The Pennsylvania State University

The Graduate School

Department of Food Science

REARRANGEMENT HOT SPOT (*rhs*) GENES ARE POTENTIAL MARKERS FOR MULTILOCUS SEQUENCE TYPING OF *ESCHERICHIA COLI* O157:H7 STRAINS

A Thesis in

Food Science

by

Kuanqing Liu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2009

The thesis of Kuanqing Liu was reviewed and approved* by the following:

Edward G. Dudley
Assistant Professor of Food Science
Thesis Co-Advisor

Stephen J. Knabel
Professor of Food Science
Thesis Co-Advisor

Chitrita DebRoy
Director of *E. coli* Reference Center

Wei Zhang
Assistant Professor of Biology
Joint Assist. Prof. of Food Microbiology
Illinois Institute of Technology
Special Signatory

Alyssa C. Bumbaugh
Assistant Professor of Biology
Shippensburg University
Special Signatory

John D. Floros
Professor of Food Science
Head of the Department of Food Science

* Signatures are on file in the Graduate School.

ABSTRACT

Escherichia coli O157:H7 is an important foodborne pathogen, notorious for its low infectious dose and its capacity to cause severe diseases such as hemorrhagic colitis and hemolytic uremic syndrome. Molecular subtyping methods are powerful tools for microbial source tracking during outbreaks. Pulsed-field gel electrophoresis is currently the "gold standard" molecular method for subtyping *E. coli* O157:H7, however, a sequence-based scheme such as multilocus sequence typing would provide a less ambiguous and more rapid, cost-effective, and portable method. Previous publications reported that the genomes of *E. coli* O157:H7 contain genes that are highly conserved between strains. This hindered the development of an effective MLST scheme for subtyping *E. coli* O157:H7. In this study, several *rhs* genes were targeted, which were recently shown to be under the strongest positive selection of genes encoded within the *E. coli* genome. Eighteen *E. coli* O157:H7 strains from Lineage I and II, and 15 *E. coli* O157:H7 strains from 8 clades were included. Examination of these *rhs* genes revealed 44 polymorphic loci (PL) and 10 sequence types (STs) among the 18 lineage-strains, and 280 PL and 12 STs among the 15 clade-strains. Phylogenetic analysis using *rhs* genes generally agreed with the Lineage I and II classification defined previously. This study also showed *E. coli* O157:H7 strains from Clade 8 fall into Lineage I/II. Additionally, unique markers were found in *rhsA* and *rhsJ* that might be used to define Clade 8 and Clade 6. Therefore, *rhs* genes may be useful in discriminating strains and determining the phylogeny of *E. coli* O157:H7.

TABLE OF CONTENTS

LIST OF TABLES ……vii

LIST OF FIGURES……viii

LIST OF ABBREVIATIONS AND DEFINITIONS…….....ix

ACKNOWLEDGEMENTS……...xii

CHAPTER ONE. Statement of the problem…….....1

CHAPTER TWO. Literature review…….....2

2.1 *E. coli* O157:H7…….....2

2.1.1 Ecology and origin of *E. coli* O157:H7…….....2

2.1.2 Pathogenesis of *E. coli* O157:H7…….....3

2.1.2.1 Diversity of virulence among *E. coli* O157:H7 phylogenetic groups…….....3

2.1.2.2 Virulence factors…….….4

2.1.2.3 Clinical manifestations……7

2.1.3 Epidemiology of *E. coli* O157:H7……...……...8

2.2 Subtyping of *E. coli* O157:H7 …….....11

2.2.1 Definitions, applications and performance criteria……11

2.2.2. Methods for subtyping *E. coli* O157:H7 during epidemiological investigations……..……...……....14

2.2.2.1 Phenotypic methods……14

2.2.2.2 Genotypic methods…….15

2.2.2.2.1 DNA-pattern-based methods……...15

2.2.2.2.2 DNA-sequence-based methods……... 22

2.3 Rearrangement hot spot (*rhs*) elements……....24

2.4 Conclusions……..26

2.5 References……..……27

CHAPTER THREE. Rearrangement hot spot (*rhs*) genes are potential markers for multilocus sequence typing of *Escherichia coli* O157:H7 strains…….50

3.1 Abstract……51

3.2 Introduction…… 52

3.3 Materials and methods ……56

3.4 Results……..59

3.5 Discussion……....78

3.6 Acknowledgements…… .85

3.7 References……....86

CHAPTER FOUR. Conclusions and future research ……94

4.1 Conclusions……94

4.2 Future research……...96

APPENDIX A: Supplementary materials……101

APPENDIX B: Determination of minimum inhibitory concentration (MIC) of polyhexamethylene biguanide (PHMB) for *E. coli* O157:H7…….. ...111

LIST OF TABLES

TABLE 2.1 Year and source(s) of reported *E. coli* O157:H7 outbreaks……..……10

TABLE 3.1 Length and GC content of *rhs* genes in the genome of *E. coli* O157:H7 strain Sakai…….65

TABLE 3.2 *E. coli* O157:H7 strains analyzed in the present study……..66

TABLE 3.3 Forward and reverse primers and cycling conditions for the 8 *rhs* genes analyzed…….67

TABLE 3.4 Sequencing primers for the 8 *rhs* genes analyzed……..68

TABLE 3.5 Sequence variations (SNPs, insertions and deletions) identified by *in silico* comparison within 7 *rhs* genes among 14 *E. coli* O157:H7 strains……...70

TABLE 3.6 SNP loci identified by sequencing within 8 *rhs* genes among lineage strains…….....71

TABLE 3.7 Summary of allelic types of the 7 *rhs* genes for 15 clade-strains...……...72

TABLE 3.8 Lineage specific polymorphism assay (LSPA) genotypes of *E. coli* O157:H7 strains in the present study……...……..73

LIST OF FIGURES

FIGURE 2.1 A prototypical *rhs* element…….....26

FIGURE 3.1 Unrooted neighbor-joining tree of 18 lineage-strains based upon the number of differences in parsimony informative sites among *rhsA*, *C*, *D*, *E*, *F*, *J* and *K*……...74

FIGURE 3.2 Lineage Specific Polymorphism Assay (LSPA) of 14 clade-strains……..76

APPENDIX FIGURE 1. Inhibitory effect of PHMB on the growth of *E. coli* strain K-12 MG1655……...112

APPENDIX FIGURE 2. Inhibitory effect of PHMB on the growth of *E. coli* O157:H7 strain FRIK920.…….113

APPENDIX FIGURE 3. Inhibitory effect of PHMB on the growth of *E. coli* O157:H7 strain Sakai.……...114

LIST OF ABBREVIATIONS AND DEFINITIONS

A/E	Attaching and Effacing
AFLP	Amplified Fragment Length Polymorphism
°C	Degree Celsius
Clone[†]	A group of isolates deriving from a common ancestor as part of a direct chain of replication and transmission from host to host or from the environment to host.
CNS	Central Nervous System
GUD[-]	β-glucuronidase negative
HUS	Hemolytic Uremic Syndrome
LB	Luria-Bertani
LSPA	Lineage Specific Polymorphism Assay
mL	milliliter
MIC	Minimum Inhibitory Concentration
MLEE	Multi-Locus Enzyme Electrophoresis
MLST	Multilocus Sequence Typing
MLVA	Multiple-Locus Variable-number tandem repeat Analysis
NCBI	National Center for Biotechnology Information

ng	nanogram
nm	nanometer
OBGS	Octamer-Based Genome Scanning
OD_{600}	Optical Density at 600 nm
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PFGE	Pulsed-field gel electrophoresis
PHMB	Polyhexamethylene Biguanide
PL	Polymorphic Loci
RAPD	Randomly Amplified Polymorphic DNA
rhs	rearrangement hot spot
rpm	revolutions per minute
SMAC	Sorbitol MacConkey
SNP	Single Nucleotide Polymorphism
SOR^-	Sorbitol negative
Strain†	Isolate(s) that exhibit distinct phenotypic and/or genotypic characteristics from other isolates of the same species
ST	Sequence Type

TTP	Thrombotic Thrombocytopenic Purpura
U	Unit
µl	microliter
WGS	Whole Genome Shotgun

† Clone and strain were defined previously by Struelens et al. (123).

ACKNOWLEDGEMENTS

I thank my parents, Xiumin Liu and Jingying Liu, who support and encourage me to study in the US. I am also grateful for the support of my brother, Kuanjian Liu and my sister-in-law, Jie Li.

I thank my advisors, Dr. Edward G. Dudley and Dr. Stephen J. Knabel. I really appreciate their generous help and guidance. I learned from them not only how to do research but also how to lead my life. They always encourage me to go for my dreams. I feel so honored to work with them. I also thank my committee members, Dr. Chitrita DebRoy, Dr. Wei Zhang and Dr. Alyssa C. Bumbaugh for their guidance and encouragement.

I thank my labmates, Chun Chen, Amrita Puri, Jia Wen and Dr. Yi Chen for their help and encouragement. I also want to give special thanks to Dr. Bindhu Verghese for her help with my research. Furthermore, I want to thank all the faculty, graduate students and staff in the Department of Food Science for their support.

At last, I thank USDA and the Department of Food Science for supporting my research.

CHAPTER ONE

STATEMENT OF THE PROBLEM

Escherichia coli O157:H7 is notorious for its low infectious dose and its ability to cause severe diseases, such as hemorrhagic colitis (HC) and hemolytic uremic syndrome (HUS). Outbreaks caused by this pathogen pose heavy burdens on public health and the economy. In order to implement effective intervention strategies to control ongoing outbreaks and prevent future outbreaks, it is imperative to know the sources of contamination and how *E. coli* O157:H7 is transmitted throughout the food system. Molecular subtyping methods are powerful tools for accomplishing this, because they allow accurate differentiation of specific clones. Pulsed-field gel electrophoresis (PFGE), a DNA-pattern-based method, is currently the gold standard approach for subtyping *E. coli* O157:H7. However, the drawbacks inherent in PFGE, such as the expert technical skill required and the generation of ambiguous data, make DNA-sequence-based approaches attractive alternatives for subtyping *E. coli* O157:H7. Multilocus sequence typing (MLST) is one such DNA-sequence-based scheme. Previous studies attempted to develop an MLST scheme for *E. coli* O157:H7 and examined various gene markers, such as housekeeping genes, virulence genes and genes encoding outer membrane proteins. However, these gene markers showed a striking lack of DNA sequence variation, which limited their utility for MLST of *E. coli* O157:H7 during epidemiological investigations. Therefore, the primary purpose of the present study was to identify gene markers that yield sufficient DNA sequence variation for MLST of *E. coli* O157:H7.

CHAPTER TWO

LITERATURE REVIEW

E. coli O157:H7 is a deadly bacterial pathogen, notorious for its low infectious dose and its ability to cause severe or even life-threatening diseases, such as HC and HUS. When outbreaks occur, it becomes critical to rapidly and accurately identify the source of contamination and the routes of transmission of foodborne pathogens, so that effective intervention strategies can be implemented to control the ongoing outbreaks and to prevent future outbreaks. Molecular subtyping methods are powerful tools for microbial source tracking and transmission-route identification during epidemiological investigations. Chapter Two is a brief summary of previous publications on *E. coli* O157:H7, outbreaks caused by O157:H7 and molecular subtyping methods available for this pathogen.

2.1 *E. coli* O157:H7

2.1.1 Ecology and origin of *E. coli* O157:H7

E. coli O157:H7 is one serotype of the species *E. coli*, with somatic (O) antigen 157 and flagellar (H) antigen 7. It is a gram-negative rod-shaped facultatively anaerobic bacterium. The first documentation of *E. coli* O157:H7 was in 1975, when a sporadic case involving a woman with bloody diarrhea was reported (127). However, it was not until 1982 when Riley et al. (113) discovered the implication of an unusual serotype of *E. coli* in two outbreaks that *E. coli* O157:H7 was recognized as an important foodborne pathogen. The main reservoir for *E. coli* O157:H7 is the gastrointestinal tract of cattle

(11, 55, 103, 136), although other animals may also carry this pathogen, such as swine (96) and sheep (79). This is probably why many *E. coli* O157:H7 outbreaks are associated with foods of cattle origin and foods which have come into contact with cattle feces.

2.1.2 Pathogenesis of *E. coli* O157:H7

E. coli O157:H7 can be distinguished biochemically from other *E. coli* strains by its inability to ferment sorbitol and its lack of β-glucuronidase activity (43). Moreover, unlike most *E. coli* strains that are harmless to their hosts, *E. coli* O157:H7 is a deadly pathogen that can cause severe or even life-threatening diseases. This section summarizes the diversity of virulence, well-studied virulence factors, clinical manifestations and epidemiology of *E. coli* O157:H7.

2.1.2.1 Diversity of virulence among *E. coli* O157:H7 phylogenetic groups

E. coli O157:H7 is known to cause severe or even life-threatening diseases, however, not all *E. coli* O157:H7 strains are equally pathogenic to humans. Kim et al. (76) used an octamer-based genome scanning (OBGS) approach to characterize a large collection of *E. coli* O157:H7 strains of bovine and human origin. They found a nonrandom distribution of bovine and human isolates, with most of the bovine isolates in one cluster designated Lineage II and with most of the human isolates in the other cluster designated Lineage I. They suggested *E. coli* O157:H7 strains from Lineage I were more virulent than Lineage II. Later, Baker et al. (4) included 10 *E. coli* O157:H7 strains isolated from bovine and human, respectively, and used gnotobiotic piglets as the animal model. They found *E. coli* O157:H7 strains of human origin caused significantly more deaths or central nervous

system (CNS) diseases in piglets than strains of cattle origin did, although no significant difference was observed in other symptoms, such as diarrhea and attaching and effacing (A/E) lesions. In 2007, Zhang et al. (143) used a comparative genome hybridization approach to investigate the genomic diversity between Lineage I and Lineage II strains. They identified another lineage designated Lineage I/II, which was originally grouped in Lineage II. The association of Lineage I/II with human infection has not been reported so far to the best of our knowledge. Later in 2008, Manning et al. (89) used a 96-locus single nucleotide polymorphism (SNP) subtyping scheme and defined 9 clades within *E. coli* O157:H7. Clade 8 strains were suggested to represent a highly virulent group based on the fact that strains from this clade were more frequently associated with the life-threatening disease HUS. Moreover, a trend of increasing incidence of outbreaks caused by Clade 8 strains was also implied in this study, as the frequency of Clade 8 strains among clinical *E. coli* O157:H7 strains in Michigan increased significantly from 2000 to 2006.

2.1.2.2 Virulence factors

The exact pathogenic mechanism(s) of *E. coli* O157:H7 during human infection have not been fully elucidated, largely due to the lack of appropriate animal models and virulence factors may function in a yet unknown complex and cooperative manner to cause illnesses. Various experimentally confirmed and putative virulence factors have been reported. Among these, Shiga toxins (Stxs) and the locus of enterocyte effacement (LEE) are the most well-studied virulence factors which have been shown to play key roles in *E. coli* O157:H7 infection.

Shiga toxin (Stx) is considered the major and defining virulence factor of *E. coli* O157:H7. This toxin was first discovered in *Shigella dysenteriae* type 1 in 1906 (101). It was first found in *E. coli* O157:H7 by O'Brien et al. (102) in 1983. Later in 1987, Scotland et al. (119) used techniques such as serum neutralization tests and probe hybridization, and discovered two variants of Stx in *E. coli* O157:H7, now designated Stx1 and Stx2. They also found that Stxs could be present in *E. coli* O157:H7 in three different combinations: only Stx1, only Stx2, and both Stx1 and Stx2. The Stx1 in *E. coli* O157:H7 is almost identical to that in *S. dysenteriae* type 1, and shares only 56% identity with Stx2 (129). Stx1 and Stx2 are encoded on lysogenic lambdoid phages that have integrated into the *E. coli* O157:H7 genome. Both Stx1 and Stx2 have a conserved A-B subunit structure, with the A subunit inhibiting protein synthesis and the B subunit binding the toxin to mammalian cell receptors (101). The key role Stxs play in *E. coli* O157:H7 infections has been shown by several studies. For example, Keusch et al. (74) reported that Stx was responsible for fluid accumulation in the rabbit ileal loop. Sjogren et al. (122) introduced a bacteriophage expressing Stx1 into a natural pathogenic *E. coli* of rabbits, and found that the modified strain caused much more severe infection in rabbits than the original strain that did not express Stx. More recently, Robinson et al. (114) showed that Stx2 promoted the intestinal colonization of *E. coli* O157:H7. In this study, they included the wild type *E. coli* O157:H7 strain 86-24 and an isogenetic mutant strain TUV86-2 whose *stx2* was mutated, and investigated the adhesion ability of these two strains both in *vitro* and in *vivo*. They found that TUV86-2 showed significantly weaker adhesion ability than 86-24. However, when Stx2 was present, TUV86-2 showed equivalent or greater adhesion ability compared to 86-24.

Stx1 and Stx2 are key virulence factors of *E. coli* O157:H7, however, they may not be equally toxigenic, as several studies suggested that Stx2 was a more potent toxin than Stx1. The first piece of evidence was from Scotland et al. (119) whose study showed that Stx2 was found in all O157 strains isolated from patients with HUS whereas Stx1 was not. Later in 1993, Tesh et al. (129) used mice to compare the toxicity of Stx1 and Stx2. In their study, they used purified Stx1 and Stx2 to inject mice, and found that the 50% lethal dose of Stx2 for mice was much lower than that of Stx1 (1 ng versus 400 ng). More recently, Siegler et al. (121) administrated purified Stx1 and Stx2 to baboons and found that all baboons that received Stx2 either died or became ill, whereas none of the baboons receiving Stx1 or saline became ill. Taking all of these together, it appears clear that Stx2 is more toxic than Stx1, at least in several animal models

LEE, a pathogenicity island is the other key virulence factor and it is responsible for inducing the attaching and effacing (A/E) phenotype. The A/E phenotype is characterized as cytoskeletal changes of the epithelial cells, effacement of microvilli, and intimate adherence of bacteria to the epithelial cell membrane (97). LEE encodes a type III secretion system (TTSS), which serves as a transportation system to export effector proteins into host cells. Additionally, LEE encodes secreted proteins designated EspA, B and D, which are part of the TTSS apparatus. Intimin and translocated intimin receptor, which are responsible for the adherence of bacteria to host cells are also encoded on LEE (41). LEE is necessary and sufficient for the A/E phenotype, as the introduction of LEE into avirulent *E. coli* strains conferred the A/E phenotype to the recipients (90). Intimin (encoded by *eae* on LEE) was demonstrated to play a role in intestinal colonization in animal models, as several studies reported that *E. coli* O157:H7 strains whose *eae* genes

were mutated failed to induce the A/E phenotype and failed to colonize the intestines of several pig models (42, 91, 131).

2.1.2.3 Clinical manifestations

E. coli O157:H7 can cause severe and life-threatening diseases with a low infectious dose. Previous publications reported that as low as 10 cells may cause illness (141). The clinical manifestations of *E. coli* O157:H7 infections include asymptomatic infection, HC, HUS and thrombotic thrombocytopenic purpura (TTP) (67). TTP is highly similar to HUS, though fewer neurological signs and fever were observed in HUS than TTP (67). Therefore, this section only discusses HC and HUS.

HC was found to be associated with *E. coli* O157:H7 as early as in 1975 in a sporadic case, where this pathogen was isolated from a female patient suffering from bloody diarrhea (112). However, it was not until 1982 that *E. coli* O157:H7 was recognized as the causative agent of HC. In 1982, Riley and colleagues (113) isolated an unusual serotype of *E. coli* from patients suffering from HC during the investigation of two outbreaks in Michigan and Oregon. Now, HC is recognized as a hallmark of *E. coli* O157:H7 infection. The incubation time of HC has not been fully identified. Some researchers suggested 3 - 4 days (113), whereas another study proposed ca. 8 days (116). The onset of HC begins with suddenly painful abdominal cramps, then watery diarrhea followed by bloody diarrhea, with little or no fever (113). Most patients recover from HC with proper supportive treatments, however, in a few cases, specific therapy is required, and severe complications and death can result (112).

Compared to HC, HUS is a more severe and life-threatening disease. The first description of HUS can be dated back to as early as 1955 (49). HUS is a complication of acute renal failure, thrombocytopenia, and microangiopathic hemolytic anemia (67). HUS has been associated with various bacterial pathogens (112), however, the link between HUS and *E. coli* O157:H7 was not discovered until 1983 by Karmali et al. (69). Most HUS victims are children (112), probably due to the fact that their immune systems have not been fully developed to combat *E. coli* O157:H7 infection. HUS can be cured with proper supportive treatments for most patients, however, death can occur in 3 - 5% of infected children and 12 - 30% recover from HUS with severe sequelae, such as renal impairment, hypertension, or CNS manifestations (97). In a word, the high mortality rate and severe sequelae make HUS a major public-health concern.

2.1.3 Epidemiology of *E. coli* O157:H7

E. coli O157:H7 is posing heavy economic and public health burdens. An estimated $301.8 - 726.0 million are spent on *E. coli* O157:H7 infections annually in the United States (14). On average, 74,000 illnesses (93) and 17 outbreaks (110) are attributable to *E. coli* O157:H7 each year in the United States. From 1982 to 2002, 350 *E. coli* O157:H7 outbreaks were reported in the United States, and 8,598 cases of O157 infection were identified (110). Among these cases, 1,493 (17.4%) were hospitalized, 354 (4.1%) developed HUS, and 40 (0.5%) died (110). Outbreaks caused by *E. coli* O157:H7 are found not only in the United States but also in other countries. A number of *E. coli* O157:H7 outbreaks have been reported in Canada (10, 15, 86, 87, 106, 126), Germany (85, 111), the United Kingdom (134), Finland (107), Sweden (117) and Japan (2, 95, 135). Most *E. coli* O157:H7 outbreaks can be traced back to foods of cattle origin (7, 36,

37, 71, 113), although contaminated water, apple juice and fresh produce have also been implicated in *E. coli* O157:H7 outbreaks (32, 125, 138). A summary of *E. coli* O157:H7 outbreaks is listed in Table 2.1.

TABLE 2.1 Year and source(s) of reported *E. coli* O157:H7 outbreaks.

Year	Source(s) or most likely source(s)	Reference
1982	Ground beef	(113)
1984	Hamburger	(116)
1985	Sandwich	(15)
1985	ND	(68)
1986	Ground beef	(104)
1988	Meat patties	(7)
1989	Water	(125)
1990	Roast beef	(22)
1990	Water	(2)
1991	Apple cider	(8)
1991	Lake water	(72)
1992	Hamburger	(37)
1992	Raw milk	(71)
1992	Ground beef	(6)
1992	Unknown	(111)
1993	Hamburger	(21)
1994	Ground beef	(19)
1994	Salami	(20)
1995	Lake water	(25)
1995	Hamburger	(30)
1995	Green salad	(109)
1995	Lettuce	(1)
1995	Roast beef and salad	(115)
1995	Deer meat	(73)
1996	Apple juice	(38)
1996	Apple cider	(34)
1996	Radish sprouts	(95)
1996	Radish sprouts	(135)
1996	Lettuce	(58)
1996	Water	(46)
1996	Apple cider	(59)
1997	Alfalfa sprouts	(12)
1997	Beef patties	(17)
1997	Water	(107)
1998	Apple cider	(126)
1998	Cheese curds	(31)
1998	Salami	(140)
1999	Water	(29)
1999	Salami	(87)
1999	Ground beef	(86)
1999	Water	(13)
1999	Beef tacos	(65)
2000	Contact with animals	(35)
2000	Contact with animals	(40)

TABLE 2.1 Year and source(s) of reported *E. coli* O157:H7 outbreaks continued.

