. In SS17, the enhanced biofilm formation observed is affected by
the deletion of *cah*, *ompA*, and *yfaL*, indicating that these genes are at the very least

critical for the enhanced biofilm production of SS17 in these conditions.

Many studies with EHEC O157:H7 or other strains of *E. coli* (such as K12 and

UPEC), have all looked at biofilm formation on plastic surfaces to understand general

mechanisms; and most are grown in rich media or in salt media. Thusly, we have, for the

first time, shown that Cah, OmpA and YfaL are critical for SS17 biofilm formation in

response to damaged leafy greens and that OmpA and YfaL directly play a role in biofilm

formation in EHEC O157:H7 in the same context. Because they are involved in biofilm

formation on spinach, it can be reasoned that they may also be involved in the adherence

phenotypes observed on spinach; in a sensory kind of adherence (they attach and that

attachment may lead to biofilm formation); and thus should be tested.

When the EDL933 mutants have been tested, we can make better conclusions

about the true nature of *cah* and *yfaL* in the context of the consequence of the nsSNPs and

the genes themselves about their relationship with biofilm formation induced by leafy

green extract. By testing other SS isolates’ ability to form biofilms (given the lack of

conservation of these genetic changes), could further strengthen these studies. What is

clear is that all of these are all autotransporters (or Type V secretion systems), and they

all play a role in biofilm formation (whether enhanced biofilm or the general

mechanism). Also, understanding how and why these autotransporters are so critical for

biofilm formation is another interesting question. Is there a signaling cascade caused by

the different interactions these autotransporters encounter, and if so what are the factors

that contribute to the genetic switch to activate biofilm genes? And of course, what are
the consequences of that in the various aspects of the cycle in EHEC O157:H7 and what if any in the context of the SS phenotype.

Even before this, there are interesting studies laid out in the natural nsSNPs and alterations to \textit{yfaL} and \textit{Cah} which might help us understand the role of autotransporters and their role in biofilm formation. For example, the truncation in \textit{Cah} would lead us to believe that this section of the C-terminal end of the protein is not-essential for biofilm formation functions of this protein. These biofilms were induced by leafy green extract, though I think it would be interesting to see if the same genes are important for biofilm formation on other surfaces such as large intestine and RAJ. There are many other genes of interest that could play a role in biofilm formation with coding and non-coding nsSNPs such as \textit{luxS}, \textit{wcaL}, and \textit{wcaM}.

In the process of all of these assays we are starting to get a glimpse of what appears to be a very sophisticated branch of the cycle of evolution of EHEC O157:H7. One isolate has shown to have a hyperadherent phenotype to the RSE cells of the RAJ ensuring a second point of colonization and potentially shedding from the cow. They have shown enhanced adherence to leafy greens, which is another major source of contamination and human infection. To couple with their leafy green adherence, when in the presence of spinach lysate, the SS17 produces a lot of biofilm comparatively, which will aid in its survival and possibly many other facets of its lifecycle henceforth. So many phenotypes and many of them shared with other SS isolates; in an elaborate system to stay naturally fit.

The Genes that did not result in a change in the phenotypes tested does not mean they are not involved in the SS phenotype and it does not mean they don’t play a
significant role in the overall lifecycle of EHEC O157:H7. There is always the possibility that these genes individually knocked out cannot affect the phenotypes observed in SS17 with the RSE model (and potentially some of them with the biofilm model), but this could be due to redundancy, magnitude of the deletion on the phenotype, and the possibility that different conditions would render different results (e.g. biofilms on plants might not require curlin fibers, but in an animal intestine they might). For example, a recent study found in the case of human infection there is potentially a role for commensals for preventing diarrheagenic \textit{E. coli} from growing in the gut in a biofilm which may modulate its pathogenesis [17]. This could indicate a larger subset of not only genes involved in various biofilms being formed, but also implications of biofilms in different environments encountered by EHEC O157:H7.

There is clearly much more work to be done to better understand and control the SS EHEC O157:H7 and EHEC collectively. To do this we must start and continue to elucidate the molecular interactions and mechanisms that mediate EHEC O157:H7 attachment and colonization with the RSE cells and the cow as a whole. We need to have a better understanding of the role of the host factors (both environmental and the microenvironment such as immune factors, microbiota, etc…) and appreciate why EHEC O157:H7 colonizes the RAJ and why it is so critical for the shedding and transmission. There are many different avenues not yet explored and they need to be if we are to move forward in finding good preventative measures against this pathogen.

First we need to finish testing all of the mutants generated in the plant adherence assays (all of the mutants) as well as the biofilm production (EDL933 mutants). All of these should be coupled with growth curves and doubling time information to rule out
phenotype changes observed being caused by differences in growth. By testing these, we can rule genes in or out as potential candidates as being involved in the phenotypes. For genes in operons, the KanR inserts should be removed through the Wanner method as well as the plasmid pKD119 should also be removed to both restore the functional operon (aside from the deleted GOI), and rule out any interplay the plasmid might have on the phenotype observed. All of these should be retested in the two adherence models to duplicate the results. Concurrently, the RSE adherence, biofilm production, and plant adherence assays should be conducted on all the representative SS strains that have not been done already to determine if the nsSNPs that were found to be conserved could be of consequence to that pattern or that particular SS isolate or group. Mutants in the genes of interest in these strains on the same assays will help future assignment of experiments. Doubling times for all of the SS isolates would also be useful information as to determine if they exhibit differences in growth in different kinds of conditions such as nutrient limiting M9, or leafy green extract.

After determining the inclusion of the KanR promoter in the construct, we reconstructed the 3’ primers to include 6-frame stop codons to prevent translational read-through. This will help ensure no pleitropic or compensatory expression complementation is not being seen. Another good test for these GOI’s is to complement them back into both the SS17 and EDL933 deletion backgrounds on an inducible promoter plasmid (such as pBAD). Together, both results showing good loss and return of function to a particular gene and its change would be good evidence for that gene being involved in the enhanced adherence in both the RSE and baby spinach model for adherence; as well as (or separately) the aggregative phenotype observed in SS EHEC
O157:H7 with the RSE cells. Once these studies have been concluded, it would be good to make the knockouts of interest in other super shedders to determine the conservation of that mechanism and/or if there is a difference in mechanisms based on origin or lineage; which could potentially be extremely important for down-the-road studies.

Once all of the proposed mutants have been tested, testing other genes of interest would be the next best step (as there may be more genes of interest not found in this study). Because none of the genes in this study produced a phenotype in the RSE model when deleted, there is a good chance that other genes are involved (or more than one gene of the original screen, which previous studies would have determined). The best way to identify what other genes that could be of interest would be to either fully sequence other super shedders (as they share these unique adherence phenotypes) or sequence the genes that already have nsSNPs to see if the nsSNP is conserved (or genes with different nsSNPs) in the other SS isolates through primer walking strategies as it will allow you to compare genes which are related and nsSNPs in those related genes (not related by function necessarily, but related biochemically or via regulation programs) and cluster them as such. Based on nsSNP analysis, there are still 31 genes with nsSNPs that could be of interest yet to be tested based on nsSNP analysis of SS17 alone.

It would also be fruitful to test the other SS isolates not previously tested on biofilm formation, leafy green and RSE adherence to start to understand the conservation of this phenotype for the genes in those strains we find interesting (goes to mechanism conservation or divergence). This could also go a long way into correlating phenotypes with specific nsSNPs and start to identify the nsSNPs required for those phenotypes observed. Full sequencing of other SS isolates will help identify candidates that can be
tested, and testing these genes will help our understanding of whether these phenotypes are characterized by a set of particular nsSNPs that are conserved or if they have separate mechanisms.

Mutations in accessory or regulatory (or regulation-related) proteins might not be conserved, but since they are part of a network of connected genes, different nsSNPs in different genes or in the same gene in different areas could have the same consequence; as could a different nsSNP in a gene that had a different nsSNP in a different isolate (so the adhesin or virulence gene itself, not a accessory protein). To generate the same gene candidate protocol, it might be good to run a proteome analysis with the SS isolates grown on the RSE cells rather than the DMEM proteome. Different interactions may change the program of these cells leading to a difference in expression. RNAseq or microarray may also be of interest as it would look at genes which are differentially expressed and can be compared to a reference to see if there may be a candidate list of genes based on expression as well as SNPs (as we did this time, but with different conditions and slightly different assays).

A transposon mutagenesis library might also be a good way at determining areas of the genome that are involved but in a less directed manner. This library would also help identify possible hypothetical proteins with no known function that might contain a SNP which might activate it or give it a function in adherence (or just not previously identified as such). This also will help identify different factors if this is a complex mechanism with multiple factors contributing to the various changes in phenotypes observed (adherence, aggregation and SS phenotypes). All of this would show if the
conservation of mechanism of the different SS isolates across lineages and origins, and how they can generate unique adherence and SS phenotypes.

Coupled with these data, an analysis of the other reference strains would go a long way to help narrow down the nsSNPs that are uncovered in these organisms to help order the nsSNPs observed. For example; if TW14359 shares a common phenotype with SS17 then the nsSNPs observed compared to TW14359 are probably not consequential for that phenotype, as opposed to EDL933 which is. Recently, non-O157:H7 isolates have been identified from SS cattle [23]. Perhaps sequencing these isolates may show a conservation of factors, or could be separate as they may have different factors at their disposal. Either way these pose an intriguing insight into the SS phenotype and may be worth looking into (if nothing else to prevent SS of these isolates too).

As previously mentioned there could also be the issue with redundancy and the possibility that this is a multi-factorial set of processes mediating different phenotypes (even if the mechanism(s)) are conserved. The best way to deal with functional redundancy is a two-pronged approach. The first is dealing with making the necessary mutations to show redundancy. By taking genes which have partial or little phenotype change in the knockout adherence assays and combining those mutations through the Wanner method, we can get simultaneous mutations which could then be tested in the systems. Genes could also be selected based on the previously mentioned method of secondary SNP conservation (within same gene or process). Once we get a good handle on which genes are involved (if more than one is) we can then test through competitive fitness tests (both in vitro & in vivo), which factors from the single mutants are the most important for the different phenotypes observed (if in fact they are separate entities).
These possibilities also need to be ruled in or out for robust replication of the results but also moving forward in prevention strategies. The more parts in the system you target, the more factors the bacteria need to change to avert these prevention strategies and too many mutations for the sake of evolution are not always fit for that system.