Year	Source(s) or most likely source(s)	Reference
2001	Japanese-style pickles	(105)
2002	Ground beef	(28)
2002	Salads and sandwiches	(10)
2002	Ground beef	(133)
2002	Fermented sausage	(117)
2003	Beef	(80)
2004	Petting zoo	(51)
2004	Ground beef	(18)
2004	Beef	(124)
2004 -2005	Contact with animals	(33)
2005	Frozen beef burgers	(77)
2005	Raw milk	(16)
2006	Spinach	(138)
2006	Lettuce	(27)
2006	Spinach	(53)
2006	Beef donair	(62)
2007	Ground beef	(26)
2007	Pizza with pepperoni	(24)
2008	Ground beef	(23)

ND: Not Determined

2.2 Subtyping of *E. coli* O157:H7

2.2.1 Definitions, applications and performance criteria

Molecular subtyping is the process of characterizing bacteria at the strain level (123). It generally has two major applications. The first is to study the long-term evolutionary history or population genetics of bacteria. For example, Feng et al. (44) used a multilocus enzyme electrophoresis (MLEE) scheme combined with allele-specific probe for the *uidA* gene that encodes the beta-glucuronidase enzyme, and proposed a stepwise evolutionary model for *E. coli* O157:H7. In this model, *E. coli* O157:H7 was shown to stem from an O55:H7-like ancestor and to diverge into SOR^- GUD^- *E. coli* O157:H7 (typical *E. coli* O157:H7) and O157:H- (atypical *E. coli* O157:H7) after acquiring Stx. Later in 1999, Kim et al. (76) used a subtyping approach termed octamer-based genome

scanning (OBGS) to investigate the evolution of *E. coli* O157:H7. They identified two distinct lineages within *E. coli* O157:H7 population designated Lineage I and II, with Lineage I strains more frequently associated with human diseases. These two studies provided insight into the evolutionary history and emergence of *E. coli* O157:H7.

The second application of molecular subtyping is to study the short-term evolutionary history of bacteria for microbial source tracking in epidemiological investigations. Many subtyping methods have been reported for identifying sources of contamination and routes of transmission of *E. coli* O157:H7. Bell et al. (6) applied PFGE to investigate a large *E. coli* O157:H7 outbreak in Washington, and found that the fingerprints of *E. coli* O157:H7 isolated from patients and hamburgers were the same. Their results confirmed that undercooked hamburgers were the source of contamination in this large *E. coli* O157:H7 outbreak. More recently, multiple-locus variable-number-tandem-repeat analysis (MLVA) was used by Cooley et al. (39) to identify the routes of transmission of *E. coli* O157:H7 in a fresh produce production region in California. They found that *E. coli* O157:H7 strains with the same MLVA genotype were isolated from different locations surrounding the produce production region (Farm A). This indicated that *E. coli* O157:H7 was widely transmitted to the environment and contamination due to this pathogen was quite prevalent. They also suggested that *E. coli* O157:H7 could persist in the environment for long periods of time, as isolates with the same MLVA genotype were isolated over several months. However, it should be kept in mind that in this study some *E. coli* O157:H7 isolates with an identical MLVA genotype were distinguishable by PFGE, which implies *E. coli* O157:H7 isolates with an identical MLVA genotype might not actually be the same strain.

Two broad categories of criteria have been proposed to evaluate molecular subtyping schemes, performance criteria and practical criteria. Performance criteria are the most important and include typeability, reproducibility, discriminatory power and epidemiologic concordance. Typeability is the proportion of strains that are assigned a type by a subtyping method; reproducibility is the ability of a subtyping method to assign the same type to a strain tested on independent, separate assays; discriminatory power is the average probability that a subtyping method will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon; epidemiologic concordance is the probability that epidemiologically related strains derived from presumably single-clone outbreaks are determined to be similar enough to be classified into the same clones (123). Equations of discriminatory power and epidemiologic concordance are given below.

$$D = 1 - \frac{1}{N(N-1)} \sum_{j=1}^{S} n_j (n_j - 1)$$

D: the index of discriminatory power; *N*: the number of unrelated strains tested; *S*: the number of different types; n_j: the number of strains belonging to the *j*th type.

$$E = N_e/N$$

E: the index of epidemiologic concordance; N_e: the number of strains assigned to epidemic clones; *N*: the number of strains tested from well-defined outbreaks.

An ideal subtyping method should have 100% typeability and 100% reproducibility so that it is applicable to all bacterial isolates and the subtyping results are repeatable within the same laboratory and between different laboratories. Discriminatory power and epidemiologic concordance are the most important performance criteria for evaluating a subtyping scheme. High discriminatory power is often desired for a subtyping method to separate outbreak-unrelated strains during epidemiological

investigations. However, a subtyping method should also possess high epidemiologic concordance to correctly cluster all outbreak-related strains and separate them from outbreak-unrelated strains. In the past, efforts were solely focused on achieving high discriminatory power. Now it is generally accepted that a subtyping method should yield both high discriminatory power and high epidemiologic concordance. In addition to those performance criteria discussed above, there are other criteria related to practical concerns, such as low cost, high throughput, high portability, and ease of use.

2.2.2. Methods for subtyping *E. coli* O157:H7 during epidemiological investigations

Molecular subtyping methods can be categorized into two broad groups, phenotypic methods and genotypic methods.

2.2.2.1 Phenotypic methods

Before the advent of genotypic approaches, phenotypic methods were the only tools for discriminating *E. coli* O157:H7 strains. Such methods utilize phenotypic differences.

Phage typing was one of the first phenotypic subtyping schemes. In this scheme, a panel of bacteriophages is used to infect bacteria and the patterns of lysis are used to discriminate bacterial strains. This method was useful for characterizing *E. coli* O157:H7 when our knowledge of bacterial genetics was limited. However, phage typing suffers from several disadvantages. First, it is a tedious protocol that is not amenable to high-throughput. Typically, a panel of 16 different bacteriophages (5, 47, 75) was used to characterize one strain, which would be a huge amount of work if the number of strains is large. Second, the interpretation of the lysis patterns is subjective and discrepancies may arise. As seen in the study of Khakhria et al. (75), the lysis patterns were qualitative

rather than quantitative. For example, "++" was given to 21 - 70 plaques whereas "+" was assigned to 5 - 20 plaques. However, it was difficult to claim a pattern with 21 plaques was truly different from that with 20 plaques. Third, the typeability of phage typing sometimes is less than 100%, as some strains are resistant to the phage panel used (54, 75). Lastly, phage typing possesses low discriminatory power for subtyping *E. coli* O157:H7 during epidemiological investigations. Khakhria et al. (75) used a panel of 16 phages and distinguished only 62 phage types among 3,273 O157 isolates. Moreover, only 13 phage types were identified among 556 *E. coli* O157:H7 isolates representing 151 outbreaks.

2.2.2.2 Genotypic methods

In contrast to phenotypic methods, genotypic subtyping approaches target genetic differences in bacteria. They are more discriminatory, more cost- and time-effective, and generally show better reproducibility. All these advantages render genotypic methods more suitable for subtyping *E. coli* O157:H7. Moreover, the advancements in bacterial genomics and DNA sequencing techniques have enhanced genotypic subtyping of *E. coli* O157:H7. Genotypic methods can be further classified into two subgroups, DNA-pattern-based and DNA-sequence-based methods.

2.2.2.2.1 DNA-pattern-based methods

A number of DNA-pattern-based approaches have been reported for *E. coli* O157:H7 and they are PFGE, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and MLVA.

PFGE is currently the gold standard approach for subtyping *E. coli* O157:H7. It was originally described by Schwartz and Cantor (118) in 1984 as a method to separate large DNA fragments of yeast up to 2000 kb. The first step in PFGE is to digest bacterial genomic DNA with a rare-cutting endonuclease that cuts the genomic DNA at specific restriction sites. In order to avoid mechanical shearing of the genomic DNA, bacterial cells are first immobilized in agarose plugs, cells are subsequently lysed, and the genomic DNA is digested by the rare-cutting restriction endonuclease. Next, the agarose slices containing the digested genomic DNA are put into wells of an agarose gel. The gel is subsequently subjected to an electrical field whose orientation is changed periodically to separate large DNA fragments that normally cannot be resolved by a constant unidirectional electrical field. PFGE possesses high discriminatory power to allow separation of *E. coli* O157:H7 strains that are indistinguishable by other subtyping methods (48, 57, 78), and has been widely used to investigate *E. coli* O157:H7 outbreaks (6, 52, 66, 71). However, several studies reported problems associated with PFGE. First, the typeability of PFGE sometimes is less than 100%. Johnson et al. (66) showed that not all *E. coli* O157:H7 strains could be subtyped by PFGE due to the repeated degradation of genomic DNA. Meng et al. (94) also reported that the PFGE profile of one *E. coli* O157:H7 strain was a repeated smear of degraded genomic DNA possibly due to the activity of endogenous nuclease. Second, PFGE banding patterns may not be stable and thus the results are sometimes difficult to interpret. Shima et al. (120) showed that the PFGE banding patterns of *E. coli* O157:H7 changed by 1 - 8 bands after repeated subculture and prolonged storage. Iguchi et al. (64) also showed that the PFGE banding patterns of *E. coli* O157:H7 changed by 1 - 5 fragments after subculturing for 50 times.

The instability of DNA banding patterns makes it sometimes difficult to correctly interpret PFGE results. Since a single nucleotide mutation may cause up to three changes in a PFGE banding pattern, Goering (50) proposed that isolates showing differences of less than 4 bands should be considered as the same strain. On the contrary, Barrett et al. (5) suggested that for highly clonal bacteria, such as *E. coli* O157:H7, a difference in more than one PFGE band may be indicative of different strains, although they also suggested that it might not be reliable to classify strains as related or unrelated by a single band difference. To resolve this discrepancy and to provide general guidelines for interpreting PFGE banding patterns, Tenover et al. (128) proposed the following suggestions: strains showing no fragment differences with the outbreak strain are part of the outbreak (Indistinguishable); strains showing 1 fragment difference with the outbreak strain are probably part of the outbreak (Closely related); strains showing 2-fragment differences with the outbreak strain are possibly part of the outbreak (Possibly related); and strains showing more than 3-fragment differences with the outbreak strain are not part of the outbreak (Unrelated). However, analysis of PFGE results cannot be completely objective even with these guidelines, and different researchers may have different interpretations of the same results, which would complicate the epidemiological investigations of *E. coli* O157:H7. Lastly, while PFGE may be useful in revealing a relatively short evolutionary history of *E. coli* O157:H7, i.e. outbreak investigation, it may not serve as a useful tool for inferring the long-term evolutionary history of *E. coli* O157:H7. PFGE targets quickly evolving genetic events, such as insertions, deletions and inversions that are not good markers for revealing the long-term evolutionary history of *E. coli* O157:H7. As reported by Shima et al. (120) and Iguchi et al. (64), PFGE

banding patterns may change even during laboratory transfers. Moreover, Laing et al. (81) showed that PFGE failed to group *E. coli* O157:H7 strains by known lineages. To sum up, PFGE is a subtyping scheme of high discriminatory power, however, it suffers from several drawbacks that limit its performance for epidemiological investigations of *E. coli* O157:H7.

RAPD is a pattern-based subtyping method, which was first described by Williams et al. (139) and Welsh et al. (137). In RAPD, arbitrary primers that are usually 10 nucleotides in length are used. These primers anneal to multiple sites of the bacterial genome, and the genomic diversity of bacterial strains is reflected as the different DNA banding patterns. RAPD is a relatively simple and rapid subtyping method. Its reproducibility is promising (9), although the stability of RAPD patterns has not been examined for long-term subculturing of *E. coli* O157:H7. Birch et al. (9) demonstrated that the discriminatory power of RAPD was higher than phage typing, as O157 strains from the same phage type possessed different RAPD profiles. Later in 1998, Katsuda et al. (70) used RAPD to characterize a total of 102 *E. coli* O157:H7 strains from cattle and they identified only 4 RAPD profiles. Their results indicated the discriminatory power of RAPD might not be satisfactory for epidemiological investigations of *E. coli* O157:H7. In 1999, Funamori et al. (48) compared RAPD with PFGE, the current gold standard subtyping method for *E. coli* O157:H7. They found that PFGE was more discriminatory than RAPD as strains belonging to the same RAPD profile were distinguishable by PFGE. Moreover, this study also showed RAPD could not separate outbreak-unrelated *E. coli* O157:H7 strains. All of the above suggested that RAPD might not be useful for subtyping *E. coli* O157:H7 during epidemiological investigations. Lastly, RAPD

banding patterns are sometimes ambiguous as smears and tilted bands were observed (9), although Birch et al. (9) claimed that the computer-aided gel analysis might reduce the subjectivity of pattern analysis.

AFLP is another DNA-pattern-based method. Zabeau and Vos (142) originally described this scheme in 1993. In AFLP, purified genomic DNA is digested with restriction enzymes and the digested DNA fragments are ligated to adaptors that are specific to each restriction enzyme used. Next, the restriction fragments ligated with specific adaptors are amplified using PCR primers that are complementary to the adaptors. Then, the PCR amplicons are separated by gel electrophoresis and the banding patterns are used to differentiate bacterial strains. AFLP is a relatively simple and fast approach and its reproducibility seems better than PFGE. Shima et al. (120) showed that the banding patterns of AFLP did not change after repeated subculturing and prolonged storage, whereas those of PFGE changed by 1 - 8 bands. The discriminatory power of AFLP varies with the number of primer combinations used. Heir et al. (57) applied both AFLP and PFGE to characterize a set of 82 O157 strains. They identified 24 AFLP profiles using one primer combination compared with 51 PFGE patterns. Therefore, PFGE seems more discriminatory than AFLP with only one primer combination. However, Zhao et al. (144) found that AFLP with three primer combinations was more discriminatory than PFGE as 37 AFLP profiles were identified among 48 *E. coli* O157:H7 strains, whereas 21 PFGE patterns were observed. Moreover, *E. coli* O157:H7 strains indistinguishable by PFGE were resolved by AFLP. Later, Tsai et al. (130) used two primer combinations in AFLP and found that AFLP possessed slightly lower discriminatory power than that of PFGE, as 26 AFLP profiles were observed in 36 *E. coli*

O157:H7 strains compared with 29 PFGE profiles. However, it is worth mentioning that in this study AFLP sometimes could resolve *E. coli* O157:H7 strains indistinguishable by PFGE and vice versa. This indicates that a combination of AFLP and PFGE could achieve higher discriminatory power than each method alone. However, caution should be taken as extremely high discriminatory power could result in low epidemiologic concordance. Additionally, the banding patterns of AFLP, like those of PFGE and RAPD are sometimes ambiguous as smears and tilted bands were observed (144).

MLVA, another pattern-based subtyping method recently started gaining popularity and was referred as the "second generation subtyping" scheme for *E. coli* O157:H7 (63). MLVA targets tandem duplications of short DNA sequences within bacterial genomes. MLVA possesses high discriminatory power for subtyping *E. coli* O157:H7. Noller et al. (98) reported an MLVA scheme that achieved equal epidemiologic concordance to and higher discriminatory power than PFGE for subtyping *E. coli* O157:H7. Later, Lindstedt et al. (83) proposed an automated MLVA scheme using capillary electrophoresis and fluorescent-labeled primers. They further improved this scheme by using multiple dyes and multiplex PCR to make MLVA a time-effective and high-throughput approach for *E. coli* O157:H7 (84). Hyytia-Trees et al. (63) in 2006 also presented an MLVA scheme by targeting 9 variable number tandem repeat loci and showed that this scheme possessed equal discriminatory power to PFGE for subtyping *E. coli* O157:H7. Additionally, MLVA is a less labor-intensive and less time-consuming protocol than PFGE. Moreover, the potential of automation and high-throughput makes MLVA an attractive alternative subtyping scheme for *E. coli* O157:H7. However, Noller et al. (98) noticed that *E. coli* O157:H7 strains from the same outbreak contained different MLVA genotypes, which

indicated that MLVA patterns may not be stable during an outbreak. Later, Noller et al. (99) found that MLVA markers even changed during laboratory transfers and one locus (TR2) was a hot spot of mutation. Therefore, difficulty may arise in interpreting the relatedness between strains having slightly different MLVA genotypes. Unlike PFGE for which general guidelines have been proposed for interpreting PFGE banding results, no such guidelines are currently available for MLVA, which probably limits its application to subtyping other bacterial pathogens. To conclude, while MLVA may not replace PFGE as the next gold standard subtyping method, it is clear that it can serve as a useful complement to PFGE in epidemiological investigations of *E. coli* O157:H7. This trend was reflected in the work of Cooley et al. (39), who used an MLVA scheme alongside PFGE to investigate the transmission and prevalence of *E. coli* O157:H7 in a produce production region in California. They found that *E. coli* O157:H7 with the same MLVA genotypes were isolated from different locations surrounding the produce production region (Farm A). This indicated that *E. coli* O157:H7 was widely transmitted and disseminated into the environment and contamination due to this pathogen was quite prevalent. They also suggested that *E. coli* O157:H7 could persist in the environment for long periods of time, as isolates with the same MLVA genotype were isolated over several months. However, it should be kept in mind that in this study some *E. coli* O157:H7 isolates with an identical MLVA genotype were distinguishable by PFGE, which implies *E. coli* O157:H7 isolates with an identical MLVA genotype might not actually be the same strain.

2.2.2.2.2 DNA-sequence-based methods

MLST is a sequence-based subtyping scheme that generates unambiguous DNA sequence results. It was first described by Maiden et al. (88) in 1998. In the original MLST scheme for *Neisseria meningitidis*, several housekeeping genes were targeted and sequenced. DNA sequence variations termed SNPs are identified to discriminate bacterial strains. MLST markers are usually distributed evenly within the genome to avoid the confounding impacts of horizontal gene transfer. Moreover, MLST gene markers are usually several hundred base pairs in length to facilitate automated Sanger sequencing. Compared with PFGE, MLST is a less labor-intensive and time-consuming protocol. MLST involves simple and common techniques, such as primer design, PCR amplification and DNA sequencing, none of which require much expert techniques and take much time. Additionally, MLST has the potential to be adapted to a high-throughput SNP-based approach. Manning et al. (89) incorporated 96 SNP loci in their subtyping scheme for *E. coli* O157:H7. Instead of sequencing the full length of gene markers possessing the 96 SNP loci, they applied real-time PCR to directly target these SNP loci. They identified 39 SNP genotypes among 528 O157 isolates and grouped these 39 SNP genotypes into 9 distinct clades. Moreover, the biggest advantage of MLST over PFGE and MLVA is that it generates unambiguous DNA sequence data, and therefore the interpretation of MLST results can be completely objective. As discussed above, due to the instability of banding patterns, the generation of smears and tilted fragments, the analysis of PFGE results can be subjective. Specific guidelines have been proposed to interpret the relatedness between test strains and outbreak strains (128), however, these guidelines are ambiguous themselves as the words “possibly” and “probably” do not have

specific meanings. In contrast, the DNA sequences generated by MLST are clear and unambiguous. No guidelines are required to interpret MLST results and thus discrepancies can be avoided.

However, due to the lack of sequence variation in selected gene markers, no satisfactory MLST approaches have been reported for subtyping *E. coli* O157:H7 during epidemiological investigations. For example, Noller et al. (100) targeted 7 housekeeping genes (*arcA*, *aroE*, *dnaE*, *mdh*, *gnd*, *gapA*, and *pgm*) and 2 membrane protein genes (*ompA* and *espA*) in their MLST scheme to subtype a total of 77 *E. coli* O157:H7 isolates. The only sequence variations identified were 2 SNPs in *ompA*, found in 5 *E. coli* O157:H7 isolates. In another study, Foley et al. (45) chose 1 housekeeping gene (*uidA)* and 3 virulence genes as their targets (*eaeA*, *hlyA*, and *fliC*) and examined 92 *E. coli* O157:H7 isolates from various sources. They found only 5 MLST types among the 92 *E. coli* O157:H7 isolates, compared with 72 distinct profiles by PFGE. Therefore, it becomes imperative to find gene markers that yield sufficient DNA sequence variations in order to develop an MLST scheme for subtyping *E. coli* O157:H7.

To conclude, although the advantages inherent in DNA-sequence-based methods render MLST an attractive subtyping approach for *E. coli* O157:H7, the remarkable lack of sequence variation among selected gene markers hinders the effectiveness of MLST for subtyping this pathogen. Therefore, in order to develop an MLST scheme for subtyping *E. coli* O157:H7 during epidemiological investigations, gene markers that yield sufficient DNA sequence variations need to be identified.

2.3 Rearrangement hot spot (*rhs*) elements

The lack of DNA sequence variation among housekeeping genes, virulence genes and genes encoding outer membrane genes observed by Noller et al. (100) and Foley et al. (45) can be explained by two possibilities. First, *E. coli* O157:H7 is thought to have emerged as a pathogen relatively recently, and sufficient time may not have passed for SNPs to accumulate to the level necessary for MLST to be effective. Second, gene markers used in these previous studies may not be under strong enough positive selective pressure and therefore allelic variants are not selected for at an appreciable rate for these gene markers to be useful in revealing the short-term evolutionary history of *E. coli* O157:H7.

Therefore, genes under strong positive selection may be candidate markers in an MLST scheme for subtyping *E. coli* O157:H7. Peterson et al. (108) used a comparative genomics approach to identify genes under strong positive selection within genomes of *E. coli* and *Shigella flexneri*. They found a number of genes that are under strong positive selection, such as *eaeH*, the gene responsible for the A/E phenotype, *ompA*, the gene encoding an outer membrane protein, and *rzpR*, the gene encoding a putative Rac prophage endopeptidase. Among these genes, rearrangement hot spot (*rhs*) genes are predicted to be under the strongest positive selection of all the coding sequences analyzed within the *E. coli* genome. Therefore, *rhs* genes are evolving more rapidly and accordingly they may yield sufficient DNA sequence variations that can allow an MLST scheme to separate *E. coli* O157:H7 strains.

The *rhs* elements were first described by Lin et al. (82) in *E. coli* strain K-12. K-12 has 5 *rhs* elements designated *rhsA*, *B*, *C*, *D*, and *E*. The prototypical *rhs* element consists of three open reading frames (ORFs), the Val-Gly dipeptide repetition (*vgr*) ORF, the *rhs* core ORF and the downstream (ds) ORF (Fig. 2.1). The *rhs* core ORF (named as *rhs* genes in the present study) is the most conserved component, while the *vgr* ORF is not conserved in all *rhs* elements (59). The prototypical *E. coli* O157:H7 strain Sakai has 9 *rhs* genes within its genome (56). The nine *rhs* genes are designated as *rhsA*, *C*, *D*, *E*, *F*, *G*, *I*, *J* and *K*. A high level of sequence homology is observed in *rhs* genes. For example, *rhsA*, *C* and *F* share a homologous region of ca. 3.7 kb. While the functions of *rhs* genes are still largely unknown, several studies shed light on this puzzle. One study by Hill et al. (61) predicted that *rhs* genes encode extracellular proteins based upon their primary sequences, however, no experimental evidence was reported to support this claim. Another study suggested RhsA might promote the intestinal colonization of *E. coli* in calves (132) and the biogenesis of group 2 capsules (92). Allen et al. (3) showed a possible role of RhsB and RhsD in increasing the resistance of *E. coli* to the biocide polyhexamethylene biguanide (PHMB). All of these together indicate that Rhs proteins might promote survival of *E. coli* during intestinal transit in some undefined manner, and that their extracellular nature might result in a strong positive selective pressure exerted by the host's immune system.