On the host side of experimentation, it would be good to identify the host factors and cellular proteins important for mediating adherence of these bacteria. For plants, the caveat to using whole spinach in the adherence assay is that it is the interaction of the bacteria with the cuticle of the plant and not the actual cells. To grow the EHEC O157:H7 SS strains with actual plant cells will let us know if the adherence seen on the leaves is mediated through a cellular attachment or through the waxy covering of the leaf. For the RSE/cattle it would be good to identify the gene(s) responsible for adherence of the SS strains to the cells. This would go to identifying other potential targets to target as the evolution of the cattle (based on breeding) might not be as fast as it is with E. coli (as mentioned previously, E. coli only share approximately 20% identity with each other), and could result in a good target to consider moving forward.

Once the cellular factors and the genes of interest have been identified in both the SS isolates and the RSE/plant cells, the questions surrounding the mechanism of adherence can then start to be answered. We can test the half-life of that interaction, the avidity of one receptor for the other, and regulation to see what the consequence of the nsSNP(s) involved could potentially be playing. We can couple this with ex vivo like studies for the cattle adherence (e.g. ligated intestines), to see the role of other processes after colonization in this micro environment. On the plant side, we should also test interactions with EHEC O157:H7 with the plant cells themselves to determine the level
of adherence of the bacteria to the leafy green. This will not only allow us to better
generate molecular based therapies against EHEC O157:H7 colonization in cattle and
plants, but could help us start to understand the different residue domains in similar
proteins in both other patho/serotypes of *E. coli* but potentially other organisms that
deploy similar mechanisms to colonize hosts either pathogenically or asymptptomatically
(like a reservoir).

Once these data are well understood and replicable, it will be time to test our
system in actual cattle and plant systems that are used commercially for consumption by
humans. Orally challenging calves that are susceptible to the EHEC O157:H7 and more
importantly, super shedders, we can then take different mutants and look for loss of 1)
Colonization, 2) Shedding, and 3) Super shedding phenotypes. If they are all interrelated
we will see one gene or related genes being responsible for all of the different super
shedders that are tested. If we knockout a gene important for initial colonization and/or
attachment we might not see colonization. By running the series of gene knockouts we
could potentially start to elucidate the different factors employed by the EHEC O157:H7
SS and non-SS isolates at different stages of the colonization of the cattle which will not
only delineate mechanism but also different targets that could be for ubiquitous and
specialized circumstances. Understanding the mechanism of honing to the RAJ and
colonization at that site I think are also critical for this prevention mechanism. Other
avenues that could be explored is the role of microbiota and quorum sensing and
chemotaxis in EHEC O157:H7, which could also have far reaching implications as it is
an emerging field of interest in humans, and is almost untouched in the field of reservoir-
pathogen biology.
Finally, once all of these pathways and mechanisms have been uncovered, we should have many targets, both microbial and host derived, which we can then target through either immunization practices, genetic approaches, or through antibody treatment and measuring their efficacy on preventing the spread of O157:H7 at the level of the colonization of the cattle. By the current estimation of most models, this will significantly reduce if not all but eliminate the EHEC O157:H7 from being able to be transmitted by the host that has been successfully treated to not harbor this pathogen and thus significantly reduce the number of cases per year, the associated morbidity and mortality, and reduce the overall impact on human health and commerce overall. The evolution of bacteria has been going on since the beginning of life, and pathogens find a way to stay relevant and pathogenic; but that doesn’t mean we can’t stop them one at a time as they come to light; which is what we are trying to do here.
Appendix A: Virulence-Related Genes in SS17

A-1: Virulence Genes in the SS17 Genome

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>Role in Virulence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>nleH1</em></td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>espR4</em></td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>espR1</em></td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>sepD</em></td>
<td>LEE-encoded type III secretion system component</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>eivA</em></td>
<td>LEE-encoded type III secretion apparatus protein</td>
<td>Functional Regulation</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>escE</em></td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>eivF</em></td>
<td>putative regulatory protein for type III secretion apparatus</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>escR</em></td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>escT</em></td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>espX6</em></td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>nleA</em></td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>lpxB</em></td>
<td>tetraacyldisaccharide-1-P synthase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>espL1</em></td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>eprK</em></td>
<td>putative lipoprotein of type III secretion apparatus</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>eprJ</em></td>
<td>putative Type III secretion apparatus protein</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>epaQ</em></td>
<td>type III secretion apparatus protein</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>epaO</em></td>
<td>type III secretion apparatus protein</td>
<td>Non-LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>mviM</em></td>
<td>Virulence factor</td>
<td>Host-cell damage Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td><em>kdtA</em></td>
<td>3-deoxy-D-manno-octulosonic-acid transferase</td>
<td></td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------------------------</td>
<td>------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>epaR1</td>
<td>putative virulence factor</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eprH</td>
<td>virulence factor</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espR3</td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espW</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espX1</td>
<td>predicted non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espY1</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxC</td>
<td>UDP-3-O-acyl N-acetylglucosamine deacetylase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxD</td>
<td>UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxA</td>
<td>UDP-N-acetylglucosamine acyltransferase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleB2</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleC</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleD</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxK</td>
<td>lipid A 4' kinase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>stx2A</td>
<td>Shiga toxin II subunit A</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>stx2B</td>
<td>Shiga toxin II subunit B</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxL</td>
<td>lauryl-acyl carrier protein (ACP)-dependent acyltransferase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espX7</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espN</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleG7</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleF</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleH2</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpxM</td>
<td>acyl (myristate) transferase</td>
<td>Endotoxin Biosynthesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espJ</td>
<td>translocated type III secretion system effector</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------</td>
<td>--------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>espM2</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eivJ</td>
<td>type III secretion apparatus protein</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eivl</td>
<td>type III secretion apparatus protein</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eivE</td>
<td>putative secreted protein</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eivG</td>
<td>type III secretion apparatus protein</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleB1</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>nleE</td>
<td>non-LEE-encoded type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espF</td>
<td>LEE-encoded type III secreted effector</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>orf29</td>
<td>LEE-encoded predicted type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escF</td>
<td>LEE-encoded type III secretion system component</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cesD2</td>
<td>predicted chaperone</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espB</td>
<td>secreted protein EspB</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espD</td>
<td>secreted protein EspD</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espA</td>
<td>secreted protein EspA</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sepL</td>
<td>LEE-encoded type III secretion system component</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escD</td>
<td>LEE-encoded type III secretion system component</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cesT</td>
<td>molecular chaperone</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Map</td>
<td>mitochondrial associated type III secreted effector protein</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cesF</td>
<td>chaperone for type III secretion of EspF</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espH</td>
<td>LEE-encoded type III secreted effector</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escQ</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>orf16</td>
<td>LEE-encoded predicted type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>orf15</td>
<td>LEE-encoded predicted type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escN</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>escV</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>orf12</td>
<td>LEE-encoded predicted type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espZ</td>
<td>translocated effector protein</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escI</td>
<td>LEE-encoded predicted type III secretion system component</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escJ</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escC</td>
<td>LEE-encoded type III needle complex subunit</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cesD</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>grlA</td>
<td>LEE-encoded positive regulator of transcription</td>
<td>Transcriptional Regulation</td>
<td>Chromosome</td>
</tr>
<tr>
<td>grlR</td>
<td>LEE-encoded negative regulator of transcription predicted lytic transglycosylase</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>etgA</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escU</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escS</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>escL</td>
<td>LEE-encoded predicted type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>orf4</td>
<td>LEE-encoded type III secretion system factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cesA</td>
<td>LEE-encoded chaperone locus of enterocyte effacement (LEE)-encoded regulator</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lcr</td>
<td>LEE-encoded chaperone locus of enterocyte effacement (LEE)-encoded regulator</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espG</td>
<td>LEE-encoded type III secreted effector outer membrane phospholipase A</td>
<td>LEE-mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>pldA</td>
<td>extracellular serine protease phasreqipase A</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espP</td>
<td>predicted non-LEE-encoded type III secreted effector Exonuclease SbcC, putative virulence factor</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>espL4</td>
<td>RTX-Family Toxin type III secretion apparatus protein</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17_2839</td>
<td>Exonuclease SbcC, putative virulence factor</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17_0525</td>
<td>RTX-Family Toxin</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>spaL</td>
<td>type III secretion apparatus protein</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------</td>
<td>----------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>spaP</td>
<td>Integral membrane protein-</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td></td>
<td>component of type III secretion</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>apparatus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>epaR2</td>
<td>putative virulence factor</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td>spaS</td>
<td>Integral membrane protein-</td>
<td>LEE-encoded effector secretion</td>
<td>Chromosome</td>
</tr>
<tr>
<td></td>
<td>component of type III</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>secretion apparatus</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ybjX</td>
<td>Virulence factor VirK</td>
<td>Host-cell damage</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-1452</td>
<td>non-LEE-encoded type III secreted</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-3932</td>
<td>type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-1699</td>
<td>non-LEE-encoded type III secreted</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-0858</td>
<td>non-LEE-encoded type III secreted</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yjbl</td>
<td>predicted non-LEE-encoded type III</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td></td>
<td>secreted effector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>yjcF</td>
<td>predicted non-LEE-encoded type III</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td></td>
<td>secreted effector</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SS17-0070</td>
<td>predicted non-LEE-encoded type III</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-0454</td>
<td>type III secreted effector</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-4752</td>
<td>non-LEE-encoded type III secreted</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-2027</td>
<td>predicted non-LEE-encoded type III</td>
<td>Non-LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-4686</td>
<td>LEE-encoded type III secretion system</td>
<td>LEE mediated host cell effects</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hlyB</td>
<td>Hemolysin toxin protein</td>
<td>Host-cell damage</td>
<td>pO157</td>
</tr>