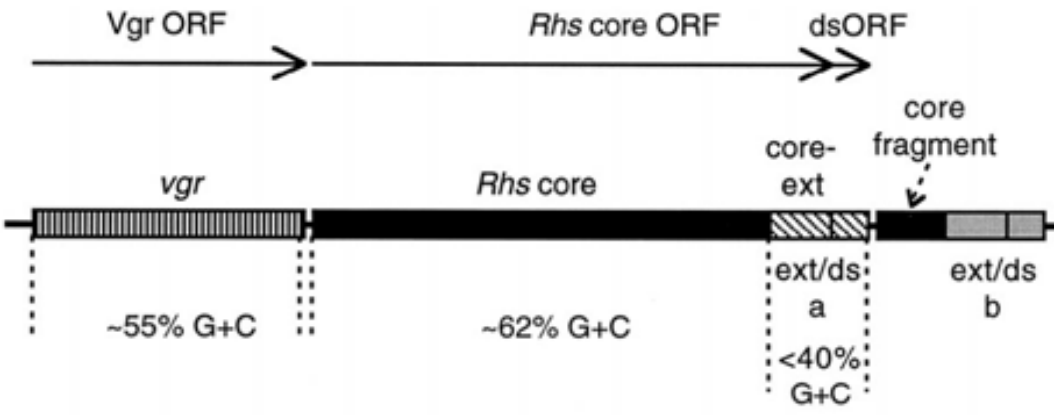


FIGURE 2.1 A prototypical *rhs* element (60). The *rhs* core with a GC content of ca. 62% and a length of ca. 3.7 kb is the most conserved component. The Val-Gly dipeptide repetition (*vgr*) open reading frame (ORF), with a GC content of ca. 55% is linked to the 5' end of the core. The *vgr* ORF is not conserved in all *rhs* elements. A mosaic segment comprising the core extension (core-ext), downstream (ds) element and core fragment is adjacent to the 3' end of the core.

2.4 Conclusions

E. coli O157:H7 is a deadly foodborne pathogen. In order to limit outbreaks caused by this pathogen, it is imperative to identify its sources and routes of transmission. Molecular subtyping methods are powerful tools for accomplishing this. Many subtyping schemes, such as PFGE, MLVA, RAPD and AFLP, have been developed for subtyping *E. coli* O157:H7, however, each method has its own pros and cons. PFGE serves as the current gold standard largely due to its high discriminatory power. However, PFGE, along with some other DNA-pattern-based subtyping schemes, such as MLVA, generates ambiguous data that are difficult to interpret. Although guidelines are available to help interpret PFGE results, discrepancies cannot be avoided. In contrast, DNA-sequence-based subtyping approaches, such as MLST, generate unambiguous data that can be interpreted without discrepancy. Moreover, the potential of being a high-throughput method also makes MLST an attractive subtyping scheme for *E. coli* O157:H7. However,

previous attempts at developing MLST schemes for *E. coli* O157:H7 were not successful as housekeeping genes, virulence genes and genes encoding outer membrane proteins lacked sufficient sequence diversity. To search for gene markers that yield sufficient DNA sequence variations for MLST of *E. coli* O157:H7, I targeted *rhs* genes, which are predicted to be under the strongest positive selection of genes analyzed within the *E. coli* genome. I speculated that *rhs* genes are evolving more rapidly than housekeeping genes and virulence genes, and thus may generate sufficient DNA sequence variation to permit MLST of *E. coli* O157:H7.

2.5 References

1. **Ackers, M. L., B. E. Mahon, E. Leahy, B. Goode, T. Damrow, P. S. Hayes, W. F. Bibb, D. H. Rice, T. J. Barrett, L. Hutwagner, P. M. Griffin, and L. Slutsker.** 1998. An outbreak of *Escherichia coli* O157:H7 infections associated with leaf lettuce consumption. J. Infect. Dis. **177:**1588-1593.

2. **Akashi, S., K. Joh, A. Tsuji, H. Ito, H. Hoshi, T. Hayakawa, J. Ihara, T. Abe, M. Hatori, T. Mori, and et al.** 1994. A severe outbreak of haemorrhagic colitis and haemolytic uraemic syndrome associated with *Escherichia coli* O157:H7 in Japan. Eur. J. Pediatr. **153:**650-655.

3. **Allen, M. J., G. F. White, and A. P. Morby.** 2006. The response of *Escherichia coli* to exposure to the biocide polyhexamethylene biguanide. Microbiology **152:**989-1000.

4. **Baker, D. R., R. A. Moxley, M. B. Steele, J. T. Lejeune, J. Christopher-Hennings, D. G. Chen, P. R. Hardwidge, and D. H. Francis.** 2007. Differences in virulence among *Escherichia coli* O157:H7 strains isolated from humans during disease outbreaks and from healthy cattle. Appl. Environ. Microbiol. **73:**7338-7346.

5. **Barrett, T. J., H. Lior, J. H. Green, R. Khakhria, J. G. Wells, B. P. Bell, K. D. Greene, J. Lewis, and P. M. Griffin.** 1994. Laboratory investigation of a multistate food-borne outbreak of *Escherichia coli* O157:H7 by using pulsed-field gel electrophoresis and phage typing. J. Clin. Microbiol. **32:**3013-3017.

6. **Bell, B. P., M. Goldoft, P. M. Griffin, M. A. Davis, D. C. Gordon, P. I. Tarr, C. A. Bartleson, J. H. Lewis, T. J. Barrett, J. G. Wells, R. Baron, and J. Kobayashi.** 1994. A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. JAMA **272:**1349-1353.

7. **Belongia, E. A., K. L. MacDonald, G. L. Parham, K. E. White, J. A. Korlath, M. N. Lobato, S. M. Strand, K. A. Casale, and M. T. Osterholm.** 1991. An outbreak of *Escherichia coli* O157:H7 colitis associated with consumption of precooked meat patties. J. Infect. Dis. **164:**338-343.

8. **Besser, R. E., S. M. Lett, J. T. Weber, M. P. Doyle, T. J. Barrett, J. G. Wells, and P. M. Griffin.** 1993. An outbreak of diarrhea and hemolytic uremic syndrome from *Escherichia coli* O157:H7 in fresh-pressed apple cider. JAMA **269:**2217-2220.

9. **Birch, M., D. W. Denning, and D. Law.** 1996. Rapid genotyping of *Escherichia coli* O157 isolates by random amplification of polymorphic DNA. Eur. J. Clin. Microbiol. Infect. Dis. **15:**297-302.

10. **Bolduc, D., L. F. Srour, L. Sweet, A. Neatby, E. Galanis, S. Isaacs, and G. Lim.** 2004. Severe outbreak of *Escherichia coli* O157:H7 in health care institutions in Charlottetown, Prince Edward Island, fall, 2002. Can. Commun. Dis. Rep. **30:**81-88.

11. **Borczyk, A. A., M. A. Karmali, H. Lior, and L. M. Duncan.** 1987. Bovine reservoir for verotoxin-producing *Escherichia coli* O157:H7. Lancet **1:**98.

12. **Breuer, T., D. H. Benkel, R. L. Shapiro, W. N. Hall, M. M. Winnett, M. J. Linn, J. Neimann, T. J. Barrett, S. Dietrich, F. P. Downes, D. M. Toney, J. L. Pearson, H. Rolka, L. Slutsker, and P. M. Griffin.** 2001. A multistate outbreak of *Escherichia coli* O157:H7 infections linked to alfalfa sprouts grown from contaminated seeds. Emerg. Infect. Dis. **7:**977-982.

13. **Bruce, M. G., M. B. Curtis, M. M. Payne, R. K. Gautom, E. C. Thompson, A. L. Bennett, and J. M. Kobayashi.** 2003. Lake-associated outbreak of *Escherichia coli* O157:H7 in Clark County, Washington, August 1999. Arch. Pediatr. Adolesc. Med. **157:**1016-1021.

14. **Buzby, J. C., Roberts, T., Lin, C.-T. J. and MacDonald J. M.** 1996. Bacterial foodborne disease: medical costs and productivity losses. Agricultural Economic Report No. 741. Washington: Economic Research Service, U.S. Department of Agriculture.

15. **Carter, A. O., A. A. Borczyk, J. A. Carlson, B. Harvey, J. C. Hockin, M. A. Karmali, C. Krishnan, D. A. Korn, and H. Lior.** 1987. A severe outbreak of *Escherichia coli* O157:H7 - associated hemorrhagic colitis in a nursing home. N. Engl. J. Med. **317:**1496-500.

16. **CDC.** 2007. *Escherichia coli* O157:H7 infection associated with drinking raw milk--Washington and Oregon, November-December 2005. Morb. Mortal. Wkly. Rep. **56:**165-167.

17. **CDC.** 1997. *Escherichia coli* O157:H7 infections associated with eating a nationally distributed commercial brand of frozen ground beef patties and burgers - Colorado, 1997. Mortal. Wkly. Rep. **46:**777-778.

18. **CDC.** 2005. *Escherichia coli* O157:H7 infections associated with ground beef from a U.S. military installation - Okinawa, Japan, February 2004. Morb. Mortal. Wkly. Rep. **54:**40-42.

19. **CDC.** 1995. *Escherichia coli* O157:H7 outbreak at a summer camp - Virginia, 1994. Morb. Mortal. Wkly. Rep. **44:**419-421.

20. **CDC.** 1995. *Escherichia coli* O157:H7 outbreak linked to commercially distributed dry-cured salami - Washington and California, 1994. Morb. Mortal. Wkly. Rep. **44:**157-160.

21. **CDC.** 1994. *Escherichia coli* O157:H7 outbreak linked to home-cooked hamburger - California, July 1993. Morb. Mortal. Wkly. Rep. **43:**213-216.

22. **CDC.** 1991. Foodborne outbreak of gastroenteritis caused by *Escherichia coli* O157:H7 - North Dakota, 1990. Morb. Mortal. Wkly. Rep. **40:**265-267.

23. **CDC.** 2008. Investigation of multistate outbreak of *E. coli* O157:H7 infections. http://www.cdc.gov/ecoli/june2008outbreak/index.html.

24. **CDC.** 2007. Investigation of outbreak of human infections caused by *E. coli* O157:H7. http://www.cdc.gov/ecoli/2007/october/103107.html.

25. **CDC.** 1996. Lake-associated outbreak of *Escherichia coli* O157:H7 - Illinois, 1995. Morb. Mortal. Wkly. Rep. **45:**437-439.

26. **CDC.** 2007. Multistate outbreak of *E. coli* O157 infections linked to Topp's brand ground beef patties. http://www.cdc.gov/ecoli/2007/october/100207.html.

27. **CDC.** 2006. Multistate outbreak of *E. coli* O157 infections, November-December 2006. http://www.cdc.gov/ecoli/2006/december/121406.htm.

28. **CDC.** 2002. Multistate outbreak of *Escherichia coli* O157:H7 infections associated with eating ground beef -United States, June-July 2002. Morb. Mortal. Wkly. Rep. **51:**637-639.

29. **CDC.** 1999. Outbreak of *Escherichia coli* O157:H7 and *Campylobacter* among attendees of the Washington County Fair-New York, 1999. Morb. Mortal. Wkly. Rep. **48:**803-805.

30. **CDC.** 1996. Outbreak of *Escherichia coli* O157:H7 infection -Georgia and Tennessee, June 1995. Morb. Mortal. Wkly. Rep. **45:**249-251.

31. **CDC.** 2000. Outbreak of *Escherichia coli* O157:H7 infection associated with eating fresh cheese curds - Wisconsin, June 1998. Morb. Mortal. Wkly. Rep. **49:**911-913.

32. **CDC.** 1996. Outbreak of *Escherichia coli* O157:H7 infections associated with drinking unpasteurized commercial apple juice - British Columbia, California, Colorado, and Washington, October 1996. Morb. Mortal. Wkly. Rep. **45:**975.

33. **CDC.** 2005. Outbreaks of *Escherichia coli* O157:H7 associated with petting zoos - North Carolina, Florida, and Arizona, 2004 and 2005. Morb. Mortal. Wkly. Rep. **54:**1277-1280.

34. **CDC.** 1997. Outbreaks of *Escherichia coli* O157:H7 infection and cryptosporidiosis associated with drinking unpasteurized apple cider - Connecticut and New York, October 1996. Morb. Mortal. Wkly. Rep. **46:**4-8.

35. **CDC.** 2001. Outbreaks of *Escherichia coli* O157:H7 infections among children associated with farm visits - Pennsylvania and Washington, 2000. Morb. Mortal. Wkly. Rep. **50:**293-297.

36. **CDC.** 1993. Preliminary report: foodborne outbreak of *Escherichia coli* O157:H7 infections from hamburgers - western United States, 1993. Morb. Mortal. Wkly. Rep. **42:**85-86.

37. **CDC.** 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers - western United States, 1992-1993. Morb. Mortal. Wkly. Rep. **42:**258-263.

38. **Cody, S. H., M. K. Glynn, J. A. Farrar, K. L. Cairns, P. M. Griffin, J. Kobayashi, M. Fyfe, R. Hoffman, A. S. King, J. H. Lewis, B. Swaminathan, R. G. Bryant, and D. J. Vugia.** 1999. An outbreak of *Escherichia coli* O157:H7 infection from unpasteurized commercial apple juice. Ann. Intern. Med. **130:**202-209.

39. **Cooley, M., D. Carychao, L. Crawford-Miksza, M. T. Jay, C. Myers, C. Rose, C. Keys, J. Farrar, and R. E. Mandrell.** 2007. Incidence and tracking of *Escherichia coli* O157:H7 in a major produce production region in California. PLoS ONE **2:**e1159.

40. **Crump, J. A., A. C. Sulka, A. J. Langer, C. Schaben, A. S. Crielly, R. Gage, M. Baysinger, M. Moll, G. Withers, D. M. Toney, S. B. Hunter, R. M. Hoekstra, S. K. Wong, P. M. Griffin, and T. J. Van Gilder.** 2002. An outbreak of *Escherichia coli* O157:H7 infections among visitors to a dairy farm. N. Engl. J. Med. **347:**555-560.

41. **Delahay, R. M., G. Frankel, and S. Knutton.** 2001. Intimate interactions of enteropathogenic *Escherichia coli* at the host cell surface. Curr. Opin. Infect. Dis. **14:**559-565.

42. **Donnenberg, M. S., S. Tzipori, M. L. McKee, A. D. O'Brien, J. Alroy, and J. B. Kaper.** 1993. The role of the *eae* gene of enterohemorrhagic *Escherichia coli* in intimate attachment *in vitro* and in a porcine model. J. Clin. Invest. **92:**1418-1424.

43. **Feng, P.** 1995. *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. Emerg. Infect. Dis. **1:**47-52.

44. **Feng, P., K. A. Lampel, H. Karch, and T. S. Whittam.** 1998. Genotypic and phenotypic changes in the emergence of *Escherichia coli* O157:H7. J. Infect. Dis. **177:**1750-1753.

45. **Foley, S. L., S. Simjee, J. Meng, D. G. White, P. F. McDermott, and S. Zhao.** 2004. Evaluation of molecular typing methods for *Escherichia coli* O157:H7 isolates from cattle, food, and humans. J. Food. Prot. **67:**651-657.

46. **Friedman, M. S., T. Roels, J. E. Koehler, L. Feldman, W. F. Bibb, and P. Blake.** 1999. *Escherichia coli* O157:H7 outbreak associated with an improperly chlorinated swimming pool. Clin. Infect. Dis. **29:**298-303.

47. **Frost, J. A., H. R. Smith, G. A. Willshaw, S. M. Scotland, R. J. Gross, and B. Rowe.** 1989. Phage-typing of vero-cytotoxin (VT) producing *Escherichia coli* O157 isolated in the United Kingdom. Epidemiol. Infect. **103:**73-81.

48. **Funamori, Y., Y. Fujinaga, K. Yokota, K. Inoue, Y. Hirai, K. Oguma, S. Kira, and K. Taketa.** 1999. Molecular typing of enterohemorrhagic *Escherichia coli* O157:H7 isolated in Okayama Prefecture using pulsed field gel electrophoresis and random amplification of polymorphic DNA. Acta. Med. Okayama **53:**193-200.

49. **Gasser, C., E. Gautier, A. Steck, R. E. Siebenmann, and R. Oechslin.** 1955. Hemolytic-uremic syndrome: bilateral necrosis of the renal cortex in acute acquired hemolytic anemia. Schweiz. Med. Wochenschr. **85:**905-909.

50. **Goering, R. V.** 1993. Molecular epidemiology of nosocomial infection: analysis of chromosomal restriction fragment patterns by pulsed-field gel electrophoresis. Infect. Control. Hosp. Epidemiol. **14:**595-600.

51. **Goode, B., C. O'Reilly, J. Dunn, K. Fullerton, S. Smith, G. Ghneim, J. Keen, L. Durso, M. Davies, and S. Montgomery.** 2009. Outbreak of *Escherichia coli* O157: H7 infections after petting zoo visits, North Carolina State Fair, October-November 2004. Arch. Pediatr. Adolesc. Med.**163:**42-48.

52. **Gouveia, S., M. E. Proctor, M. S. Lee, J. B. Luchansky, and C. W. Kaspar.** 1998. Genomic comparisons and Shiga toxin production among *Escherichia coli* O157:H7 isolates from a day care center outbreak and sporadic cases in southeastern Wisconsin. J. Clin. Microbiol. **36:**727-733.

53. **Grant, J., A. M. Wendelboe, A. Wendel, B. Jepson, P. Torres, C. Smelser, and R. T. Rolfs.** 2008. Spinach-associated *Escherichia coli* O157:H7 outbreak, Utah and New Mexico, 2006. Emerg. Infect. Dis. **14:**1633-1636.

54. **Grif, K., H. Karch, C. Schneider, F. D. Daschner, L. Beutin, T. Cheasty, H. Smith, B. Rowe, M. P. Dierich, and F. Allerberger.** 1998. Comparative study of five different techniques for epidemiological typing of *Escherichia coli* O157. Diagn. Microbiol. Infect. Dis. **32:**165-176.

55. **Hancock, D. D., T. E. Besser, M. L. Kinsel, P. I. Tarr, D. H. Rice, and M. G. Paros.** 1994. The prevalence of *Escherichia coli* O157:H7 in dairy and beef cattle in Washington State. Epidemiol. Infect. **113:**199-207.

56. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11-22.

57. **Heir, E., B. A. Lindstedt, T. Vardund, Y. Wasteson, and G. Kapperud.** 2000. Genomic fingerprinting of shigatoxin-producing *Escherichia coli* (STEC) strains: comparison of pulsed-field gel electrophoresis (PFGE) and fluorescent amplified-fragment-length polymorphism (FAFLP). Epidemiol. Infect. **125:**537-548.

58. **Hilborn, E. D., J. H. Mermin, P. A. Mshar, J. L. Hadler, A. Voetsch, C. Wojtkunski, M. Swartz, R. Mshar, M. A. Lambert-Fair, J. A. Farrar, M. K. Glynn, and L. Slutsker.** 1999. A multistate outbreak of *Escherichia coli* O157:H7 infections associated with consumption of mesclun lettuce. Arch. Intern. Med. **159:**1758-1764.

59. **Hilborn, E. D., P. A. Mshar, T. R. Fiorentino, Z. F. Dembek, T. J. Barrett, R. T. Howard, and M. L. Cartter.** 2000. An outbreak of *Escherichia coli* O157:H7 infections and haemolytic uraemic syndrome associated with consumption of unpasteurized apple cider. Epidemiol. Infect. **124:**31-36.

60. **Hill, C. W.** 1999. Large genomic sequence repetitions in bacteria: lessons from rRNA operons and Rhs elements. Res. Microbiol. **150:**665-674.

61. **Hill, C. W., C. H. Sandt, and D. A. Vlazny.** 1994. Rhs elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. Mol. Microbiol. **12:**865-871.

62. **Honish, L., I. Zazulak, R. Mahabeer, K. Krywiak, R. Leyland, N. Hislop, and L. Chui.** 2007. Outbreak of *Escherichia coli* O157:H7 gastroenteritis associated with consumption of beef donairs, Edmonton, Alberta, May-June 2006. Can. Commun. Dis. Rep. **33:**14-19.

63. **Hyytia-Trees, E., S. C. Smole, P. A. Fields, B. Swaminathan, and E. M. Ribot.** 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 (STEC O157). Foodborne Pathog. Dis. **3:**118-131.

64. **Iguchi, A., R. Osawa, J. Kawano, A. Shimizu, J. Terajima, and H. Watanabe.** 2002. Effects of repeated subculturing and prolonged storage at room temperature of enterohemorrhagic *Escherichia coli* O157:H7 on pulsed-field gel electrophoresis profiles. J. Clin. Microbiol. **40:**3079-3081.

65. **Jay, M. T., V. Garrett, J. C. Mohle-Boetani, M. Barros, J. A. Farrar, R. Rios, S. Abbott, R. Sowadsky, K. Komatsu, R. Mandrell, J. Sobel, and S. B. Werner.** 2004. A multistate outbreak of *Escherichia coli* O157:H7 infection linked to consumption of beef tacos at a fast-food restaurant chain. Clin. Infect. Dis. **39:**1-7.

66. **Johnson, J. M., S. D. Weagant, K. C. Jinneman, and J. L. Bryant.** 1995. Use of pulsed-field gel electrophoresis for epidemiological study of *Escherichia coli* O157:H7 during a food-borne outbreak. Appl. Environ. Microbiol. **61:**2806-2808.

67. **Karmali, M. A.** 1989. Infection by verocytotoxin-producing *Escherichia coli*. Clin. Microbiol. Rev. **2:**15-38.

68. **Karmali, M. A., G. S. Arbus, N. Ish-Shalom, P. C. Fleming, D. Malkin, M. Petric, R. Cheung, S. Louie, G. R. Humphreys, and M. Strachan.** 1988. A family outbreak of hemolytic-uremic syndrome associated with verotoxin-producing *Escherichia coli* serotype O157:H7. Pediatr. Nephrol. **2:**409-414.

69. **Karmali, M. A., B. T. Steele, M. Petric, and C. Lim.** 1983. Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. Lancet **1:**619-620.

70. **Katsuda, K., T. Sameshima, M. Akiba, and M. Nakazawa.** 1998. Characterisation of *Escherichia coli* O157:H7 (H-) derived from cattle using random amplification of polymorphic DNA analysis. Res. Vet. Sci. **65:**177-178.

71. **Keene, W. E., K. Hedberg, D. E. Herriott, D. D. Hancock, R. W. McKay, T. J. Barrett, and D. W. Fleming.** 1997. A prolonged outbreak of *Escherichia coli* O157:H7 infections caused by commercially distributed raw milk. J. Infect. Dis. **176:**815-818.