<tr>
<td>hlyD</td>
<td>Hemolysin toxin protein</td>
<td>Host-cell damage</td>
<td>pO157</td>
</tr>
<tr>
<td>hlyA</td>
<td>Hemolysin toxin protein</td>
<td>Host-cell damage</td>
<td>pO157</td>
</tr>
<tr>
<td>espP</td>
<td>Type V secretion protein</td>
<td>Host-cell damage</td>
<td>pO157</td>
</tr>
<tr>
<td>hlyC</td>
<td>Hemolysin toxin protein</td>
<td>Host-cell damage</td>
<td>pO157</td>
</tr>
</tbody>
</table>
Table A-2: Adherence Related Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>Role in Virulence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>tccP</td>
<td>Tir-cytoskeleton coupling protein (TccP)</td>
<td>Effector of Adherence</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yidE</td>
<td>Mediator of hyperadherence</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ybgO</td>
<td>Predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Inv</td>
<td>Adherence and invasion OMP</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ypfA</td>
<td>AIDA-I Pertactin-like adhesin</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpfE</td>
<td>Putative fimbrial subunit</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimA</td>
<td>Major type 1 subunit fimbrin (pilin)</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yraJ</td>
<td>Predicted outer membrane protein</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>htrE</td>
<td>Putative fimbrial usher protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yagW</td>
<td>CFA/I Fimbrial adhesin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>efa1</td>
<td>Central fragment of efa1</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yeeJ</td>
<td>Putative adhesin</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yfaL</td>
<td>Type V AidA-like adhesin</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimB</td>
<td>Tyrosine recombinase of on/off regulator of fimA</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ycbF</td>
<td>Predicted periplasmic pilin chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgG</td>
<td>Assembly/transport component</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgF</td>
<td>Predicted transport protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgE</td>
<td>Predicted transport protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgD</td>
<td>Transcriptional regulator for 2nd curli operon</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgB</td>
<td>Curli nucleator protein, minor subunit in curli</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgA</td>
<td>Cryptic curli major subunit</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>csgC</td>
<td>Predicted curli production protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Crl</td>
<td>Curli-operon regulator</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>------------------</td>
<td>----------</td>
</tr>
<tr>
<td>yfcS</td>
<td>predicted periplasmic pilus chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>iha</td>
<td>IrgA-Homologue adhesin/enterobactin receptor</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimH</td>
<td>minor fimbrial subunit, D-mannose specific adhesin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yfcU</td>
<td>putative fimbrial usher</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sfmC</td>
<td>pilin chaperone, periplasmic predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sfmA</td>
<td>predicted outer membrane export usher protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sfmD</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sfmH</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sfmF</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ybgP</td>
<td>predicted pili assembly protein</td>
<td>Adhesin Biogenesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ybgD</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ecpE</td>
<td>predicted chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yadC</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yadK</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yadL</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yadM</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ecpD</td>
<td>putative fimbrial chaperone protein</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yadN</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fmiC</td>
<td>F9 fimbriae usher protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimZ</td>
<td>predicted DNA-binding transcriptional regulator</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yeC</td>
<td>predicted periplasmic pilin chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fmiA</td>
<td>major subunit of F9 fimbriae</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fmiB</td>
<td>F9 fimbriae chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fmiD</td>
<td>F9 fimbriae adhesin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpfD</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>lpfA</td>
<td>putative major fimbrial subunit</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fiml</td>
<td>fimbrial protein involved in type 1 pilus biosynthesis periplasmic chaperone,</td>
<td>Adhesin Biogenesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimC</td>
<td>required for type 1 fimbriae</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimD</td>
<td>outer membrane usher</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimF</td>
<td>fimbrial morphology</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product Description</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>----------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>fimG</td>
<td>minor subunit of type 1 fimbriae</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ycbQ</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ycbR</td>
<td>predicted periplasmic pilin chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ycbT</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ycbV</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ompA</td>
<td>outer membrane protein A (3a;II*:G;d)</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ydeS</td>
<td>putative F9 fimbriae protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ydeR</td>
<td>putative F9 fimbriae protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yehD</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yfcV</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yraH</td>
<td>predicted fimbrial-like adhesin protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yral</td>
<td>predicted periplasmic pilin chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eae</td>
<td>intimin adherence protein</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>tir</td>
<td>translocated intimin receptor protein</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>rfaH</td>
<td>DNA-binding transcriptional antiterminator</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hcpC</td>
<td>assembly protein in type IV pilin biogenesis, protein with nucleoside domain</td>
<td>Adhesin Biogenesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hcpB</td>
<td>triphosphate hydrolase domain</td>
<td>Adhesin Biogenesis</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hcpA</td>
<td>type IV major pilin subunit Calcium-Binding Antigen-43 Homologue</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cah</td>
<td>Calcium-Binding Antigen-43 Homologue</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17_0342</td>
<td>AidA-like adhesin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Hha</td>
<td>Regulator of FlII/Curlin genes</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>cpxR</td>
<td>Regulator of Type IV pili/Curli by Cell-Interaction</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ihfA</td>
<td>Regulates expression of Type I Fim and Flagella</td>
<td>Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yagZ</td>
<td>Mat fimbriillin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yagY</td>
<td>predicted chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yagX</td>
<td>predicted usher protein</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td><strong>Gene</strong></td>
<td><strong>Product</strong></td>
<td><strong>Role in Virulence</strong></td>
<td><strong>Location</strong></td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------------</td>
<td>-----------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>etpC</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpD</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpE</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpF</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpG</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpH</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpI</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpJ</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpK</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpL</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpM</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpN</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>etpO</td>
<td>type II secretion protein</td>
<td>Effector of Adherence</td>
<td>pO157</td>
</tr>
<tr>
<td>SS17-4768</td>
<td>putative major fimbrial subunit</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fimC</td>
<td>putative fimbrial chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-4767</td>
<td>putative fimbrial chaperone</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-4764</td>
<td>putative fimbrial protein</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-1639</td>
<td>porcine attaching-effacing</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>etpG</td>
<td>Type II secretion protein</td>
<td>Effector of Adherence</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-4513</td>
<td>PapC-like Fimbrial Usher</td>
<td>Usher Protein</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-2873</td>
<td>Putative Fimbrial-like adhesin</td>
<td>Fimbrial-like adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ppdB</td>
<td>Prepilin peptidase</td>
<td>Chaperone</td>
<td>Chromosome</td>
</tr>
<tr>
<td>eaeH</td>
<td>Putative adhesin with Ig-Like class 3 domain</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-0333</td>
<td>Putative adhesin</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-0524</td>
<td>Putative adhesin with Ig-Like class 3 domain</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>SS17-0342</td>
<td>AidA-like adhesin</td>
<td>Non-Fimbrial Adhesin</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------</td>
<td>------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>toxB</td>
<td>Cytotoxin B</td>
<td>Non-Fimbrial Adhesin</td>
<td>pO157</td>
</tr>
<tr>
<td></td>
<td>O-Antigen short chain regulator</td>
<td>Regulator</td>
<td></td>
</tr>
<tr>
<td>wzzB</td>
<td></td>
<td></td>
<td>Chromosome</td>
</tr>
</tbody>
</table>
A-3: Virulence Associated Genes in SS17

Table A-3: Virulence Associated Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>Role in Virulence</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>feoB</td>
<td>Ferrous iron transporter, protein B</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terF</td>
<td>Phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>phoB</td>
<td>Response regulator Phosphorus regulation system</td>
<td>Signalling</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terD</td>
<td>Phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>rcsA</td>
<td>Colanic acid synthesis regulator subunit A</td>
<td>Gene Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>rcsB</td>
<td>Colanic acid synthesis regulator subunit B</td>
<td>Gene Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaG</td>
<td>Colanic acid</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaE</td>
<td>Colanic acid glycosyl transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaL</td>
<td>Colanic acid glycosyl transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaK</td>
<td>Colanic acid biosynthesis protein</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaJ</td>
<td>Colanic acid lipid carrier transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaI</td>
<td>Colanic acid glycosyl transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaG</td>
<td>GDP-L-fucose synthetase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaF</td>
<td>Colanic acid acetyltransferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaD</td>
<td>Colanic acid polymerase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaC</td>
<td>Colanic acid glycosyl transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaB</td>
<td>Colanic acid acetyltransferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaA</td>
<td>Colanic acid glycosyl transferase</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>wcaM</td>
<td>predicted colanic acid biosynthesis protein</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------</td>
<td>----------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>bssR</td>
<td>Biofilm regulator</td>
<td>Gene Regulator</td>
<td>Chromosome</td>
</tr>
<tr>
<td>pgaD</td>
<td>Biofilm PGA synthesis auxiliary protein</td>
<td>Biofilm/Capsule</td>
<td>Chromosome</td>
</tr>
<tr>
<td>pgaA</td>
<td>Biofilm PGA outer membrant secretin</td>
<td></td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuT</td>
<td>Periplasmic-heme binding protein</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuW</td>
<td>Protein associated with heme-uptake</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuX</td>
<td>Heme-based iron utilization protein</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuY</td>
<td>Flavin Reductase</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuU</td>
<td>Heme ABC transporter</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuS</td>
<td>heme oxygenase</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>chuA</td>
<td>outer membrane heme/hemoglobin receptor</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>afuC</td>
<td>putative ATP-binding component of a transport system</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>afuB</td>
<td>putative component of transport system for ferric iron</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>afuA</td>
<td>periplasmic ferric iron-binding protein</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sbmA</td>
<td>predicted transporter</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>tolC</td>
<td>Type I Secretion system transmembrane channel protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>motA</td>
<td>Stator of flagellar motor</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>motB</td>
<td>Stator of flagellar motor</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliR</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliQ</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliO</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliN</td>
<td>Flagellar motor switch protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliM</td>
<td>Flagellar motor switch protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliL</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliK</td>
<td>Hook length control protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliJ</td>
<td>flagellar protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliI</td>
<td>Flagellum-specific ATP synthase</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliH</td>
<td>Flagellar assembly protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliG</td>
<td>Motor switch protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliF</td>
<td>M-Ring protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliS</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliD</td>
<td>Flagellar Hook-associated protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliC</td>
<td>flagellar filament structural protein (flagellin)</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliT</td>
<td>predicted chaperone</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>fliP</td>
<td>flagellar biosynthesis protein</td>
<td>Motility</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ureA</td>
<td>putative urease structural subunit A (gamma)</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>Gene</td>
<td>Product</td>
<td>Role in Virulence</td>
<td>Location</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------</td>
<td>-------------------</td>
<td>------------</td>
</tr>
<tr>
<td>ureB</td>
<td>putative urease structural subunit B (beta)</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ureC</td>
<td>putative urease structural subunit C (alpha)</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ureE</td>
<td>putative nickel metallochaperone</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ureF</td>
<td>putative urease accessory protein F</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ureG</td>
<td>putative GTP hydrolase</td>
<td>Nutrient Acquisition</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terW</td>
<td>unknown associated with putative tellurite resistance</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terZ</td>
<td>phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terA</td>
<td>phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terB</td>
<td>phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terC</td>
<td>phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>terE</td>
<td>phage inhibition, colicin &amp; tellurite resistance protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hslJ</td>
<td>heat-inducible protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sapA</td>
<td>predicted antimicrobial peptide transporter subunit</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sapB</td>
<td>predicted antimicrobial peptide transporter subunit</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sapC</td>
<td>predicted antimicrobial peptide transporter subunit</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sapD</td>
<td>peptide transport system ATP-binding protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sapF</td>
<td>peptide transport system ATP-binding protein</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>mdtI</td>
<td>multidrug efflux system transporter</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>qseE</td>
<td>AI-3 Sensory Histidine kinase 2-component regulator</td>
<td>Signalling</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yfhA</td>
<td>AI-3 Sensory Histidine kinase 2-component regulator</td>
<td>Signalling</td>
<td>Chromosome</td>
</tr>
<tr>
<td>hydN</td>
<td>formate dehydrogenase-H, [4Fe-4S] ferredoxin subunit</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yhiF</td>
<td>predicted DNA-binding transcriptional regulator</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
<tr>
<td>luxS</td>
<td>Auto-inducer 2 synthase</td>
<td>Signalling</td>
<td>Chromosome</td>
</tr>
<tr>
<td>sdiA</td>
<td>Auto-inducer 1-effected transcriptional regulator</td>
<td>Signalling</td>
<td>Chromosome</td>
</tr>
<tr>
<td>yhiE</td>
<td>acid-induced positive regulator of glutamate- acid resistance</td>
<td>Resistance</td>
<td>Chromosome</td>
</tr>
</tbody>
</table>
## Appendix B: Primer Tables