72. **Keene, W. E., J. M. McAnulty, F. C. Hoesly, L. P. Williams, Jr., K. Hedberg, G. L. Oxman, T. J. Barrett, M. A. Pfaller, and D. W. Fleming.** 1994. A swimming-associated outbreak of hemorrhagic colitis caused by *Escherichia coli* O157:H7 and *Shigella sonnei*. N. Engl. J. Med. **331:**579-584.

73. **Keene, W. E., E. Sazie, J. Kok, D. H. Rice, D. D. Hancock, V. K. Balan, T. Zhao, and M. P. Doyle.** 1997. An outbreak of *Escherichia coli* O157:H7 infections traced to jerky made from deer meat. JAMA **277:**1229-1231.

74. **Keusch, G. T., G. F. Grady, L. J. Mata, and J. McIver.** 1972. The pathogenesis of *Shigella* diarrhea. I. Enterotoxin production by *Shigella dysenteriae* I. J Clin Invest **51:**1212-1218.

75. **Khakhria, R., D. Duck, and H. Lior.** 1990. Extended phage-typing scheme for *Escherichia coli* O157:H7. Epidemiol. Infect. **105:**511-520.

76. **Kim, J., J. Nietfeldt, and A. K. Benson.** 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc. Natl. Acad. Sci. U S A **96:**13288-13293.

77. **King, L. A., A. Mailles, P. Mariani-Kurkdjian, C. Vernozy-Rozand, M. P. Montet, F. Grimont, N. Pihier, H. Devalk, F. Perret, E. Bingen, E. Espie, and V. Vaillant.** 2009. Community-wide outbreak of *Escherichia coli* O157:H7 associated with consumption of frozen beef burgers. Epidemiol. Infect. **137:**889-896.

78. **Krause, U., F. M. Thomson-Carter, and T. H. Pennington.** 1996. Molecular epidemiology of *Escherichia coli* O157:H7 by pulsed-field gel electrophoresis and comparison with that by bacteriophage typing. J. Clin. Microbiol. **34:**959-961.

79. **Kudva, I. T., P. G. Hatfield, and C. J. Hovde.** 1996. *Escherichia coli* O157:H7 in microbial flora of sheep. J. Clin. Microbiol. **34:**431-433.

80. **Laine, E. S., J. M. Scheftel, D. J. Boxrud, K. J. Vought, R. N. Danila, K. M. Elfering, and K. E. Smith.** 2005. Outbreak of *Escherichia coli* O157:H7 infections associated with nonintact blade-tenderized frozen steaks sold by door-to-door vendors. J. Food Prot. **68:**1198-1202.

81. **Laing, C., C. Pegg, D. Yawney, K. Ziebell, M. Steele, R. Johnson, J. E. Thomas, E. N. Taboada, Y. Zhang, and V. P. Gannon.** 2008. Rapid determination of *Escherichia coli* O157:H7 lineage types and molecular subtypes by using comparative genomic fingerprinting. Appl. Environ. Microbiol. **74:**6606-6615.

82. **Lin, R. J., M. Capage, and C. W. Hill.** 1984. A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. **177:**1-18.

83. **Lindstedt, B. A., E. Heir, E. Gjernes, T. Vardund, and G. Kapperud.** 2003. DNA fingerprinting of Shiga-toxin producing *Escherichia coli* O157 based on Multiple-locus variable-number tandem-repeats analysis (MLVA). Ann. Clin. Microbiol. Antimicrob. **2:**12.

84. **Lindstedt, B. A., T. Vardund, and G. Kapperud.** 2004. Multiple-locus variable-number tandem-repeats analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. J. Microbiol. Methods **58:**213-222.

85. **Ludwig, K., H. Ruder, M. Bitzan, S. Zimmermann, and H. Karch.** 1997. Outbreak of *Escherichia coli* O157:H7 infection in a large family. Eur. J. Clin. Microbiol. Infect. Dis. **16:**238-241.

86. **Macdonald, C., J. Drew, R. Carlson, S. Dzogan, S. Tataryn, A. Macdonald, A. Ali, R. Amhed, R. Easy, C. Clark, and F. Rodgers.** 2000. Outbreak of *Escherichia coli* O157:H7 leading to the recall of retail ground beef--Winnipeg, Manitoba, May 1999. Can. Commun. Dis. Rep. **26:**109-111.

87. **MacDonald, D. M., M. Fyfe, A. Paccagnella, A. Trinidad, K. Louie, and D. Patrick.** 2004. *Escherichia coli* O157:H7 outbreak linked to salami, British Columbia, Canada, 1999. Epidemiol. Infect. **132:**283-289.

88. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U S A **95:**3140-3145.

89. **Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U S A **105:**4868-4873.

90. **McDaniel, T. K., and J. B. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399-407.

91. **McKee, M. L., A. R. Melton-Celsa, R. A. Moxley, D. H. Francis, and A. D. O'Brien.** 1995. Enterohemorrhagic *Escherichia coli* O157:H7 requires intimin to colonize the gnotobiotic pig intestine and to adhere to HEp-2 cells. Infect. Immun. **63:**3739-3744.

92. **McNulty, C., J. Thompson, B. Barrett, L. Lord, C. Andersen, and I. S. Roberts.** 2006. The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. Mol. Microbiol. **59:**907-922.

93. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5:**607-625.

94. **Meng, J., S. Zhao, T. Zhao, and M. P. Doyle.** 1995. Molecular characterisation of *Escherichia coli* O157:H7 isolates by pulsed-field gel electrophoresis and plasmid DNA analysis. J. Med. Microbiol. **42:**258-263.

95. **Michino, H., K. Araki, S. Minami, S. Takaya, N. Sakai, M. Miyazaki, A. Ono, and H. Yanagawa.** 1999. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am. J. Epidemiol. **150:**787-796.

96. **Nakazawa, M., and M. Akiba.** 1999. Swine as a potential reservoir of shiga toxin-producing *Escherichia coli* O157:H7 in Japan. Emerg. Infect. Dis. **5:**833-834.

97. **Nataro, J. P., and J. B. Kaper.** 1998. Diarrheagenic *Escherichia coli*. Clin. Microbiol. Rev. **11:**142-201.

98. **Noller, A. C., M. C. McEllistrem, A. G. Pacheco, D. J. Boxrud, and L. H. Harrison.** 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. J. Clin. Microbiol. **41:**5389-5397.

99. **Noller, A. C., M. C. McEllistrem, K. A. Shutt, and L. H. Harrison.** 2006. Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Microbiol. **44:**374-377.

100. **Noller, A. C., M. C. McEllistrem, O. C. Stine, J. G. Morris, Jr., D. J. Boxrud, B. Dixon, and L. H. Harrison.** 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J. Clin. Microbiol. **41:**675-679.

101. **O'Brien, A. D., and R. K. Holmes.** 1987. Shiga and Shiga-like toxins. Microbiol. Rev. **51:**206-220.

102. **O'Brien, A. D., T. A. Lively, T. W. Chang, and S. L. Gorbach.** 1983. Purification of *Shigella dysenteriae* 1 (Shiga)-like toxin from *Escherichia coli* O157:H7 strain associated with haemorrhagic colitis. Lancet **2:**573.

103. **Orskov, F., I. Orskov, and J. A. Villar.** 1987. Cattle as reservoir of verotoxin-producing *Escherichia coli* O157:H7. Lancet **2:**276.

104. **Ostroff, S. M., P. M. Griffin, R. V. Tauxe, L. D. Shipman, K. D. Greene, J. G. Wells, J. H. Lewis, P. A. Blake, and J. M. Kobayashi.** 1990. A statewide outbreak of *Escherichia coli* O157:H7 infections in Washington State. Am. J. Epidemiol. **132:**239-247.

105. **Ozeki, Y., T. Kurazono, A. Saito, T. Kishimoto, and M. Yamaguchi.** 2003. A diffuse outbreak of enterohemorrhagic *Escherichia coli* O157:H7 related to the Japanese-style pickles in Saitama, Japan. Kansenshogaku Zasshi **77:**493-498.

106. **Panaro, L., D. Cooke, and A. Borczyk.** 1990. Outbreak of *Escherichia coli* O157:H7 in a nursing home--Ontario. Can. Dis. Wkly. Rep. **16:**90-92.

107. **Paunio, M., R. Pebody, M. Keskimaki, M. Kokki, P. Ruutu, S. Oinonen, V. Vuotari, A. Siitonen, E. Lahti, and P. Leinikki.** 1999. Swimming-associated outbreak of *Escherichia coli* O157:H7. Epidemiol. Infect. **122:**1-5.

108. **Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen.** 2007. Genes under positive selection in *Escherichia coli*. Genome Res. **17:**1336-1343.

109. **Preston, M., A. Borczyk, and R. Davidson.** 1997. Hospital outbreak of *Escherichia coli* O157:H7 associated with a rare phage type--Ontario. Can. Commun. Dis. Rep. **23:**33-36; discussion 36-37.

110. **Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow.** 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerg. Infect. Dis. **11:**603-609.

111. **Reida, P., M. Wolff, H. W. Pohls, W. Kuhlmann, A. Lehmacher, S. Aleksic, H. Karch, and J. Bockemuhl.** 1994. An outbreak due to enterohaemorrhagic *Escherichia coli* O157:H7 in a children day care centre characterized by person-to-person transmission and environmental contamination. Zentralbl. Bakteriol. **281:**534-543.

112. **Riley, L. W.** 1987. The epidemiologic, clinical, and microbiologic features of hemorrhagic colitis. Annu. Rev. Microbiol. **41:**383-407.

113. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681-685.

114. **Robinson, C. M., J. F. Sinclair, M. J. Smith, and A. D. O'Brien.** 2006. Shiga toxin of enterohemorrhagic *Escherichia coli* type O157:H7 promotes intestinal colonization. Proc. Natl. Acad. Sci. U S A **103:**9667-9672.

115. **Rodrigue, D. C., E. E. Mast, K. D. Greene, J. P. Davis, M. A. Hutchinson, J. G. Wells, T. J. Barrett, and P. M. Griffin.** 1995. A university outbreak of *Escherichia coli* O157:H7 infections associated with roast beef and an unusually benign clinical course. J. Infect. Dis. **172:**1122-1125.

116. **Ryan, C. A., R. V. Tauxe, G. W. Hosek, J. G. Wells, P. A. Stoesz, H. W. McFadden, Jr., P. W. Smith, G. F. Wright, and P. A. Blake.** 1986. *Escherichia coli* O157:H7 diarrhea in a nursing home: clinical, epidemiological, and pathological findings. J. Infect. Dis. **154:**631-638.

117. **Sartz, L., B. De Jong, M. Hjertqvist, L. Plym-Forshell, R. Alsterlund, S. Lofdahl, B. Osterman, A. Stahl, E. Eriksson, H. B. Hansson, and D. Karpman.** 2008. An outbreak of *Escherichia coli* O157:H7 infection in southern Sweden associated with consumption of fermented sausage; aspects of sausage production that increase the risk of contamination. Epidemiol. Infect. **136:**370-380.

118. **Schwartz, D. C., and C. R. Cantor.** 1984. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. Cell **37:**67-75.

119. **Scotland, S. M., G. A. Willshaw, H. R. Smith, and B. Rowe.** 1987. Properties of strains of *Escherichia coli* belonging to serogroup O157 with special reference to production of vero cytotoxins VT1 and VT2. Epidemiol. Infect. **99:**613-624.

120. **Shima, K., Y. Wu, N. Sugimoto, M. Asakura, K. Nishimura, and S. Yamasaki.** 2006. Comparison of a PCR-restriction fragment length polymorphism (PCR-RFLP) assay to pulsed-field gel electrophoresis to determine the effect of repeated subculture and prolonged storage on RFLP patterns of Shiga toxin-producing *Escherichia coli* O157:H7. J. Clin. Microbiol. **44:**3963-3968.

121. **Siegler, R. L., T. G. Obrig, T. J. Pysher, V. L. Tesh, N. D. Denkers, and F. B. Taylor.** 2003. Response to Shiga toxin 1 and 2 in a baboon model of hemolytic uremic syndrome. Pediatr. Nephrol. **18:**92-96.

122. **Sjogren, R., R. Neill, D. Rachmilewitz, D. Fritz, J. Newland, D. Sharpnack, C. Colleton, J. Fondacaro, P. Gemski, and E. Boedeker.** 1994. Role of Shiga-like toxin I in bacterial enteritis: comparison between isogenic *Escherichia coli* strains induced in rabbits. Gastroenterology **106:**306-317.

123. **Struelens, M. J.** 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin. Microbiol. Infect. **2:**2-11.

124. **Sutcliffe, P., L. Picard, B. Fortin, D. Malaviarachchi, J. Hohenadel, and B. O'Donnell.** 2004. *Escherichia coli* O157:H7 outbreak at a summer hockey camp, Sudbury 2004. Can. Commun. Dis. Rep. **30:**189-194.

125. **Swerdlow, D. L., B. A. Woodruff, R. C. Brady, P. M. Griffin, S. Tippen, H. D. Donnell, Jr., E. Geldreich, B. J. Payne, A. Meyer, Jr., and J. G. Wells.** 1992. A waterborne outbreak in Missouri of *Escherichia coli* O157:H7 associated with bloody diarrhea and death. Ann. Intern. Med. **117:**812-819.

126. **Tamblyn, S., J. deGrosbois, D. Taylor, and J. Stratton.** 1999. An outbreak of *Escherichia coli* O157:H7 infection associated with unpasteurized non-commercial, custom-pressed apple cider--Ontario, 1998. Can. Commun. Dis. Rep. **25:**113-117; discussion 117-120.

127. **Tarr, P. I.** 1995. *Escherichia coli* O157:H7: clinical, diagnostic, and epidemiological aspects of human infection. Clin. Infect. Dis. **20:**1-8; quiz 9-10.

128. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233-2239.

129. **Tesh, V. L., J. A. Burris, J. W. Owens, V. M. Gordon, E. A. Wadolkowski, A. D. O'Brien, and J. E. Samuel.** 1993. Comparison of the relative toxicities of Shiga-like toxins type I and type II for mice. Infect. Immun. **61:**3392-3402.

130. **Tsai, T. Y., W. C. Luo, F. T. Wu, and T. M. Pan.** 2005. Molecular subtyping for *Escherichia coli* O157: H7 isolated in Taiwan. Microbiol. Immunol. **49:**579-588.

131. **Tzipori, S., F. Gunzer, M. S. Donnenberg, L. de Montigny, J. B. Kaper, and A. Donohue-Rolfe.** 1995. The role of the *eaeA* gene in diarrhea and neurological complications in a gnotobiotic piglet model of enterohemorrhagic *Escherichia coli* infection. Infect. Immun. **63:**3621-3627.

132. **Van Diemen, P. M., F. Dziva, M. P. Stevens, and T. S. Wallis.** 2005. Identification of enterohemorrhagic *Escherichia coli* O26:H- genes required for intestinal colonization in calves. Infect. Immun. **73:**1735-1743.

133. **Vogt, R. L., and L. Dippold.** 2005. *Escherichia coli* O157:H7 outbreak associated with consumption of ground beef, June-July 2002. Public Health Rep. **120:**174-178.

134. **Wall, P. G., R. J. McDonnell, G. K. Adak, T. Cheasty, H. R. Smith, and B. Rowe.** 1996. General outbreaks of vero cytotoxin producing *Escherichia coli* O157 in England and Wales from 1992 to 1994. Commun. Dis. Rep. Rev. **6:**R26-33.

135. **Watanabe, Y., K. Ozasa, J. H. Mermin, P. M. Griffin, K. Masuda, S. Imashuku, and T. Sawada.** 1999. Factory outbreak of *Escherichia coli* O157:H7 infection in Japan. Emerg. Infect. Dis. **5:**424-428.

136. **Wells, J. G., L. D. Shipman, K. D. Greene, E. G. Sowers, J. H. Green, D. N. Cameron, F. P. Downes, M. L. Martin, P. M. Griffin, S. M. Ostroff, M. E. Potter, R. V. Tauxe, and I. K. Wachsmuth.** 1991. Isolation of *Escherichia coli* serotype O157:H7 and other Shiga-like-toxin-producing *E. coli* from dairy cattle. J. Clin. Microbiol. **29:**985-989.

137. **Welsh, J., and M. McClelland.** 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. **18:**7213-7218.

138. **Wendel, A. M., D. H. Johnson, U. Sharapov, J. Grant, J. R. Archer, T. Monson, C. Koschmann, and J. P. Davis.** 2009. Multistate outbreak of *Escherichia coli* O157:H7 infection associated with consumption of packaged spinach, August-September 2006: the Wisconsin investigation. Clin. Infect. Dis. **48:**1079-1086.

139. **Williams, J. G., A. R. Kubelik, K. J. Livak, J. A. Rafalski, and S. V. Tingey.** 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic. Acids Res. **18:**6531-6535.

140. **Williams, R. C., S. Isaacs, M. L. Decou, E. A. Richardson, M. C. Buffett, R. W. Slinger, M. H. Brodsky, B. W. Ciebin, A. Ellis, and J. Hockin.** 2000. Illness outbreak associated with *Escherichia coli* O157:H7 in Genoa salami. *E. coli* O157:H7 Working Group. CMAJ **162:**1409-1413.

141. **Willshaw, G. A., J. Thirlwell, A. P. Jones, S. Parry, R. L. Salmon, and M. Hickey.** 1994. Vero cytotoxin-producing *Escherichia coli* O157 in beefburgers linked to an outbreak of diarrhoea, haemorrhagic colitis and haemolytic uraemic syndrome in Britain. Lett. Appl. Microbiol. **19:**304-307.

142. **Zabeau, M. and P. Vos** 1993. European Patent Application, EP0534858.

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

144. **Zhao, S., S. E. Mitchell, J. Meng, S. Kresovich, M. P. Doyle, R. E. Dean, A. M. Casa, and J. W. Weller.** 2000. Genomic typing of *Escherichia coli* O157:H7 by semi-automated fluorescent AFLP analysis. Microbes Infect. **2:**107-113.

CHAPTER THREE

REARRANGEMENT HOT SPOT (*rhs*) GENES ARE POTENTIAL MARKERS FOR MULTILOCUS SEQUENCE TYPING OF *ESCHERICHIA COLI* O157:H7 STRAINS

K. Liu, S. J. Knabel, E. G. Dudley*

Department of Food Science, The Pennsylvania State University, University Park, Pennsylvania 16802

* Corresponding author. Mailing address: 326 Food Science Building, The Pennsylvania State University, University Park, PA 16802, US. Phone: (814) 867-0439.

Email: egd100@psu.edu

3.1 Abstract

Escherichia coli O157:H7 is an important foodborne pathogen, notorious for its low infectious dose and its capacity to cause severe diseases such as hemorrhagic colitis and hemolytic uremic syndrome. Molecular subtyping methods are powerful tools for microbial source tracking during outbreaks. Pulsed-field gel electrophoresis is currently the "gold standard" molecular method for subtyping *E. coli* O157:H7, however a sequence-based scheme such as multilocus sequence typing (MLST) would provide a more rapid, cost-effective and portable method. Previous publications reported that strains of *E. coli* O157:H7 are highly clonal, with little sequence variation found between strains. This hindered the development of an effective MLST scheme for subtyping *E. coli* O157:H7. In this study, several *rhs* genes were targeted, which were recently shown to be under the strongest positive selection of genes encoded within the *E. coli* genome. Eighteen *E. coli* O157:H7 strains from Lineage I and II, and 15 *E. coli* strains from 8 clades were included. Examination of these *rhs* genes revealed 44 polymorphic loci (PL) and 10 sequence types (STs) among the 18 lineage-strains, and 280 PL and 12 STs among the 15 clade-strains. Phylogenetic analysis using *rhs* genes generally agreed with the Lineage I and II classification defined previously. This study also showed *E. coli* O157:H7 strains from Clade 8 fall into Lineage I/II. Additionally, unique markers were found in *rhsA* and *rhsJ* that might be used to define Clade 8 and Clade 6. Therefore, *rhs* genes may be useful in discriminating strains and determining the phylogeny of *E. coli* O157:H7.

3.2 Introduction

Escherichia coli O157:H7 was first described in 1983 as the causative agent of a foodborne outbreak attributed to contaminated ground beef patties (35), and it has subsequently emerged as a very important foodborne pathogen. On average, an estimated 74,000 illnesses (30) and 17 outbreaks (34) are attributable to *E. coli* O157:H7 each year in the United States. *E. coli* O157:H7 has a low infectious dose, with previous research suggesting that ingestion of only 10 cells may cause illness (40, 43). Diseases caused by *E. coli* O157:H7, such as hemorrhagic colitis and hemolytic uremic syndrome can be very severe or even life-threatening.

Cattle are believed to be the main reservoir for *E. coli* O157:H7 (4, 14, 42), although other animals may also carry this organism (6, 20). Outbreaks may occur when foods such as ground beef patties and fresh produce come in contact with fecal material or contaminated animal hides (11, 20). In order to limit ongoing outbreaks and prevent future outbreaks, it is imperative to identify the sources of *E. coli* O157:H7 and its routes of transmission throughout the food system.

Molecular subtyping methods combined with traditional epidemiological approaches are powerful tools for identifying sources and routes of transmission of foodborne pathogens. An ideal molecular subtyping method should be time- and cost-effective, and should also produce data that are easy to interpret and compare between different laboratories. Most importantly, such methods should also possess high discriminatory power (i.e. able to separate unrelated strains) and high epidemiologic concordance (i.e. able to group related strains together) (17). Several approaches exist for subtyping *E. coli* O157:H7, such as bacteriophage typing (1), amplified fragment length polymorphism

(AFLP) (19), multiple-locus variable-number tandem repeats analysis (MLVA) (26), and pulsed-field gel electrophoresis (PFGE) (7). Among these methods, PFGE, a DNA-pattern-based method, is currently the "gold standard" approach widely used by public health laboratories. PFGE possesses high discriminatory power. It is generally applicable to all bacteria and it requires little knowledge of genome composition. However, PFGE suffers from several drawbacks. First, it is a labor intensive protocol that requires expert technical skill (17). Second, PFGE sometimes generates ambiguous banding patterns which are difficult to interpret, especially when a large number of isolates are studied on the same gel (24). Third, PFGE banding patterns are not easily comparable between laboratories (17). Fourth, PFGE banding patterns may change during laboratory transfer (18). Lastly, PFGE may not be suitable for phylogenetic analysis (23). The ability of a subtyping method to determine phylogenic relatedness is important, as not all members of the *E. coli* O157:H7 serotype are equally pathogenic to humans. For example, Kim et al. (21) and Zhang et al. (47) identified 3 lineages of *E. coli* O157:H7 designated Lineage I, Lineage II and Lineage I/II. Lineage I is suggested to be more virulent to humans (5, 21). More recently, Manning et al. (28) identified 9 genetically distinct clades of *E. coli* O157:H7, and proposed that Clade 8 represents a highly virulent group of *E. coli* O157:H7.

Multilocus sequence typing (MLST) is a subtyping method that discriminates between strains of a bacterial species by identifying DNA sequence differences in 6 - 8 targeted genes. MLST has several advantages for subtyping bacterial pathogens when compared to PFGE. For example, it is less labor intensive and generates unambiguous DNA sequence data that can be easily compared between laboratories. Furthermore, MLST is

suitable for phylogenetic analysis of bacterial pathogens, such as *E. coli* O157:H7 (28). Satisfactory MLST schemes exist for other bacterial pathogens (27, 45), however due to the lack of sequence variations in previously targeted gene markers (13, 32), no successful MLST approaches are reported for subtyping *E. coli* O157:H7 in epidemiological investigations. For example, Noller et al. (32) and Foley et al. (13) found a lack of sequence diversity among housekeeping genes and genes encoding outer membrane proteins and virulence factors in *E. coli* O157:H7 strains. These two studies also indicated that housekeeping genes and virulence genes may not be useful for revealing the short-term evolution of *E. coli* O157:H7 during outbreaks. Afterwards, Zhang et al. (46) used microarray-based whole genome scanning to investigate the extent to which single nucleotide polymorphisms (SNPs) contribute to the genomic diversity of *E. coli* O157:H7. They identified 906 SNPs among 11 *E. coli* O157:H7 strains, although 56% of these SNPs were found exclusively in two atypical O157 strains. Later Manning et al. (28) expanded on this work and developed a subtyping scheme based upon the interrogation of 96 putative SNP loci in 83 O157 genes. This approach was more discriminatory than previous MLST schemes and separated 528 O157 strains into 39 distinct SNP genotypes. However, a simpler approach that targets a few informative gene markers would be more feasible and applicable for rapid strain discrimination and phylogenetic determination in epidemiological investigations.

Therefore, we were interested in screening genomes of *E. coli* O157:H7 for these informative markers to develop an MLST scheme. A recent analysis of *E. coli* genomes predicted that rearrangement hot spot (*rhs*) genes are under the strongest positive selection of all coding sequences analyzed (33). Therefore, we hypothesized that these

genes would display significant sequence variations for subtyping O157 strains. The *rhs* genes are related by DNA sequence and were first discovered as elements mediating tandem duplication of the *glyS* locus in *E. coli* K-12 (25). The function(s) of these genes are currently unknown. *E. coli* O157:H7 strains may have multiple *rhs* genes encoded within their genomes. For example, there are 9 *rhs* genes within the genome of the prototypical *E. coli* O157:H7 strain Sakai and these genes are designated *rhsA, C, D, E, F, G, I, J* and *K* (Table 3.1) (15). Three of these 9 *rhs* genes, *rhsF, J* and *K* were previously studied by Zhang et al. (46), and a number of SNPs were identified among these 3 genes. However no studies have been conducted to comprehensively investigate *rhs* genes as markers in an MLST scheme for subtyping *E. coli* O157:H7.

The primary purpose of the present study was to investigate whether there are sufficient DNA sequence variations among *rhs* genes to develop an MLST approach for subtyping *E. coli* O157:H7. In this study, a greater level of DNA sequence variations was observed among *rhs* genes than in gene markers targeted in previous studies (13, 32). Furthermore, phylogenetic analysis using these *rhs* genes generally agreed with the lineage classification of *E. coli* O157:H7 strains defined by a previous study (21). The present study also reports for the first time evidence that *E. coli* O157:H7 strains from Clade 8 are classified as Lineage I/II, which is a different lineage from well-studied *E. coli* O157:H7 outbreak strains such as EDL933 and Sakai. Therefore, we suggest that outbreaks of *E. coli* O157:H7 are caused by two lineages of this pathogen - Lineage I and Lineage I/II.

3.3 Materials and methods

Bacterial strains. All *E. coli* O157:H7 strains used in this study are listed in Table 3.2. Bacterial strains were stored at -80°C in 10% glycerol. When needed, strains were streaked directly onto Sorbitol MacConkey agar (SMAC) plates made of Difco MacConkey Agar Base (Becton, Dickinson and Company, Sparks, MD) supplemented to 1% (w/v) D-Sorbitol (Alfa Aesar, Ward Hill, MA) and incubated at 37°C overnight.

***In silico* comparison of *E. coli* O157:H7 *rhs* genes.** DNA sequences for *rhs* genes were obtained from whole genome shotgun (WGS) sequences of 14 *E. coli* O157:H7 strains deposited in the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov). These strains are EC4024 (accession number ABJT00000000), EC4042 (accession number ABHM00000000), EC4045 (accession number ABHL00000000), EC4076 (accession number ABHQ00000000), EC4113 (accession number ABHP00000000), EC4115 (accession number CP001164), EC4196 (accession number ABHO00000000), EC4206 (accession number ABHK00000000), EC4401 (accession number ABHR00000000), EC4486 (accession number ABHS00000000), EC4501 (accession number ABHT00000000), EC508 (accession number ABHW00000000), EC869 (accession number ABHU00000000), and Sakai (accession number BA000007) (15, 22). At the time of manuscript preparation, these genomes were at various stages of assembly, and consisted of between one and several hundred contigs. The DNA sequences of the 9 *rhs* genes from *E. coli* O157:H7 strain Sakai were extracted using Artemis (Version 9.0) (37). The Basic Local Alignment Search Tool (3) was used to identify the corresponding *rhs* gene within each of the whole shotgun sequences. We excluded *rhsG* from the search as complete sequences of *rhsG* were not found in all WGS

sequences at that time. Since *rhs* genes share a high level of sequence identity to one another, each contig encoding an *rhs* gene was aligned against the Sakai genome using the Artemis Comparison Tool (Version 6.0) (8) to assure each gene was given the correct letter designation based upon its location within the genome. The *rhs* genes of the 14 strains were aligned using the Seqman program of the Lasergene software package (DNAStar, Madison, WI).

Amplification of *rhs* genes. Primers for amplifying and sequencing *rhs* genes were designed based upon the *E. coli* O157:H7 strain Sakai genome using Primer 3.0 (http://frodo.wi.mit.edu/). PCR templates were prepared by resuspending one colony from SMAC plates in 20 µl sterile distilled water. PCR reactions contained 1.0 - 1.5 µl of this cell suspension as template. Buffer, nucleotides, primer, and *Taq* polymerase (New England Biolabs, Ipswich, MA) concentrations recommended by the manufacturer were used. Primers and cycling conditions are listed in Table 3.3.

DNA Sequencing. After PCR, products for sequencing were prepared by adding 1/20 volume exonuclease I (10 U/µl, USB Corp. Cleveland, OH) and 1/20 volume of shrimp alkaline phosphatase (1 U/µl, USB Corp). The mixture was incubated at 37°C for 45 min to degrade the primers and unincorporated dNTPs and then at 80°C for 15 min to inactivate the enzymes. Afterwards, PCR products were sent to the Nucleic Acid Facility Center at the Pennsylvania State University (University Park, PA) for sequencing on an ABI Hitachi 3730XL DNA Analyzer. In order to obtain the full sequence of each *rhs* gene, sequencing was done with several sequencing primers whose products shared ca. 150 bp overlap. Sequencing primers for each *rhs* gene are listed in Table 3.4. Sequences of *rhs* genes were aligned and assembled using Seqman. Most of the length of the

consensus sequences reported are based upon the sequencing of one strand. However, in all cases where SNPs unique to one strain were identified, or where the four color chromatograms did not display unambiguous sequence reads, these regions were sequenced again using a reverse primer to generate double-stranded reads.

Lineage Specific Polymorphism Assay (LSPA). LSPA was carried out either in one single multiplex PCR according to Yang et al. (44) or two multiplex PCRs according to Ziebell et al. (48) when any of the 6 genomic markers was not amplified efficiently in one multiplex PCR. Templates were prepared as described above for the amplification of *rhs* genes. When LSPA was conducted using one single multiplex PCR, a 10 µl aliquot of PCR product was loaded onto a 6% polyacrylamide gel. When LSPA was conducted using two multiplex PCRs, equal volumes of PCR products from both reactions were mixed together, and a 10 µl aliquot of this mixture was loaded onto a 6% polyacrylamide gel. DNA fragments were visualized after ethidium bromide staining using an EC3 500 BioImaging Systems (UVP, Upland, CA)**.**

Phylogenetic analysis. SNPs in the 8 *rhs* genes of each *E. coli* O157:H7 strain (Table 3.6) were concatenated and the parsimony informative (PI) sites (31) were used to construct an unrooted neighbor-joining tree (1,000 replications) using MEGA 4.0 (39).

Nucleotide sequence accession number. Nucleotide sequences for *rhs* genes were deposited into GenBank under the accession numbers FJ839695 to FJ839816 and FJ871402 to FJ871406.

3.4 Results

***In silico* comparison of 8 *rhs* genes from 14 *E. coli* O157:H7 strains.** Previous attempts at identifying sequence variations within *E. coli* O157:H7 genes (13, 32) were initiated when only two full genome sequences were available. The recent online release of whole genome shotgun (WGS) sequence data for 13 additional strains (EC4206, EC4045, EC4042, EC4115, EC4196, EC4113, EC4076, EC4024, EC4401, EC4486, EC4501, EC508 and EC869) permitted a more comprehensive screen for single nucleotide polymorphisms (SNPs). While sequencing and/or assembling errors may exist within these WGS sequences, they still provided a powerful tool for testing hypotheses before sequencing a larger collection of strains. Eight of these 13 strains are from an outbreak traced to contaminated spinach in California in 2006 (10) and 2 from an outbreak traced to contaminated lettuce from Taco Bell restaurants in 2006 (9), which allowed us to investigate whether *rhs* genes have identical sequences within strains from the same outbreak. We began our study by comparing the sequences of 9 *rhs* genes (*rhsA*, *C*, *D*, *E*, *F*, *G*, *I*, *J* and *K*) from these 13 strains and from the full genome sequence previously reported for *E. coli* O157:H7 strain Sakai (15). Alignment of these gene sequences showed that there were 51 polymorphic loci (PL) within 7 of the *rhs* genes (*rhsA*, *C*, *D*, *E*, *F*, *I*, and *K*) (Table 3.5). No PL were identified in *rhsJ* between any of the strains analyzed. Sequences of *rhsG* were not analyzed, as complete sequences for this gene were not identified in any of the 13 WGS sequenced strains. Based upon the sequence variations (SNPs, insertions and deletions) 8 sequence types (STs) were evident among 12 strains (Table 3.5) (EC4501 and EC508 were not included in this comparison because complete sequences for *rhsC* were not found within the online database). The

2006 spinach outbreak strains contained 5 STs (defined as ST-I1, -I2, -I3, -I4, and -I5), and the Taco-Bell strains possessed 2 STs (ST-I2 and -I6) (Table 3.5). A few DNA sequence variations were observed within the same outbreak (Table 3.5), but these variations might be whole genome sequencing and/or assembling errors. Strain EC869 was ST-I7, and strain Sakai was ST-I8 (Table 3.5). These results suggest that sequence variations identified within *rhs* genes may be useful for discriminating between strains of *E. coli* O157:H7.

Comparison of 8 *rhs* genes from 18 lineage-strains. We next wanted to test whether sequence differences within *rhs* genes would permit discrimination between other strains of *E. coli* O157:H7. Strains were chosen from a previously described collection designated USA 40 (44) from the lab of Dr. Andrew Benson at the University of Nebraska-Lincoln (Table 3.2). Dr. Benson's lab defined two phylogenetic groups of *E. coli* O157:H7 designated Lineage I and Lineage II (21), and we initially sequenced the full length of *rhsA* from 9 Lineage I and 9 Lineage II strains (Table 3.6). SNPs identified with this collection of strains were found at the same nucleotide positions as the SNPs identified in the above *in silico* comparison (Table 3.5). Therefore, instead of sequencing the full length of the other *rhs* genes from these strains, we only sequenced regions surrounding the SNP loci identified in Table 3.5. Additionally, during preliminary sequencing of *rhs* genes from FDA517 and FRIK1997 (data not shown), we identified 8 additional SNPs between nucleotide position 612 to 807 in *rhsJ*. Therefore, this region of *rhsJ* was also sequenced in the remaining lineage strains. Partial sequencing of these 8 *rhs* genes revealed 44 SNP loci and 10 STs among the 18 strains analyzed (Table 3.6). Among Lineage II strains, 6 STs were evident: FRIK1985, FRIK2000, FRIK920, and

FRIK944 were ST-L1; FRIK1990, FRIK2001, NE037, FDA508, and FDA517 each possessed a unique ST (ST-L2, -L3, -L4, -L5 and -L6, respectively) (Table 3.6). Among Lineage I strains, 4 STs were evident: 93-001, 95-003, NE018 and FRIK1986 each was classified as ST-L7; FDA507, FDA518 and FRIK1275 shared ST-L8; FRIK523 and FRIK1997 were classified as ST-L9 and ST-L10, respectively (Table 3.6). Based upon our findings, we suggest *rhs* genes have sufficient sequence variations to distinguish between many members of this strain collection.

Phylogenetic analysis of lineage strains using 8 *rhs* genes. A phylogenetic tree based upon the parsimony informative sites in Table 3.6 was constructed (Fig. 3.1). Two distinct groups were identified, with all Lineage I strains and the 3 Lineage II strains of human origin (NE037, FDA508 and FDA517) in Group 1, and with the remaining 6 Lineage II strains, all of bovine origin, in Group 2. As NE037, FDA508 and FDA517 were more closely related to Lineage I than to Lineage II, a previously published assay designated lineage specific polymorphism assay (LSPA) (44) was conducted to further characterize these 18 strains (Table 3.8). The LSPA classifies strains of *E. coli* O157:H7 based upon the amplicon size obtained from PCR amplification of six genomic loci (*folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*). Generally one or two amplicon sizes are observed, and strains are classified into LSPA genotypes by a six-digit binary designation. Our LSPA results showed that the 9 Lineage I strains were all LSPA genotype 111111, and the 6 Lineage II strains of cattle origin were LSPA genotype 222222 (FRIK944, FRIK920, FRIK2000 and FRIK1990), 222221 (FRIK1985), and 222211 (FRIK2001) (Table 3.8). The 3 Lineage II strains of human origin were classified as LSPA genotype 211111 (FDA508 and FDA517) and as LSPA genotype 211131

(NE037) (Table 3.8).

Comparison of 7 *rhs* genes from 15 clade-strains. Recently, Manning et al. (28) used a 96-locus SNP typing scheme to identify 9 phylogenetic groups of *E. coli* O157:H7, and designated these groups as Clade 1 through 9. As sufficient DNA sequence variations were observed in *rhs* genes among lineage strains and phylogenetic analysis using *rhs* genes generally confirmed the phylogeny of the lineage strains determined previously (21), we next tested whether this method could be used to distinguish between strains from different clades. A strain collection comprising of at least one member of each clade except for Clade 5 was used. Clade 5 strains were not included, as research since the publication of the paper by Manning et al. (28) suggested that strains originally classified as Clade 5 may have been mixed cultures resulting in the distinct phylogeny observed (S. D. Manning, personal communication). Seven *rhs* genes (*rhsA*, *C*, *D*, *E*, *F*, *I* and *J*) were targeted. We excluded *rhsK* as only one SNP, found in strain FDA508 was identified previously (Table 3.6). The full lengths of 7 *rhs* genes from 15 clade-strains were obtained and sequence variations are summarized in Table 3.7 (refer to Supplementary Tables 1 through 7 for detailed information). Analysis of the 7 *rhs* genes revealed 12 STs among the 15 clade-strains, and these STs were designated ST-C1 through ST-C12. Sakai (Clade 1), TW10022 (Clade 1) and 93-111 (Clade 2) shared ST-C1 (Table 3.7). TW11039 and TW11052 (Clade 4) shared ST-C5 (Table 3.7). Each of the remaining strains possessed a unique ST (Table 3.7). We identified 280 PL, with *rhsA, C, D, E, F, I,* and *J* containing 23, 71, 60, 54, 62, 2, and 8 PL respectively (Appendix A Supplementary Table 1 to 7). Therefore, sufficient DNA sequence variations were also observed among these clades strains, and strains that belong to the same clade may not

necessarily share the same *rhs* sequence type.

Interestingly, 4 unique SNPs at nucleotide positions 3468, 3478, 3479 and 3516 were identified in *rhsA* of the Clade 8 strain M1-Spinach (Appendix A Supplementary Table 1). These 4 SNPs were also identified by *in silico* comparison of *E. coli* O157:H7 strains from the spinach and Taco-Bell outbreaks (Table 3.5). Riordan et al. (36) previously suggested that Clade 8 strains could be distinguished from other clade strains by the presence of an adenine residue at nucleotide position 539 of a hypothetical protein designated ECs2357. By this method, we confirmed the spinach and Taco-Bell outbreak strains were Clade 8 (data not shown) using *in silico* analysis. More interestingly, FDA508, the Lineage II strain of human origin also possessed these 4 SNPs in *rhsA* (Table 3.6) and it was classified as Clade 8 according to Riordan et al. (36) (data not shown).

***E. coli* O157:H7 strains from Clade 8 fall into Lineage I/II**. The correlation between the LSPA genotypes and lineages was first shown by Yang et al. (44), who classified LSPA genotype 111111 as Lineage I, and classified LSPA genotype 211111, 212111 and 222222 as Lineage II. These 4 LSPA genotypes comprised ca. 89% of isolates among a collection of 1429 strains. Later, Zhang et al. (47) classified *E. coli* O157:H7 strains with LSPA genotype 211111 as a third lineage - Lineage I/II. Previous studies showed that Lineage I and Lineage II strains are more commonly isolated from human and cattle, respectively (21). However Lineage I/II strains do not appear to show a host preference (44).

We noted from our data that the *rhsA* sequence of strain FDA508, which is characterized as LSPA genotype 211111 (Table 3.8) and Clade 8 in our study, is identical

to that from Clade 8 strains associated with the California spinach and the Taco Bell outbreaks (Table 3.5, 6 and Appendix A Supplementary Table 1). Based upon this finding, we hypothesized that Clade 8 strains would be LSPA genotype 211111 and would accordingly fall into Lineage I/II. The 6-gene LSPA (44) showed that the Clade 8 strain M1-spinach was LSPA genotype 211111 (Table 3.8 and Fig. 3.2). Furthermore, *in silico* analysis of the six LSPA alleles (*folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*) revealed that the 8 spinach and 2 Taco-Bell outbreak strains were also LSPA genotype 211111 (data not shown), and these strains were determined to be Clade 8 by *in silico* analysis of ECs2357 (36) (data not shown). Therefore, the Clade 8 strains in our study belong to a separate lineage (designated Lineage I/II) than other well-studied outbreak strains such as Sakai, EDL933 and 93-111, which are all LSPA genotype 111111.

Correlation of *E. coli* O157:H7 LSPA genotypes with clade designation. We next asked whether strains from other clades are also LSPA genotype 211111. The LSPA genotypes of O157 strains from other clades were determined (Table 3.8 and Fig. 3.2). Two major LSPA genotypes were observed, with *E. coli* O157:H7 strains from Clade 1, 2, 3 and 4 possessing LSPA genotype 111111, and with strains TW09109 (Clade 6), TW11102 (Clade 6), and TW10245 (Clade 7) typed as LSPA genotype 211111 (Table 3.8 and Fig. 3.2). Strains G5101 (Clade 9), TW07763 (Clade 9) and TW01663 (Clade 7), each possessed a unique LSPA genotype and were classified as LSPA genotype 311111, 212111 and 222222, respectively (Table 3.8 and Fig. 3.2). Therefore, LSPA genotype 211111 is not confined to Clade 8, but may also be characteristic of Clade 6 and possibly Clade 7 strains.

TABLE 3.1. Length and GC content of *rhs* genes in the genome of *E. coli* O157:H7 strain Sakai

Gene	Gene designation[1]	Length (bp)[2]	GC (%)[2]
rhsA	ECs4470	4230	59.0
rhsC	ECs0729	4200	59.5
rhsD	ECs0560	4197	61.1
rhsE	ECs2061	4203	61.0
rhsF	ECs4864	4185	58.5
rhsG	ECs0237	4215	61.2
rhsI	ECs0242	1761	56.6
rhsJ	ECs0603	1335	53.3
rhsK	ECs0605	4851	58.0

[1] Gene designations for the 9 *rhs* genes were defined by Hayashi et al. (15).
[2] Length and GC content of *rhs* genes were calculated by Artemis 9.0 (37).

TABLE 3.2. *E. coli* O157:H7 strains analyzed in the present study

Strain	Classification[1]	Source[2]	Reference
FRIK1985	Lineage II	Dr. Andrew Benson	21
FRIK1990	Lineage II	Dr. Andrew Benson	21
FRIK2000	Lineage II	Dr. Andrew Benson	21
FRIK2001	Lineage II	Dr. Andrew Benson	21
FRIK920	Lineage II	Dr. Andrew Benson	21
FRIK944	Lineage II	Dr. Andrew Benson	21
NE037	Lineage II	Dr. Andrew Benson	21
FDA508	Lineage II	Dr. Andrew Benson	21
FDA517	Lineage II	Dr. Andrew Benson	21
93-001	Lineage I	Dr. Andrew Benson	21
95-003	Lineage I	Dr. Andrew Benson	21
FDA507	Lineage I	Dr. Andrew Benson	21
FDA518	Lineage I	Dr. Andrew Benson	21
NE018	Lineage I	Dr. Andrew Benson	21
FRIK523	Lineage I	Dr. Andrew Benson	21
FRIK1275	Lineage I	Dr. Andrew Benson	21
FRIK1986	Lineage I	Dr. Andrew Benson	21
FRIK1997	Lineage I	Dr. Andrew Benson	21
Sakai	Clade 1	Dr. Wei Zhang	15
TW10022	Clade 1	STEC Center	28
93-111	Clade 2	STEC Center	28
TW11308	Clade 2	STEC Center	28
EDL933	Clade 3	STEC Center	28
TW11346	Clade 3	STEC Center	28
TW11039	Clade 4	STEC Center	28
TW11052	Clade 4	STEC Center	28
TW09109	Clade 6	STEC Center	28
TW11102	Clade 6	STEC Center	28
TW10245	Clade 7	STEC Center	28
TW01663	Clade 7	STEC Center	28
M1-Spinach	Clade 8	Dr. Wei Zhang	10
G5101	Clade 9	STEC Center	28
TW07763	Clade 9	STEC Center	28

[1] Lineage designations were previously determined by octamer-based genome scanning (21); Clade designations were determined by Manning et al. (28) using an SNP-based approach.

[2] STEC Center: Shiga Toxin *Escherichia coli* Center at Michigan State University.