### Appendix B-1: Recombineering Primers

Table B-1: Recombineering Primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Target</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>iha</td>
<td>5' end</td>
<td>gatggcagtggaacggattctcagctggatagcgttttgTCCCGTCAGTCACGGTATAT</td>
</tr>
</tbody>
</table>
|      | 3' end | tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
## Appendix B-2: Screening Primers

### Table B-2: KanR Screening Primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>iha</td>
<td>iha-A</td>
<td>GGTGGTGACCCTGCATTCTG</td>
</tr>
<tr>
<td></td>
<td>iha-B</td>
<td>GGAGCATGTGAGCCAGAAGACG</td>
</tr>
<tr>
<td>ompA</td>
<td>ompA-A</td>
<td>CGATGTGTTGCTGCAGCTGG</td>
</tr>
<tr>
<td></td>
<td>ompA-B</td>
<td>CGGGAATGGGTTCAGGGCATC</td>
</tr>
<tr>
<td>csgG</td>
<td>csgG-A</td>
<td>CTGCCTTAGTGAGTCAGTG</td>
</tr>
<tr>
<td></td>
<td>csgG-B</td>
<td>GACAGCTCTCTTGGCACCAAC</td>
</tr>
<tr>
<td>wzzB</td>
<td>wzzB-A</td>
<td>TCTCTTTGGAGCGACTAAGAC</td>
</tr>
<tr>
<td></td>
<td>wzzB-B</td>
<td>GCAGGTATTCAGCCTTCTCC</td>
</tr>
<tr>
<td>cah</td>
<td>cah-A</td>
<td>GACGGACTGACATGAAGGC</td>
</tr>
<tr>
<td></td>
<td>cah-B</td>
<td>GACGGACTGACATGAAGGC</td>
</tr>
<tr>
<td>eaeH</td>
<td>eaeH-A</td>
<td>GCGATGGGTAGTGGCAAGTC</td>
</tr>
<tr>
<td></td>
<td>eaeH-B</td>
<td>CTATGAAGGTGAGTGGGAGC</td>
</tr>
<tr>
<td>eivA</td>
<td>eivA-A</td>
<td>CCGCTCCAGGACTGCTTAAG</td>
</tr>
<tr>
<td></td>
<td>eivA-B</td>
<td>TGTCCGGCTAAACGGCTGAAG</td>
</tr>
<tr>
<td>yehA</td>
<td>yehA-A</td>
<td>GACGCACCTGGACCTATATGG</td>
</tr>
<tr>
<td></td>
<td>yehA-B</td>
<td>GACGGACTGACATGAAGGC</td>
</tr>
<tr>
<td>yfaL</td>
<td>yfaL-A</td>
<td>GGGCAGTATGAGTGGTGGGAC</td>
</tr>
<tr>
<td></td>
<td>yfaL-B</td>
<td>TGAACGGCAAGAATAGGTTGG</td>
</tr>
<tr>
<td>eae</td>
<td>eae-A</td>
<td>CTCAGAATCGCGGTACGATG</td>
</tr>
<tr>
<td></td>
<td>eae-B</td>
<td>CTCTCTCTCTCTCTCTCTG</td>
</tr>
<tr>
<td></td>
<td>KanR-C</td>
<td>GTATTTCCCTCGCTCAGGC</td>
</tr>
<tr>
<td></td>
<td>KanR-D</td>
<td>GATGTTGGACGAGTGGGAAT</td>
</tr>
<tr>
<td></td>
<td>KanR-C2</td>
<td>GCCTGAGCGAGACGAAATAC</td>
</tr>
<tr>
<td></td>
<td>KanR-D2</td>
<td>GATGTTGGACGAGTGGGAAT</td>
</tr>
</tbody>
</table>
Appendix B-3: nsSNP Sequencing Primers