TABLE 3.3. Forward and reverse primers and cycling conditions for the 8 *rhs* genes analyzed

Gene		Primer sequence (5'-3')	Cycling conditions[1]
rhsA	Forward	AGCAGACAAATAACCCGATA	95°C for 4 min, 30 cycles of 94°C for 30 s, 50°C for 30 s, and 72°C for 7 min
	Reverse	GCATGCTTTCATACATCTACA	
rhsC	Forward	AGCAGACAAATAACCCGATA	95°C for 4 min, 30 cycles of 94°C for 30 s, 57.2°C for 30 s, and 72°C for 7 min
	Reverse	TTTCGCAATCAGCTATAACA	
rhsD	Forward	TGCTAACACAGGGATATGAA	95°C for 4 min, 35 cycles of 94°C for 30 s, 52.5°C for 30 s, and 72°C for 7 min
	Reverse	TTACGCATAAAACAACAATGA	
rhsE	Forward	CCTGATGAAAAGCAGGATAC	95°C for 5 min, 35 cycles of 95°C for 45 s, 55°C for 1 min, and 72°C for 5 min
	Reverse	CTAGCTGGCTCTATGTCCAC	
rhsF	Forward	GCTAGGCTTCTGGTTCTTATT	95°C for 4 min, 30 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 7 min
	Reverse	ACAACATTGTATTGCCTTCA	
rhsI	Forward	GAACAAAAGAAAGAGCCAAA	95°C for 4 min, 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 3 min
	Reverse	CCCAAGGTAAACAACAAATAA	
rhsJ	Forward	TGGAGATGGAAAAATAAAA	95°C for 4 min, 35 cycles of 94°C for 30 s, 53.3°C for 15 s, and 72°C for 2.5 min
	Reverse	GTAAAGAAACCGAACACC	
rhsK	Forward	GCTGACATAAACGAGGGATA	95°C for 4 min, 30 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 7 min
	Reverse	ATATTGCCGCCTTATTGTT	

[1] All reactions had a final extension at 72°C for 10 min.

TABLE 3.4. Sequencing primers for the 8 *rhs* genes analyzed

Gene	Primer sequence (5'-3')[1]	Priming sites[2]
rhsA	CTGCGTGGTCAGTACCAG	756 - 739
	GAAGCGGATGAGGTGCTG	580 - 597
	GGTGGTGAAAAAGGAACAC	1227 - 1245
	AAACGGGACACAGTACGAT	1836 - 1854
	GAACTCACCACCGCTTATAC	2494 - 2513
	CCACAATACAGAACGACAGA	3104 - 3123
	TAAATGTAGATCCGCAAGGT	3719 - 3738
rhsC	CTGCGTGGTCAGTACCAG	756 - 739
	GAAGCGGATGAGGTGCTG	580 - 597
	GGTGGTGAAAAAGGAACAC	1227 - 1245
	AAACGGGACACAGTACGAT	1836 - 1854
	GAACTCACCACCGCTTATAC	2494 - 2513
	CCACAATACAGAACGACAGA	3104 - 3123
	TTAAATAATGGCGGATATGG	3760 - 3779
rhsD	AAGCGGTGTGCTGTTTAC	796 - 779
	CTGACGTACAGGCGTGAG	658 - 675
	GTATGACGCGGCAGGAAG	1293 - 1310
	AGTATGGAATACGACCTTGC	1921 - 1940
	CTGAACAGCCAGGTGTATG	2569 - 2587
	AAACAGACAACGGTGAGC	3206 - 3223
	TAGGTGGGACTGATCAATTT	3872 - 3891
rhsE	TCAGCACCAGACGGAACT	748 - 731
	TGGATACTGGGGTGGTCA	550 - 567
	CTACCGCTACGACGATACC	1095 - 1113
	GTTATGAATACGACCGCTTT	1673 - 1692
	GTGCATTACGGTTATGATGA	2239 - 2258
	TATGTCTACCGACACGATGA	2830 - 2849
	GAAGAAATCCGGGCAGAC	3337 - 3354
	TTGATATTACTGGCAGTGGA	3830 - 3849
rhsF	CTGCGTGGTCAGTACCAG	756 - 739
	GTACTGACCGGGCTGGTG	622 - 639
	AGTATGAGAAAGACCGCATC	1142 - 1161
	AGGAAGGGCTGAGTCAGTA	1697 - 1715
	TACGGGTATGATGAGAAAGG	2224 - 2243
	ACAGCACCCTCAGCATGT	2753 - 2770
	AAGTGAAATCCTGGCTGAC	3309 - 3327
	TATTGGAATATTTGCATTGG	3771 - 3790
rhsI	AGGTCACCTCCGGTTTAC	646 - 629
	GAGCCGCTACCTCTATGAC	537 - 555

	CAACCAGCTTAATGAGGAGA	1092 - 1111
rhsJ	TGGAGATGGAAAAATAAAA	*
	GAAAATCCGCTTCACGTA	354 - 371
	GTAAAGAAACCGAACACC	**
rhsK	AGGACTGCTGGGAAAACT	1017 - 1034
	CCAGAAACAGAGCGGTTC	3727 - 3744

[1] Primers used in partial sequencing of lineage strains are underlined. The remaining primers were used for sequencing of the full length of *rhs* genes.
[2] Numbers represent the priming sites of each primer in the corresponding *rhs* gene of *E. coli* O157:H7 strain Sakai. The priming site is given from the 5' to the 3' of each sequencing primer.
[*] Forward primer is 5' of the coding sequence of *rhsJ*.
[**] Reverse primer is 3' of the coding sequence of *rhsJ*.

TABLE 3.5. Sequence variations (SNPs, insertions and deletions) identified by *in silico* comparison within 7 *rhs* genes among 14 *E. coli* O157:H7 strains.

SNP loci in *rhsA, C, D, E, F, I and K*

Strains	*rhsA*[1]									*rhsC*						*rhsD*					
	0018	0134	1720	2812	3468	3478	3479	3516	3860	0060	0153	0203	0219	1693	2812	1378	1896	2160	2206	2238	2300
EC4206	G	C	G	T	C	A	A	G	C	T	C	G	C	T	C	G	-	A	A	G	C
EC4024	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	.	.	.	.
EC4115	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	.	.	.	.
EC4113	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
EC4045	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
EC4042	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
EC4196	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	-	.	.	.	.
EC4076	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	-	.	.	.	.
EC4401	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	.	.	.	.
EC4486	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	-	-	-	-	G
EC4501	.	T	.	C	G	G	C	A	.	#	#	#	#	#	#	A	-	.	.	.	.
EC508	.	.	.	.	.	.	.	.	.	#	#	#	#	#	#	A	*	.	.	.	.
EC869[3]	A	.	T	.	G	G	C	A	A	C	G	C	T	C	T	A	-	.	.	.	.
Sakai[2]	.	T	.	C	G	G	C	A	.	.	.	.	.	C	.	A	-	.	.	.	.

Strains	*rhsE*															
	0492	0499	0653	0670	0702	0726	0759	0777	0794	0810	0811	0815	0816	0935	0946	0958
EC4206	-	-	G	G	A	G	G	G	C	C	C	C	C	A	G	C
EC4024	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4115	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4113	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4045	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4042	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4196	-	-	-	-	-	C	-	-	-	-	-	-	.	.	.	.
EC4076	C	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4401	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4486	-	-	.	.	.	.	.	.	.	.	.	.	.	.	.	.
EC4501	-	-	.	.	.	.	.	.	.	.	.	.	A	.	.	.
EC508	-	-	.	.	.	.	.	.	.	.	.	.	A	.	.	.
EC869[3]	-	-	.	.	.	.	.	.	.	.	.	.	A	G	C	T
Sakai[2]	-	-	.	.	.	.	.	.	.	.	.	.	A	.	.	.

Strains	*rhsF*				*rhsI*					*rhsK*					Outbreak[5]	ST[4]
	0134	1720	2812	3672	0120	0133	0210	0248	1185	1129	1141	1158	1248	4172		
EC4206	C	G	T	G	C	T	C	C	C	A	A	A	–	C	Spinach	I1
EC4024	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I2
EC4115	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I2
EC4113	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I3
EC4045	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I3
EC4042	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I3
EC4196	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I4
EC4076	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Spinach	I5
EC4401	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Taco-Bell	I2
EC4486	.	.	.	.	.	.	.	.	.	-	-	-	C	.	Taco-Bell	I6
EC4501	T	.	C	.	.	.	.	.	T	-	-	-	C	A	Taco-John	NA
EC508	.	.	.	A	.	.	.	.	.	-	-	-	C	.		NA
EC869[3]	.	T	.	.	T	C	T	T	.	-	-	-	C	A		I7
Sakai[2]	T	.	C	.	.	.	.	.	T	-	-	-	C	A	Sakai	I8

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots. Gaps at each SNP locus are indicated by dashes.
[2] Sequences of *rhsD* and *rhsE* for strain Sakai were obtained by sequencing in this study. Sequences of the other *rhs* genes for strain Sakai were extracted from GenBank (accession number BA000007) using Artemis (Version 9) (37).
[3] The first 101 nucleotides of *rhsI* from strain EC869 were missing.
[4] ST: Sequence Type. "I" stands for *in silico* comparison and numbers are assigned to different sequence types.
[5] Spinach stands for the 2006 spinach *E. coli* O157:H7 outbreak (10). Taco-Bell stands for the 2006 Taco-Bell *E. coli* O157:H7 outbreak (9). Taco-John stands for the 2006 Taco-John *E. coli* O157:H7 outbreak (12). Sakai stands for the *E. coli* O157:H7 strain Sakai (15).
Complete sequences of *rhsC* of EC4501 and EC508 were not available.
* There is an insertion from nucleotide position 1896 to 1949 in *rhsD* of EC508.

TABLE 3.6. SNP loci identified by sequencing within 8 *rhs* genes among lineage strains

	SNP loci in *rhsA, C, D, E, F, I and J*								
	rhsA[1]	*rhsC*	*rhsD*	*rhsE*	*rhsF*	*rhsI*	*rhsJ*	*rhsK*	
	0 0 1 2 3 3 3 3 3	0 0 0 0 0 1 2	1 1 1 1 1 1 2 2	0 0 0 1	0 0 1 2 3	0 1	0 0 0 0 0 0 0 0	4		
	0 1 7 8 4 4 4 5 8	0 1 1 2 2 4 8	6 8 8 9 9 9 6 6	9 9 9 3	0 1 7 8 6	3 1	6 6 6 7 7 7 7 8	1		
	1 3 2 1 6 7 7 1 6	6 5 6 0 1 3 1	8 4 8 2 3 3 5 6	3 4 5 1	1 3 2 1 7	8 8	1 8 9 1 4 6 6 0	6		
Strains	8 4 0 2 8 8 9 6 0	0 3 9 3 9 5 2	7 2 7 9 2 8 0 7	3 4 6 4	8 4 0 2 2	7 5	2 8 6 7 5 5 8 7	9	Lineage[2]	ST[3]
FRIK1985	A C T T G G C A A	C G G C T G C	C C G A C T C C	G C T C	G C T T G	# #	A G C A T A T A	A	II	L1
FRIK2000	. . . .	. . . .	. . . .	. . . .	. . . . .	# #	. . . .	.	II	L1
FRIK920	. . . .	. . . .	. . . .	. . . .	. . . . .	# #	. . . .	.	II	L1
FRIK944	. . . .	. . . .	. . . .	. . . .	. . . . .	# #	. . . .	.	II	L1
FRIK1990	. . . .	. . . .	T	A G C .	A	# #	. . . .	.	II	L2
FRIK2001*	G	. . A	. . . .	. . . A	. . . . .	# #	. . . .	.	II	L3
NE037	G . G C	T C . G C A .	. T A G T C T T	A G C .	. . G . .	T C	G C G G C G C G	.	II	L4
FDA508**	G . G . C A A G C	T C . G C . T	. . . .	A G C .	. . G . A	C C	. . . .	C	II	L5
FDA517	G . G C	T C . G C . .	. . . .	A G C .	. . G . .	T C	G C G G C G C G	.	II	L6
93-001	G T G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L7
95-003	G T G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L7
NE018	G T G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L7
FRIK1986	G T G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L7
FDA507	G . G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L8
FDA518	G . G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L8
FRIK1275	G . G C C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L8
FRIK523	G . G C	T C . G C . .	. . . .	A G C .	. T G C .	C C	. . . .	.	I	L9
FRIK1997	G . G C C A A . C	T C . G C . .	. . . .	A G C .	. T G C .	C T	. . . .	.	I	L10

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.

[2] Lineage designations were defined by Kim et al. (21).

[3] ST: Sequence Type. "L" stands for lineage and numbers are assigned to different sequence types.

Amplification of *rhsI* from FRIK1985, FRIK1990, FRIK2000, FRIK2001, FRIK920 and FRIK944 was not successful.

*There is a deletion from nucleotide position 3377 to 3463 in *rhsA* of strain FRIK2001.

**There is a 54-nucleotide insertion at the nucleotide position 1958 in *rhsD* of strain FDA508.

TABLE 3.7. Summary of allelic types of the 7 *rhs* genes for 15 clade-strains*

Clade[3]	Strains	Allelic types of *rhs* genes[1]							ST of strains[2]
		rhsA	*rhsC*	*rhsD*	*rhsE*	*rhsF*	*rhsI*	*rhsJ*	
1	Sakai	1	1	1	1	1	1	1	C1
1	TW10022	1	1	1	1	1	1	1	C1
2	93-111	1	1	1	1	1	1	1	C1
2	TW11308	1	1	2	1	1	1	1	C2
3	EDL933	2	1	1	1	1	1	1	C3
3	TW11346	2	2	3	1	1	1	1	C4
4	TW11039	3	1	1	1	2	2	1	C5
4	TW11052	3	1	1	1	2	2	1	C5
6	TW09109	3	1	4	1	3	3	2	C6
6	TW11102	3	1	1	1	3	3	2	C7
7	TW10245	4	3	1	2	4	2	1	C8
7	TW01663	5	4	1	3	5	NA	1	C9
8	M1-Spinach	6	5	5	4	3	2	1	C10
9	G5101	7	6	6	5	6	2	1	C11
9	TW07763	8	7	7	6	7	2	1	C12

[1] Numbers are assigned within each *rhs* gene to represent different allelic types.
[2] ST: Sequence Type. "C" stands for clade and numbers are assigned to different sequence types.
[3] Clades designation were determined by Manning et al. (28).
NA: Not Available. Amplification of *rhsI* from TW01663 was not successful.
* Please refer to Appendix A Supplementary Table 1 to 7 for full data.

Table 3.8. Lineage specific polymorphism assay (LSPA) genotypes of *E. coli* O157:H7 strains in the present study

Strain	Classification[1]	LSPA genotype[2]
Sakai[*]	Clade 1	111111
TW10022	Clade 1	111111
93-111	Clade 2	111111
TW11308	Clade 2	111111
EDL933[*]	Clade 3	111111
TW11346	Clade 3	111111
TW11039	Clade 4	111111
TW11052	Clade 4	111111
TW09109	Clade 6	211111
TW11102	Clade 6	211111
TW10245	Clade 7	211111
TW01663	Clade 7	222222
M1-Spinach	Clade 8	211111
G5101	Clade 9	311111
TW07763	Clade 9	212111
93-001[†]	Lineage I	111111
95-003[†]	Lineage I	111111
FDA507[†]	Lineage I	111111
FDA518[†]	Lineage I	111111
NE018	Lineage I	111111
FRIK523	Lineage I	111111
FRIK1275	Lineage I	111111
FRIK1986	Lineage I	111111
FRIK1997	Lineage I	111111
FRIK944	Lineage II	222222
FRIK920[*†]	Lineage II	222222
FRIK1990[†]	Lineage II	222222
FRIK2000	Lineage II	222222
FRIK2001[†]	Lineage II	222211
FRIK1985[†]	Lineage II	222221
NE037[†]	Lineage II	211131
FDA508[†]	Lineage II	211111
FDA517	Lineage II	211111

[1] Clade designations were defined by Manning et al. (28). Lineage designations were defined by Kim et al. (21).

[2] LSPA: Lineage Specific Polymorphism Assay. Six gene markers in LSPA are listed in the order: *folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*. Numbers are assigned for each gene marker to represent different alleles.

[*] Sakai, EDL933 and FRIK920 whose LSPA genotypes were determined by Zhang et al. (47), were used as standards for LSPA genotype 111111 and LSPA genotype 222222, respectively.

[†] The LSPA genotypes of these strains were determined previously by Yang et al. (44), and are confirmed in the present study.

FIGURE 3.1 Unrooted neighbor-joining tree of 18 lineage-strains based upon the number of differences in parsimony informative sites among *rhsA, C, D, E, F, J* and *K.* SNPs identified in each *rhs* gene (Table 3.6) were concatenated to represent the SNP sequence type for each *E. coli* O157:H7 strain. Parsimony informative sites were used to construct the phylogenetic tree. Since sequence information of *rhsI* was not available for all strains, *rhsI* was not included. Bootstrap values (1,000 replications) above 75 are shown at the interior branches. Substitutions per site are indicated by the bar at the bottom. The strains marked by open circle are Lineage II strains and strains marked with solid circle are Lineage I strains, whose designations were defined previously by Kim et al. (21). LSPA genotypes are indicated to the right of the strain designations.

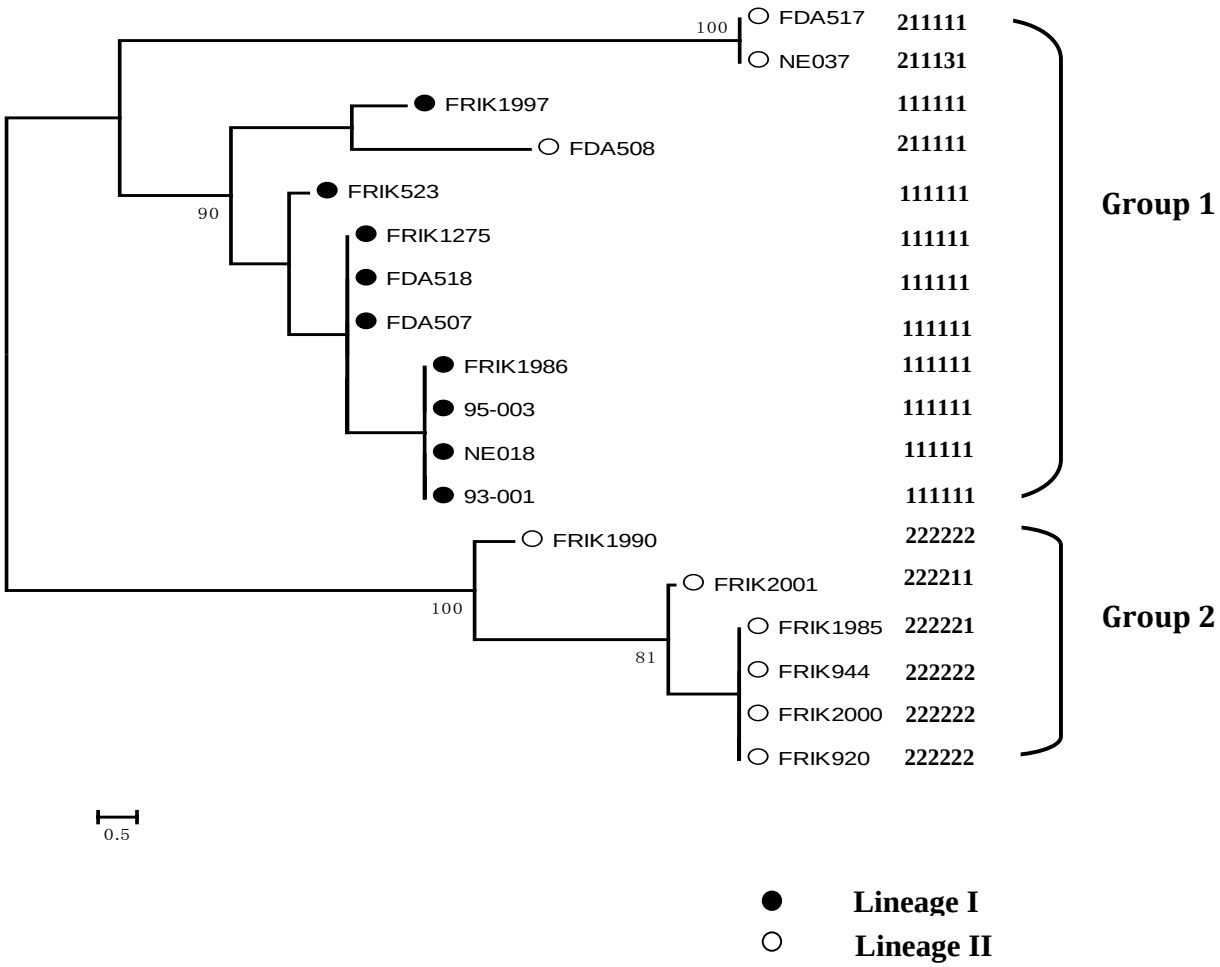


FIGURE 3.1 Unrooted neighbor-joining tree of 18 lineage-strains based upon the number of differences in parsimony informative sites among *rhsA, C, D, E, F, J* and *K*.

FIGURE 3.2 Lineage Specific Polymorphism Assay (LSPA) of 14 clade-strains. Fourteen *E. coli* O157:H7 strains from 8 clades were subjected to LSPA. The positions of the 6 gene markers (*folD-sfmA*, *Z5395*, *yhcG*, *rtcB*, *rbsB*, and *iclR-arp*) are indicated on the left. Molecular size marker (50 bp) is loaded in the middle and sizes are indicated on the right. Strains from the 8 clades are indicated at the bottom. FRIK920 and EDL933 that were reported as LSPA genotype 222222 and LSPA genotype 111111 (47) were used as standards.

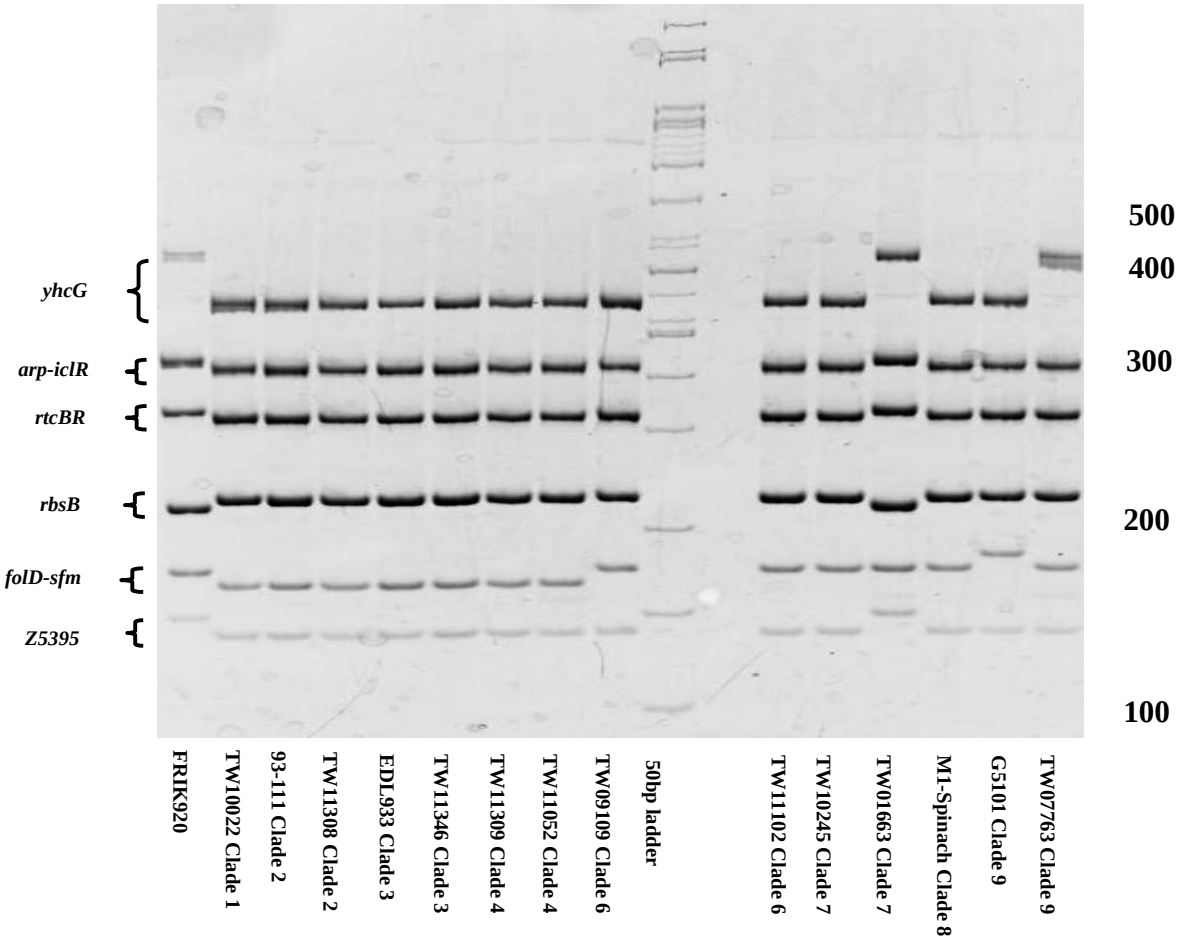


FIGURE 3.2 Lineage Specific Polymorphism Assay (LSPA) of 14 clade-strains.

3.5 Discussion

Foodborne outbreaks pose a great financial and public-health burden (38). To control ongoing outbreaks and prevent future outbreaks, it is critical to identify sources and routes of transmission of foodborne pathogens. Molecular subtyping methods are powerful tools for microbial source tracking. An ideal molecular subtyping method should be cost effective, high throughput, and most importantly it should possess high discriminatory power and epidemiologic concordance (17). Several subtyping methods exist for *E. coli* O157:H7 (1, 19, 26), including PFGE (7), which is generally regarded as the "gold standard" method for subtyping *E. coli* O157:H7. However the drawbacks inherent in PFGE, such as low portability and ambiguity of results, render a DNA-sequence-based approach more appropriate. MLST, a DNA-sequence-based subtyping approach (27), generates unambiguous DNA sequence data which are easily portable between laboratories. Additionally MLST is a much simpler protocol than PFGE. Furthermore, MLST is more suitable for phylogenetic analysis. The ability of a subtyping method to reveal the phylogeny of pathogens is imperative as several studies have shown different phylogenetic groups of *E. coli* O157:H7 are not equally pathogenic to humans (21, 28).

A useful MLST scheme in epidemiological investigations of outbreaks demands that gene markers targeted possess sufficient DNA sequence variations to discriminate between unrelated bacterial strains. Various gene markers were studied previously (13, 32), but these markers showed little sequence variation. For example, Noller et al. (32) targeted 7 housekeeping genes (*arcA*, *aroE*, *dnaE*, *mdh*, *gnd*, *gapA*, and *pgm*) and 2 membrane protein genes (*ompA* and *espA*) in their MLST scheme to type a total of 77 *E.*

coli O157:H7 isolates. The only sequence variations identified were 2 SNPs in *ompA*, found in 5 *E. coli* O157:H7 isolates. In another study, Foley et al. (13) chose 1 housekeeping gene (*uidA*) and 3 virulence genes as their targets (*eaeA*, *hlyA*, and *fliC*) and they examined 92 *E. coli* O157:H7 isolates from various sources. They found only 5 MLST types among the 92 *E. coli* O157:H7 isolates, compared with 72 distinct profiles by PFGE. The remarkable lack of sequence variation observed by these researchers could be explained by two theories that are not mutually exclusive. First, the beta-glucuronidase negative, sorbitol-negative *E. coli* O157:H7 strains are believed to have emerged from a common ancestor approximately 40,000 years ago (46), which may not be sufficient for SNPs in housekeeping and virulence genes to accumulate to levels necessary for MLST to be effective. Secondly, gene markers used in these previous studies may not be under strong enough positive selective pressure and therefore allelic variants are not selected for at an appreciable rate for these gene markers to be useful in revealing the short-term evolution history of *E. coli* O157:H7 during epidemiological investigations.

Therefore, in our MLST scheme we examined several *rhs* genes, which were identified as the genes that are under the strongest positive selection in a survey of *E. coli* genomes (33). Our study showed that these *rhs* genes possessed significantly more sequence variations than housekeeping and virulence genes used in previous studies (13, 32) (Table 3.6 and Table 3.7). This suggests that *rhs* genes could serve as potential markers in sequence-based methods for subtyping *E. coli* O157:H7 during epidemiological investigations. Additionally, based upon our results, we also suggest that genes under the strongest positive selection may be good gene markers for revealing

the short-term evolutionary history of *E. coli* O157:H7 and other highly clonal bacterial pathogens. However, it should also be kept in mind that *rhs* genes alone may not be able to separate all unrelated *E. coli* O157:H7 outbreak strains. Therefore, other gene markers should be included to increase discriminatory power. We are currently using high-throughput sequencing of *E. coli* O157:H7 genomes to identify such targets.

Significant DNA sequence variations exist between homologous *rhs* genes, but we are unable to explain why *rhs* genes are under such strong positive selection. While the function(s) of *rhs* genes are still unknown, several studies shed light on this puzzle. One study by Hill et al. (16) predicted that *rhs* genes encode extracellular proteins based upon their primary sequences, however we are unfamiliar with any experimental evidence supporting this claim. Another study suggested *rhsA* gene may promote the intestinal colonization of *E. coli* in calves (41) and the biogenesis of group 2 capsules (29). Allen et al. (2) showed a possible role of *rhsB* and *rhsD* gene in increasing the resistance of *E. coli* to the biocide polyhexamethylene biguanide. Taking all these together, we speculate that Rhs proteins promote survival of *E. coli* during intestinal transit in some undefined manner, and that their extracellular nature results in a strong positive selective pressure from the host's immune system.

The capability of a subtyping method to discriminate unrelated *E. coli* O157:H7 strains is imperative, but it is also important for a subtyping method to determine phylogeny, as previous studies suggested that not all *E. coli* O157:H7 strains are equally pathogenic towards humans (21, 28). A previous study showed that PFGE failed to reveal the phylogenetic relatedness of *E. coli* O157:H7 strains (23). In the present study, phylogenetic analysis of 18 lineage *E. coli* O157:H7 strains using *rhs* genes generally

agrees with the phylogeny determined previously (21), although the 3 human-origin *E. coli* O157:H7 strains from Lineage II (NE037, FDA508 and FDA517) clustered together with the 9 Lineage I strains (Fig. 3.1). However, using the 6-gene LSPA assay we determined that two of these strains are classified in the recently described Lineage I/II, and that NE037 has an LSPA genotype (211131) distinct from Lineage I, I/II and II. Our MLST scheme correctly clustered together strains with LSPA genotype 111111, which is the most prevalent LSPA genotype in Lineage I , and grouped together LSPA genotypes 222222, 222221 and 222211 which belong to Lineage II (47). Our data suggest that NE037 (LSPA genotype 211131) falls into Lineage I/II, however further studies are needed to confirm this.

Though our MLST scheme, to a large extent, confirmed the phylogeny of the 18 lineage strains defined by a previous study (21), it did not accurately classify strains according to clades described by Manning et al. (28). We initially hypothesized that strains from the same clade would share the same sequence type and therefore this MLST scheme could serve as a quick method to distinguish the 9 clades. However, our results showed that strains from the same clade possessed different sequence types except for Clade 1 and Clade 4 (Table 3.7). This could be explained by two possibilities. First, differences in the sequences of *rhs* genes may not accurately reflect the evolutionary history of *E. coli* O157:H7. Second, differences in the sequences of *rhs* genes may define additional phylogenetic groups within a clade. Therefore, to distinguish these two possibilities, we suggest a larger number of strains from each clade be studied.

Despite of the fact that O157 strains were not classified according to their clade designations using *rhs* genes, several interesting observations are worth mentioning.

First of all, the 2 Clade 1 strains, Sakai and TW10022 shared the same sequence type, but were not distinguishable from 93-111, a Clade 2 strain (Table 3.7). Recently, Riordan et al. (36) targeted SNPs in 4 genes and developed a quick subtyping method to discriminate clades strains, and they also reported problems separating Clade 1 from Clade 2. Taken together, these observations suggest Clade 1 and 2 are highly related and better methods are needed to differentiate these two clades. Second of all, within Clade 7, TW01663 is characterized as LSPA 222222 (Lineage II), whereas TW10245 is defined as LSPA 211111 (Lineage I/II) (Table 3.8) (Fig. 3.2). Since two different lineages would not be expected to coexist within the same clade, further study is needed to investigate whether there are indeed two lineages in Clade 7. Furthermore, G5101 was defined as LSPA genotype 311111, which has not been previously reported (Table 3.8) (Fig. 2). Further studies are needed to determine the prevalence of this LSPA genotype in *E. coli* O157:H7.

In addition to the above observations, the present study also provided data suggesting that *E. coli* O157:H7 strains from Clades 6, 7 and 8 are possible members of Lineage I/II. The original classification of the LSPA genotype 211111 grouped these strains into Lineage II (44), a group more commonly associated with strains that are of low virulence to humans. Through full genome phylogenetic analysis, Zhang et al. (47) defined an additional group, Lineage I/II, and suggested that this lineage comprises LSPA genotype 211111. Their data also suggest that Lineage I/II (LSPA genotype 211111) strains are more closely related to Lineage I strains (LSPA genotype 111111) than they are to Lineage II strains (LSPA genotype 222222, 222211, and 222212) (47). Our data support this, suggesting that like Lineage I strains, Lineage I/II strains are also responsible for

causing severe outbreaks in human (Table 3.8). Our data also provide further insight into the observation of Manning et al. (28), that *E. coli* O157:H7 strains EDL933 and Sakai, which are both LSPA genotype 111111, share ca. 97% of their gene content, but only ca. 89% gene content with a strain from the 2006 Spinach outbreak, which we defined in this study as LSPA genotype 211111. Comparative genomic analysis of Lineage I, I/II and II strains may lend insight into differences in pathogenicity and host specificity of these strains.

Lastly, this study identified sequence markers that may be useful for differentiating strains of different lineages and clades. First, as noted previously by Zhang et al. (47), we showed that the complete gene for *rhsI* is not present in Lineage II strains, providing further evidence that this marker might be useful for separating these strains from Lineage I and I/II. Furthermore, the present study also identified possible SNP markers for identifying Clade 6 and Clade 8. Four SNPs identified in *rhsA* at nucleotide positions 3468, 3478, 3479 and 3516 could potentially be used as markers for Clade 8 (Appendix A Supplementary Table 1). These 4 SNPs were observed in the spinach and Taco Bell outbreak strains (Table 3.5) as well as in strain FDA508 (Table 3.6), which were all confirmed to be Clade 8 by either sequencing or *in silico* comparison of ECs2357 (36) (data not shown). Eight unique SNPs in *rhsJ* were identified for the 2 Clade 6 strains TW09109 and TW11102 (Appendix A Supplementary Table 7). Interestingly, NE037 and FDA517 also possess these 8 SNPs in *rhsJ* (Table 3.6). However due to the lack of information on SNPs defining Clade 6, we are not able to determine at this time whether NE037 and FDA517 also belong to Clade 6.

In conclusion, our study identified significant DNA sequence variations in 7 *rhs*

genes (*rhsA, C, D, E, F, I* and *J*) that could potentially serve as markers for subtyping *E. coli* O157:H7. Future studies are required to test this MLST scheme on a larger collection of *E. coli* O157:H7 strains from well-defined outbreaks in order to examine whether this method might effectively subtype *E. coli* O157:H7 strains during epidemiological investigations. As we currently do not have the full lengths of *rhsC*, *D*, *E*, *F*, *I*, *J* and *K* for all the lineage strains (Table 3.6), it is possible that additional SNPs may be found by sequencing the full lengths of these genes. Additionally, the present study also suggested that *E. coli* O157:H7 strains from Clade 8 fall into Lineage I/II and that unique markers in *rhsA* and *J* could be used for distinguishing Clade 8 and Clade 6. A larger collection of *E. coli* O157:H7 strains should be tested to determine whether these SNPs are unique to Clade 8 and Clade 6.

3.6 Acknowledgements

This work was supported by a United States Department of Agriculture Cooperative State Research, Education, and Extension Service (USDA-CSREES) grant 2008-34163-19283 (E.G.D and S.J.K) and by start-up Funds from the Department of Food Science and the College of Agricultural Sciences at the Pennsylvania State University (E.G.D).

We would like to thank Dr. Andrew Benson at the University of Nebraska-Lincoln, Dr. Thomas Whittam and Beth Whittam at the Michigan State University, and Dr. Wei Zhang at the Illinois Institute of Technology for kindly providing *E. coli* O157:H7 strains. We also want to thank Dr. Bindhu Verghese for helping with the phylogenetic analysis of *rhs* sequences. We would also like to thank Dr. Robert Roberts at the Pennsylvania State University, Jia Wen, Amrita Puri and Chun Chen for reading this manuscript before submission.

3.7 References

1. **Ahmed, R., C. Bopp, A. Borczyk, and S. Kasatiya.** 1987. Phage-typing scheme for *Escherichia coli* O157:H7. J. Infect. Dis. **155:**806-809.
2. **Allen, M. J., G. F. White, and A. P. Morby.** 2006. The response of *Escherichia coli* to exposure to the biocide polyhexamethylene biguanide. Microbiology **152:**989-1000.
3. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.
4. **Armstrong, G. L., J. Hollingsworth, and J. G. Morris, Jr.** 1996. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. Epidemiol. Rev. **18:**29-51.
5. **Baker, D. R., R. A. Moxley, M. B. Steele, J. T. Lejeune, J. Christopher-Hennings, D. G. Chen, P. R. Hardwidge, and D. H. Francis.** 2007. Differences in virulence among *Escherichia coli* O157:H7 strains isolated from humans during disease outbreaks and from healthy cattle. Appl. Environ. Microbiol. **73:**7338-7346.
6. **Beutin, L., D. Geier, H. Steinruck, S. Zimmermann, and F. Scheutz.** 1993. Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. J. Clin. Microbiol. **31:**2483-2488.

7. **Byrne, C. M., I. Erol, J. E. Call, C. W. Kaspar, D. R. Buege, C. J. Hiemke, P. J. Fedorka-Cray, A. K. Benson, F. M. Wallace, and J. B. Luchansky.** 2003. Characterization of *Escherichia coli* O157:H7 from downer and healthy dairy cattle in the upper Midwest region of the United States. Appl. Environ. Microbiol. **69:**4683-4688.

8. **Carver, T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21:**3422-3423.

9. **CDC.** 2006. Multistate outbreak of *E. coli* O157 infections, November-December 2006. http://www.cdc.gov/ecoli/2006/december/121406.htm.

10. **CDC.** 2006. Ongoing multistate outbreak of *Escherichia coli* serotype O157:H7 infections associated with consumption of fresh spinach—United States. Morbid. Mortal. Wkly. Rep. **55:**1045-1046.

11. **Elder, R. O., J. E. Keen, G. R. Siragusa, G. A. Barkocy-Gallagher, M. Koohmaraie, and W. W. Laegreid.** 2000. Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. Proc. Natl. Acad. Sci. U S A **97:**2999-3003.

12. **FDA.** 2007. FDA and states closer to identifying source of *E. coli* contamination associated with illnesses at Taco John's restaurants. http://www.fda.gov/bbs/topics/NEWS/2007/NEW01546.html.

13. **Foley, S. L., S. Simjee, J. Meng, D. G. White, P. F. McDermott, and S. Zhao.** 2004. Evaluation of molecular typing methods for *Escherichia coli* O157:H7 isolates from cattle, food, and humans. J. Food Prot. **67:**651-657.

14. **Hancock, D. D., T. E. Besser, M. L. Kinsel, P. I. Tarr, D. H. Rice, and M. G. Paros.** 1994. The prevalence of *Escherichia coli* O157:H7 in dairy and beef cattle in Washington State. Epidemiol. Infect. **113:**199-207.

15. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11-22.

16. **Hill, C. W., C. H. Sandt, and D. A. Vlazny.** 1994. *Rhs* elements of *Escherichia coli*: a family of genetic composites each encoding a large mosaic protein. Mol. Microbiol. **12:**865-871.

17. **Hyytia-Trees, E. K., K. Cooper, E. M. Ribot, and P. Gerner-Smidt.** 2007. Recent developments and future prospects in subtyping of foodborne bacterial pathogens. Future Microbiol. **2:**175-185.

18. **Iguchi, A., R. Osawa, J. Kawano, A. Shimizu, J. Terajima, and H. Watanabe.** 2002. Effects of repeated subculturing and prolonged storage at room temperature of enterohemorrhagic *Escherichia coli* O157:H7 on pulsed-field gel electrophoresis profiles. J. Clin. Microbiol. **40:**3079-3081.

19. **Iyoda, S., A. Wada, J. Weller, S. J. Flood, E. Schreiber, B. Tucker, and H. Watanabe.** 1999. Evaluation of AFLP, a high-resolution DNA fingerprinting method, as a tool for molecular subtyping of enterohemorrhagic *Escherichia coli* O157:H7 isolates. Microbiol. Immunol. **43:**803-806.

20. **Jay, M. T., M. Cooley, D. Carychao, G. W. Wiscomb, R. A. Sweitzer, L. Crawford-Miksza, J. A. Farrar, D. K. Lau, J. O'Connell, A. Millington, R. V. Asmundson, E. R. Atwill, and R. E. Mandrell.** 2007. *Escherichia coli* O157:H7 in feral swine near spinach fields and cattle, central California coast. Emerg. Infect. Dis. **13:**1908-1911.

21. **Kim, J., J. Nietfeldt, and A. K. Benson.** 1999. Octamer-based genome scanning distinguishes a unique subpopulation of *Escherichia coli* O157:H7 strains in cattle. Proc. Natl. Acad. Sci. U S A **96:**13288-13293.

22. **Kotewicz, M. L., M. K. Mammel, J. E. LeClerc, and T. A. Cebula.** 2008. Optical mapping and 454 sequencing of *Escherichia coli* O157:H7 isolates linked to the US 2006 spinach-associated outbreak. Microbiology **154:**3518-3528.

23. **Laing, C., C. Pegg, D. Yawney, K. Ziebell, M. Steele, R. Johnson, J. E. Thomas, E. N. Taboada, Y. Zhang, and V. P. Gannon.** 2008. Rapid determination of *Escherichia coli* O157:H7 lineage types and molecular subtypes by using comparative genomic fingerprinting. Appl. Environ. Microbiol. **74:**6606-6615.

24. **Lau, Y. J., B. S. Hu, W. L. Wu, Y. H. Lin, H. Y. Chang, and Z. Y. Shi.** 2000. Identification of a major cluster of *Klebsiella pneumoniae* isolates from patients with liver abscess in Taiwan. J. Clin. Microbiol. **38:**412-414.

25. **Lin, R. J., M. Capage, and C. W. Hill.** 1984. A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. **177:**1-18.

26. **Lindstedt, B. A., E. Heir, E. Gjernes, T. Vardund, and G. Kapperud.** 2003. DNA fingerprinting of Shiga-toxin producing *Escherichia coli* O157 based on multiple-locus variable-number tandem-repeats analysis (MLVA). Ann. Clin. Microbiol. Antimicrob. **2:**12.

27. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U S A **95:**3140-3145.

28. **Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U S A **105:**4868-4873.

29. **McNulty, C., J. Thompson, B. Barrett, L. Lord, C. Andersen, and I. S. Roberts.** 2006. The cell surface expression of group 2 capsular polysaccharides in *Escherichia coli*: the role of KpsD, RhsA and a multi-protein complex at the pole of the cell. Mol. Microbiol. **59:**907-922.

30. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg. Infect. Dis. **5:**607-625.

31. **Nei, M., and S. Kumar.** 2000. Chapter 7 Phylogenetic inference: Maximum parsimony methods, p.119-122. *In* M. Nei and S. Kumar (ed.), Molecular Evolution and Phylogenetics, Oxford University Press, New York, NY.

32. **Noller, A. C., M. C. McEllistrem, O. C. Stine, J. G. Morris, Jr., D. J. Boxrud, B. Dixon, and L. H. Harrison.** 2003. Multilocus sequence typing reveals a lack of diversity among *Escherichia coli* O157:H7 isolates that are distinct by pulsed-field gel electrophoresis. J. Clin. Microbiol. **41:**675-679.

33. **Petersen, L., J. P. Bollback, M. Dimmic, M. Hubisz, and R. Nielsen.** 2007. Genes under positive selection in *Escherichia coli*. Genome Res. **17:**1336-1343.

34. **Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow.** 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982-2002. Emerg. Infect. Dis. **11:**603-609.

35. **Riley, L. W., R. S. Remis, S. D. Helgerson, H. B. McGee, J. G. Wells, B. R. Davis, R. J. Hebert, E. S. Olcott, L. M. Johnson, N. T. Hargrett, P. A. Blake, and M. L. Cohen.** 1983. Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. N. Engl. J. Med. **308:**681-685.

36. **Riordan, J. T., S. B. Viswanath, S. D. Manning, and T. S. Whittam.** 2008. Genetic differentiation of *Escherichia coli* O157:H7 clades associated with human disease by real-time PCR. J. Clin. Microbiol. **46:**2070-2073.

37. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944-945.

38. **Stewart, J. R., J. W. Santo Domingo, T. J. Wade.** 2007. Fecal pollution, public health, and microbial source tracking, p.3-9. *In* J. W. Santo Domingo and M. J. Sadowsky (ed.), Microbial Source Tracking. American Society for Microbiology, Washington, DC.

39. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596-1599.

40. **Tilden, J., Jr., W. Young, A. M. McNamara, C. Custer, B. Boesel, M. A. Lambert-Fair, J. Majkowski, D. Vugia, S. B. Werner, J. Hollingsworth, and J. G. Morris, Jr.** 1996. A new route of transmission for *Escherichia coli*: infection from dry fermented salami. Am. J. Public Health **86:**1142-1145.

41. **van Diemen, P. M., F. Dziva, M. P. Stevens, and T. S. Wallis.** 2005. Identification of enterohemorrhagic *Escherichia coli* O26:H- genes required for intestinal colonization in calves. Infect. Immun. **73:**1735-1743.

42. **Wells, J. G., L. D. Shipman, K. D. Greene, E. G. Sowers, J. H. Green, D. N. Cameron, F. P. Downes, M. L. Martin, P. M. Griffin, S. M. Ostroff, M. E. Potter, R. V. Tauxe and I. K. Wachsmuth.** 1991. Isolation of *Escherichia coli* serotype O157:H7 and other Shiga-like-toxin-producing *E. coli* from dairy cattle. J. Clin. Microbiol. **29:**985-989.

43. **Willshaw, G. A., J. Thirlwell, A. P. Jones, S. Parry, R. L. Salmon, and M. Hickey.** 1994. Vero cytotoxin-producing *Escherichia coli* O157 in beef burgers linked to an outbreak of diarrhea, hemorrhagic colitis and hemolytic uremic syndrome in Britain. Lett. Appl. Microbiol. **19:**304-307.