Table B-3: nsSNP Sequencing Primers

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Sequence 5'-3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>csGG</td>
<td>csgG-S1A</td>
<td>CCCGCCAGATTTGACATTGC</td>
</tr>
<tr>
<td></td>
<td>csgG-S1B</td>
<td>GACAAACCTCGACCATCCAG</td>
</tr>
<tr>
<td>wzBB</td>
<td>wzzB-S1A</td>
<td>CGTCCCAGAGCAATGTTC</td>
</tr>
<tr>
<td></td>
<td>wzzB-S1B</td>
<td>CATTGTGGCTATTGCTTGG</td>
</tr>
<tr>
<td>caHH</td>
<td>cah-S1A</td>
<td>GAACAGAGGTTGCCGTATG</td>
</tr>
<tr>
<td></td>
<td>cah-S1B</td>
<td>GAACGTGCGGGATCAATGC</td>
</tr>
<tr>
<td>eeEH</td>
<td>eaeH-S1A</td>
<td>GACGGCAAATGCAAACAGAC</td>
</tr>
<tr>
<td></td>
<td>eaeH-S1B</td>
<td>CGTATAAGACACGGTTC</td>
</tr>
<tr>
<td>eivA</td>
<td>eivA-S1A</td>
<td>CCAGCATCCGCCTCAAGAAG</td>
</tr>
<tr>
<td></td>
<td>eivA-S1B</td>
<td>CAATGCACGTGGACCTTTC</td>
</tr>
<tr>
<td>yehA</td>
<td>yehA-S1A</td>
<td>CCACCTGCTCATCTTTGCTT</td>
</tr>
<tr>
<td></td>
<td>yehA-S1B</td>
<td>GATTCGGCGAACAGTAAGG</td>
</tr>
<tr>
<td>yfaL</td>
<td>yfaL-S1A</td>
<td>GAAACACGGCGTTAGTAC</td>
</tr>
<tr>
<td></td>
<td>yfaL-S1B</td>
<td>GTATTGCATCGGTTGGAC</td>
</tr>
<tr>
<td>Iha*</td>
<td>Iha-S1A</td>
<td>GATGCCTTGGAACATGGAC</td>
</tr>
<tr>
<td></td>
<td>Iha-S2A</td>
<td>CGGATGCCAGGTAAATTC</td>
</tr>
<tr>
<td></td>
<td>Iha-S1B</td>
<td>CAACACGCATACGAGATAG</td>
</tr>
<tr>
<td></td>
<td>Iha-S2B</td>
<td>CTCTTTCGCTGAATTTAC</td>
</tr>
<tr>
<td>ompA*</td>
<td>ompA-S1A</td>
<td>GCAGGCAATTGCTGGTAAAG</td>
</tr>
<tr>
<td></td>
<td>ompA-S2A</td>
<td>GTCCAGTCTCGTAGTATTG</td>
</tr>
<tr>
<td></td>
<td>ompA-S1B</td>
<td>CCTGACGGAGTTCACTTCTT</td>
</tr>
<tr>
<td></td>
<td>ompA-S2B</td>
<td>GACAAACACATCGGTCGAC</td>
</tr>
</tbody>
</table>

*Iha and ompA have 2 primers for each direction for walking because they did not have previously described nsSNPs in SS17.
Appendix C: Genetic Analyses of SS strains

Appendix C-1: PFGE Analysis of SS Isolates

Figure C-1: PFGE Analysis and Clustering of SS Isolates
All 102 original SS isolates along with the 2 recent environmental SS isolates (RM11333 and 11326) were PFGE typed in order to determine their likeness on a gross genetic scale. The isolate name, phage type and shedding rate are listed flanking the PFGE pattern of that particular isolate.[9].
References


Found in both Typical and Atypical Enteropathogenic *E. coli* Strains. Applied and Environmental Microbiology 79: 411-414.


to Plastic but Not for Binding to Epithelial Cells. Applied and Environmental Microbiology 74: 2384-2390.