44. **Yang, Z., J. Kovar, J. Kim, J. Nietfeldt, D. R. Smith, R. A. Moxley, M. E. Olson, P. D. Fey, and A. K. Benson.** 2004. Identification of common subpopulations of non-sorbitol-fermenting, beta-glucuronidase-negative *Escherichia coli* O157:H7 from bovine production environments and human clinical samples. Appl. Environ. Microbiol. **70:**6846-6854.

45. **Zhang, W., B. M. Jayarao, and S. J. Knabel.** 2004. Multi-virulence-locus sequence typing of *Listeria monocytogenes.* Appl. Environ. Microbiol. **70:**913-920.

46. **Zhang, W., W. Qi, T. J. Albert, A. S. Motiwala, D. Alland, E. K. Hyytia-Trees, E. M. Ribot, P. I. Fields, T. S. Whittam, and B. Swaminathan.** 2006. Probing genomic diversity and evolution of *Escherichia coli* O157 by single nucleotide polymorphisms. Genome Res. **16:**757-767.

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

48. **Ziebell, K., M. Steele, Y. Zhang, A. Benson, E. N. Taboada, C. Laing, S. McEwen, B. Ciebin, R. Johnson, and V. Gannon.** 2008. Genotypic characterization and prevalence of virulence factors among Canadian *Escherichia coli* O157:H7 strains. Appl. Environ. Microbiol. **74:**4314-4323.

CHAPTER FOUR

CONCLUSIONS AND FUTURE RESEARCH

4.1 Conclusions

This study showed that *rhs* genes possess sufficient DNA sequence variations to incorporate into an MLST scheme for subtyping *E. coli* O157:H7. Sequence analysis of *rhs* genes revealed 44 PL and 10 STs among the 18 lineage-strains, and 280 PL and 12 STs among the 15 clade-strains. Compared with gene markers used in previous publications, *rhs* genes demonstrated much more DNA sequence diversity. Therefore, I suggest *rhs* genes could serve as useful markers in an MLST scheme for subtyping *E. coli* O157:H7 during epidemiological investigations.

My data also showed that *rhs* genes might be useful in inferring the phylogeny of *E. coli* O157:H7. Phylogenetic analysis of the 18 lineage-strains using *rhs* genes generally agreed with a previous study that originally defined Lineage I and II. However, *rhs* genes were not informative in revealing the phylogeny of the 15 clade-strains largely due to the fact that strains included in the present study are too limited. Twelve STs were identified among the 15 clade-strains, out of which 10 strains represented 10 unique STs. Most of the SNPs that could be used to differentiate strains of different STs were represented in only one strain (singleton), and are not useful for inferring the phylogeny of the 15 clade-strains as only parsimony informative sites (SNPs that are represented in at least two strains) are useful for the construction of a neighbor-joining tree. Therefore, a larger collection of *E. coli* O157:H7 strains from all clades should be included to examine the usefulness of *rhs* genes in revealing the phylogeny of clade-strains.

Additionally, my data suggested for the first time that Clade 8 strains, which possibly represent a highly virulent group of *E. coli* O157:H7, fall into Lineage I/II. Furthermore, my data also indicated that *E. coli* O157:H7 outbreaks might be caused by two lineages of *E. coli* O157:H7 - Lineage I and I/II. Lastly, markers were identified in *rhsA* and *rhsJ*, which I suggested might be useful for defining *E. coli* O157:H7 strains from Clade 8 and Clade 6.

In summary, I conclude that *rhs* genes might be useful markers for revealing the short-term evolutionary history of *E. coli* O157:H7 during outbreaks. Therefore, an MLST scheme targeting *rhs* genes may serve as a useful tool for both inferring the phylogenetic relatedness of *E. coli* O157:H7 and for epidemiological investigations of *E. coli* O157:H7 outbreaks.

4.2 Future research

Future studies are required to investigate whether this MLST scheme targeting *rhs* genes would be useful for epidemiological investigations of *E. coli* O157:H7 outbreaks. Significant DNA-sequence variations were identified within *rhs* genes among well-studied *E. coli* O157:H7 strains. However, my data revealed that *rhs* genes alone could not separate all outbreak-unrelated strains. For example, *E. coli* O157:H7 strains of the 2006 spinach outbreak are indistinguishable from those of the 2006 Taco-Bell outbreak. I included *E. coli* O157:H7 strains from only a few outbreaks and have not tested this MLST scheme on other outbreaks. Therefore, I am interested in examining the usefulness of this MLST scheme in epidemiological investigations of *E. coli* O157:H7 outbreaks by including more strains from other well-defined outbreaks. Seven *rhs* genes should be amplified and sequenced for each *E. coli* O157:H7 strain. Sequence diversity (SNPs) should then be identified. Discriminatory power and epidemiologic concordance can be calculated accordingly to evaluate the effectiveness of this MLST scheme for subtyping *E. coli* O157:H7 in epidemiological investigations.

Future studies are required to find gene markers (preferably ca. 600 - 700 bp) that are shorter and yield more DNA sequence variations to develop a feasible and more discriminatory MLST scheme for subtyping *E. coli* O157:H7. Though *rhs* genes possess sufficient DNA sequence variations for separating well-studied *E. coli* O157:H7 strains, they are not amenable for a cost- and time-effective MLST scheme because *rhs* genes are too long for easy sequencing. The average length of *rhs* genes is ca. 4 kb, which would take the current Sanger sequencers 7 - 8 runs using internal primers to obtain the full length. Moreover, my data also showed that *rhs* genes alone could not separate all

outbreak-unrelated strains, indicating the discriminatory power of this MLST scheme might not be satisfactory. Therefore, Dr. Dudley's laboratory is currently using high-throughput sequencing techniques to search for gene markers that are shorter and possess more DNA sequence variations to allow the development of a more efficient and discriminatory MLST scheme.

Future studies are required to investigate the stability of *rhs* genes as markers in this MLST scheme. Several studies reported that the banding patterns of PFGE and MLVA were not stable during laboratory subculturing, which suggested that the genomic content of *E. coli* O157:H7 might change during outbreaks. Accordingly, a question arises for this MLST scheme: how stable are these *rhs* genes during outbreaks? In order to answer this question, ideally, I want to compare the *rhs* sequences of the ancestral strain(s) and its descendant strains from an outbreak. However, the ancestral strain(s) are no longer available. Therefore, this experiment cannot be conducted. An alternative may be to mimic the transmission of *E. coli* O157:H7 during outbreaks by passing *E. coli* O157:H7 from cattle (the reservoir host) to farm compost (the environment) to foods (transmission vehicles) and finally to animal models such as gnotobiotic piglets (infection host). The genomic DNA of an *E. coli* O157:H7 strain from the reservoir host, the environment, the transmission vehicles and the infection host can be extracted and amplified to obtain the *rhs* sequences. Next, the sequences of *rhs* genes of that *E. coli* O157:H7 from each transmission point can be compared to infer whether *rhs* genes would be stable during outbreaks.

Future studies are required to examine whether *rhs* genes would support the current clade designations. Due to the limited number of strains chosen from each clade,

phylogenetic analysis using *rhs* genes was not informative in revealing the phylogeny of the clade strains. I identified 12 STs among the 15 clade-strains, out of which 10 strains represented 10 unique STs. Most of the SNPs that could be used to differentiate strains of different STs were represented in only one strain (singleton), and are not useful for inferring the phylogeny of the 15 clade-strains as only parsimony informative sites (SNPs that are represented in at least two strains) are useful for the construction of a neighbor-joining tree. Therefore, a larger collection of O157 strains including at least 3 strains from each clade should be used. The sequences of *rhs* genes of each strain should then be obtained and a phylogenetic tree based on the *rhs* sequences would be constructed to examine whether *rhs* genes would support the clade classification.

Future studies are required to test the hypothesis that *E. coli* O157:H7 outbreaks are caused by two lineages of *E. coli* O157:H7, Lineage I and I/II. My preliminary data showed that Lineage I and I/II are the most prevalent genotypes among the 15 clade-strains, all of which are associated with human disease. However, the number of strains included in my study is too small. Therefore, a larger collection of *E. coli* O157:H7 outbreak strains from different states of the US and from different nations should be included, and lineage specific polymorphism assay (LSPA) should be applied to identify the lineage designation of these strains. The distribution of Lineage I and I/II among *E. coli* O157:H7 outbreak strains could be obtained to test whether *E. coli* O157:H7 human infections are caused by Lineage I and I/II. Dr. Dudley's laboratory is currently using LSPA to characterize clinical isolates of *E. coli* O157:H7 from the Pennsylvania Department of Health.

Future studies are required to test whether markers in *rhsA* and *rhsJ* could be used to distinguish Clade 8 and Clade 6 from other clades, respectively. My data suggested that 4 SNPs in *rhsA* and 8 SNPs in *rhsJ* are unique to Clade 8 and Clade 6, respectively. A strain collection comprising at least 3 strains from each clade should be included, and *rhsA* and *J* should be amplified and be partially sequenced to examine whether the markers in *rhsA* and *J* are unique to Clade 8 and 6, respectively.

Lastly, future studies are required to unveil the function(s) of *rhs* genes. A previous study predicted *rhs* genes were under the strongest positive selection of all the coding sequences analyzed within the *E. coli* genome, however, no studies have provided any explanations as to why this is the case. Strong positive selection indicates *rhs* genes are playing important roles in *E. coli*. Previous publications reported *rhs* genes might be involved in intestinal colonization of *E. coli* in cattle and might be involved in resistance of *E. coli* to the biocide PHMB. Moreover, *rhs* genes were predicted to encode extracellular proteins based on their amino acid sequences. Taking all these together, I speculated that Rhs proteins promote survival of *E. coli* during intestinal transit in some undefined manner, and that their extracellular nature results in strong positive selective pressure exerted by the host's immune system. To test this hypothesis, it is critical to first determine the condition(s) under which *rhs* genes are expressed. The *rhs* promoter(s) should be fused to a reporter gene (*lacZ*) and the recombined vector should be introduced into a recipient strain. The recipient strain with the recombined vector would grow under different conditions, such as LB broth supplemented with bile salt or PHMB, or LB at low pH, to resemble the intestinal transit of *E. coli* O157:H7 in human or animal hosts. The activity of the *rhs* promoter(s) should be examined. If *rhs* genes are expressed under

any of these conditions, I am interested in knocking out the *rhs* gene(s) in *E. coli* O157:H7 and doing complementation experiments to elucidate the possible function(s) of *rhs* genes in *E. coli* O157:H7.

APPENDIX A

SUPPLEMENTARY MATERIALS

SUPPLEMENTARY TABLE 1. SNP loci in *rhsA* identified by sequencing clade strains

	rhsA[1]																										
	0	0	1	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	0	1	7	8	9	0	1	1	1	1	1	2	3	3	3	4	4	4	4	5	5	6	8				
	1	3	2	1	7	2	0	2	2	7	8	3	7	8	9	0	6	7	7	1	6	4	6				
Strain	8	4	0	2	0	1	8	3	5	7	3	1	5	0	2	8	8	8	9	6	2	2	0	Clade[2]		AT[3]	
Sakai	G	T	G	C	C	A	A	A	C	T	C	G	T	C	G	A	G	G	C	A	C	C	C	1		1	
TW10022	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1		1	
93-111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2		1	
TW11308	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2		1	
EDL933	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3		2	
TW11346	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3		2	
TW11039	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4		3	
TW11052	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4		3	
TW09109	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6		3	
TW11102	.	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6		3	
TW10245	A	C	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.		7		4	
TW01663	A	C	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	7		5	
M1-Spinach	.	C	.	T	.	.	.	.	.	.	.	.	.	.	C	A	A	G	.	.	.			8		6	
G5101	A	C	.	.	T	G	G	G	G	C	T	C	C	A	A	G	.	.	.	.	.	G	.	9		7	
TW07763*	A	C	-	-	-	-	G	G	G	C	T	C	C	A	A	G	.	.	.	.	.	G	.	9		8	

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.
* TW07763 has a deletion from nucleotide position 745 to 3099 and the deleted nucleotides are indicated by dashes.

SUPPLEMENTARY TABLE 2. SNP loci in *rhsC* identified by sequencing clade strains

	rhsC[1]																																					
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
	0	1	1	2	2	7	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	1	2	2	2	2	3	4	4	4	4	4	4	5	5	5		
	6	2	5	0	1	9	9	1	1	1	1	2	2	5	7	8	8	8	8	9	9	8	0	7	9	9	1	1	5	6	6	7	7	0	0	0		
Strain	0	6	3	3	9	0	1	2	5	6	8	1	3	4	1	4	7	8	9	0	2	0	9	8	3	5	8	5	5	1	2	5	6	0	2	3	Clade[2]	AT[3]
Sakai	T	A	C	G	C	A	T	G	T	T	A	T	C	G	T	T	T	G	G	C	A	A	T	C	C	G	A	G	A	G	G	G	G	C	C	T	1	1
TW10022	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
93-111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
TW11308	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
EDL933	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11346	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	2
TW11039	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	1
TW11052	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	1
TW09109	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	1
TW11102	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	1
TW10245	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	3
TW01663	C	.	G	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	4
M1-Spinach	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	5
G5101	C	.	G	C	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	6
TW07763	C	G	G	.	.	T	C	C	C	C	G	A	.	A	C	C	C	A	A	T	C	G	C	T	T	A	C	A	G	C	A	C	C	A	A	C	9	7

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.

SUPPLEMENTARY TABLE 2 (continued). SNP loci in *rhsC* identified by sequencing clade strains

Strain	*rhsC*[1]																																			Clade[2]	AT[3]
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3		
	5	5	5	6	6	6	6	6	6	6	6	7	7	7	7	7	8	3	8	8	8	8	9	0	0	1	4	4	4	4	4	5	5	5	6		
	0	6	9	2	2	2	6	7	7	7	9	0	1	4	6	7	1	3	1	3	8	9	5	1	9	0	6	7	7	7	8	1	6	7	5		
	4	6	6	2	4	9	8	1	4	8	3	4	0	9	0	5	0	0	2	2	9	2	1	5	9	8	8	7	8	9	9	6	4	9	5		
Sakai	G	A	G	G	T	C	C	C	G	A	C	G	T	C	C	C	A	C	C	G	G	T	A	G	G	A	C	T	A	A	A	G	T	G	A	1	1
TW10022	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
93-111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
TW11308	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
EDL933	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11346	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	C	.	.	.	.	.	3	2
TW11039	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	1
TW11052	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	1
TW09109	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	1
TW11102	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	1
TW10245	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	3
TW01663	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	4
M1-Spinach	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	5
G5101	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	6
TW07763	A	C	A	C	G	T	T	T	T	G	.	C	C	G	T	T	G	.	.	A	C	C	C	A	C	G	G	C	.	C	T	A	C	T	T	9	7

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.

SUPPLEMENTARY TABLE 3. SNP loci in *rhsD* identified by sequencing clade strains

Strain	*rhsD*[1]	Clade[2]	AT[3]
	0 0 0 0 0 0 0 0 1 1 1 1 1 1 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4		
	3 6 6 6 6 6 6 9 3 8 8 9 9 9 4 6 6 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 4 4 5 5 8 0		
	4 3 3 4 6 8 8 8 7 4 8 2 3 3 9 5 6 2 2 2 2 3 3 4 4 5 5 6 8 0 1 2 2 3 4 5 8 0 0 0 1 1 2 4 5 5 6 7 8 8 9 9 0 3 4 6 2 3 3 7		
Strain	2 4 9 5 7 1 4 7 8 2 7 9 2 8 3 0 7 1 4 7 9 2 3 2 8 1 3 9 7 8 1 0 4 8 4 3 7 1 2 7 3 9 8 0 2 7 1 6 0 5 1 7 0 0 8 0 6 2 0 8	Clade[2]	AT[3]
Sakai	T A G C A C T G A C G A C T G C C C C G C G G G T C G A G C G C C A C G G C A A C T G C G A G C G G T A T A A C A G - A	1	1
TW10022	. . . . - .	1	1
93-111	. . . . - .	2	1
TW11308	. . . . A - .	2	2
EDL933	. . . . - .	3	1
TW11346	. . . . T T . . . . - .	3	3
TW11039	. . . . - .	4	1
TW11052	. . . . - .	4	1
TW09109	. . . . T A G T C . T T A G A T C A C G G A G T T C A A G T A A G G G G C A G A C C T A C C G C G C T C . - .	6	4
TW11102	. . . . - .	6	1
TW10245	. . . . - .	7	1
TW01663	. . . . - .	7	1
M1-Spinach	. . . . G . . . . - .	8	5
G5101	. C A T C A C . . . . A . . . . - .	9	6
TW07763	C . . . . . . T . . . . G T	9	7

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai.
Identical nucleotides at each SNP locus are indicated by dots. Gaps at each SNP locus are indicated by dashes.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.

SUPPLEMENTARY TABLE 4. SNP loci in *rhsE* identified by sequencing clade strains

Strain	*rhsE*[1]	Clade[2]	AT[3]
	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 3 0 0 0 0 1 3 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 0 0 0 1 1 1 3 3 3 5 8 4 6 7 7 2 5 9 9 1 4 3 3 4 6 8 8 2 4 5 5 8 8 9 9 9 0 0 1 1 1 1 2 3 3 7 8 8 9 1 3 4 5 4 6 7 1 1 3 5 1 5 6 2 0 4 6 8 1 5 4 3 6 7 2 4 9 5 7 1 4 9 4 3 9 6 8 0 8 9 7 8 4 5 6 7 8 3 8 3 2 5 1 1 3 4 6 7 4 4 1 3 1 7 4 3 8 5 6 4 7 1 7		
Sakai	C C A T C T A G C A C T G G G G G A A T T C C A C C C T A G T A T C C A G C T C C A C G G C C G T G G T C T	1	1
TW10022	. . . . .	1	1
93-111	. . . . .	2	1
TW11308	. . . . .	2	1
EDL933	. . . . .	3	1
TW11346	. . . . .	3	1
TW11039	. . . . .	4	1
TW11052	. . . . .	4	1
TW09109	. . . . .	6	1
TW11102	. . . . .	6	1
TW10245	. . . . . G C T C G T G G A A A T A	7	2
TW01663	. . . . . G C T	7	3
M1-Spinach	. . . . . C	8	4
G5101†	T T C C T C C A T C A C T A A A - - T G C T T . G T T C C C C G C T T . . . . . A A A . . C	9	5
TW07763‡	T T C C T - C A T C A C T A A A - - T G C T T . G T T C C C C G C T T . . . . . A . . A A . C T C	9	6

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai.
Identical nucleotides at each SNP locus are indicated by dots.
Gaps at each SNP locus are indicated by dashes.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.
† G5101 has a deletion from nucleotide position 780 to 783 and the other deletion from nucleotide position 792 to 794.
‡ TW07763 has one deletion from nucleotide position 153 to 377, a second deletion from nucleotide position 780 to 783, and a third deletion from nucleotide position 792 to 794.

SUPPLEMENTARY TABLE 5. SNP loci in *rhsF* identified by sequencing clade strains

rhsF[1]

Strain	0060	0126	0134	0153	0790	0791	0912	0915	0916	0918	0921	0923	0954	0971	0984	0987	0988	0989	0990	0992	1180	1209	1278	1293	1295	1318	1415	1455	1461	1462	Clade[2]	AT[3]
Sakai	T	A	T	C	A	T	G	T	T	A	T	C	G	T	T	T	G	G	C	A	A	T	C	C	G	A	G	A	G	G	1	1
TW10022	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
93-111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
TW11308	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
EDL933	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11346	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11039	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	2
TW11052	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	4	2
TW09109	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	3
TW11102	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	6	3
TW10245	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	4
TW01663	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	7	5
M1-Spinach	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	8	3
G5101	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	6
TW07763	C	G	C	G	T	C	C	C	C	G	A	.	A	C	C	C	A	A	T	C	G	C	T	T	A	C	A	G	C	A	9	7

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.

[2] Clade designations were defined by Manning et al.

[3] AT: Allelic Type.

SUPPLEMENTARY TABLE 5 (continued). SNP loci in *rhsF* identified by sequencing clade strains

Strain	*rhsF*[1]																																Clade[2]	AT[3]
	1475	1476	1500	1502	1503	1504	1566	1596	1622	1624	1629	1668	1671	1674	1678	1704	1710	1720	1749	1760	1775	1810	2812	2832	2889	2892	2951	3015	3099	4008	4083	4088		
Sakai	G	G	C	C	T	G	A	G	G	T	C	C	C	G	A	G	T	G	C	C	C	A	C	G	G	T	A	G	G	A	A	G	1	1
TW10022	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1
93-111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
TW11308	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	1
EDL933	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11346	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	1
TW11039	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	4	2
TW11052	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	4	2
TW09109	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	6	3
TW11102	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	6	3
TW10245	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	7	4
TW01663	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	T	.	.	.	.	.	.	.	.	.	7	5
M1-Spinach	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	8	3
G5101	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	9	6
TW07763	C	C	A	A	C	A	C	A	C	G	T	T	T	T	G	C	C	.	G	T	T	G	.	A	C	C	C	A	C	G	-	A	9	7

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai.
Identical nucleotides at each SNP locus are indicated by dots.
Gaps at each SNP locus are indicated by dashes.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.

SUPPLEMENTARY TABLE 6. SNP loci in *rhsI* identified by sequencing clade strains

	rhsI[1]		
	0 1		
	3 1		
	8 8		
Strain	7 5	**Clade**[2]	**AT**[3]
Sakai	C T	1	1
TW10022	. .	1	1
93-111	. .	2	1
TW11308	. .	2	1
EDL933	. .	3	1
TW11346	. .	3	1
TW11039	. C	4	2
TW11052	. C	4	2
TW09109	T C	6	3
TW11102	T C	6	3
TW10245	. C	7	2
TW01663	# #	7	NA
M1-Spinach	. C	8	2
G5101	. C	9	2
TW07763	. C	9	2

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai.
Identical nucleotides at each SNP locus are indicated by dots.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.
#Amplification of *rhsI* from strain TW01663 failed.

SUPPLEMENTARY TABLE 7. SNP loci in *rhsJ* identified by sequencing clade strains

	rhsJ[1]									
	6	6	6	7	7	7	7	8		
	1	8	9	1	4	6	6	0		
Strain	2	8	6	7	5	5	8	7	Clade[2]	AT[3]
Sakai	A	G	C	A	T	A	T	A	1	1
TW10022	.	.	.	.	.	.	.	.	1	1
93-111	.	.	.	.	.	.	.	.	2	1
TW11308	.	.	.	.	.	.	.	.	2	1
EDL933	.	.	.	.	.	.	.	.	3	1
TW11346	.	.	.	.	.	.	.	.	3	1
TW11039	.	.	.	.	.	.	.	.	4	1
TW11052	.	.	.	.	.	.	.	.	4	1
TW09109	G	C	G	G	C	G	C	G	6	2
TW11102	G	C	G	G	C	G	C	G	6	2
TW10245	.	.	.	.	.	.	.	.	7	1
TW01663	.	.	.	.	.	.	.	.	7	1
M1-Spinach	.	.	.	.	.	.	.	.	8	1
G5101	.	.	.	.	.	.	.	.	9	1
TW07763	.	.	.	.	.	.	.	.	9	1

[1] SNP positions are based upon the corresponding *rhs* gene of the *E. coli* O157:H7 strain Sakai. Identical nucleotides at each SNP locus are indicated by dots.
[2] Clade designations were defined by Manning et al.
[3] AT: Allelic Type.

Reference

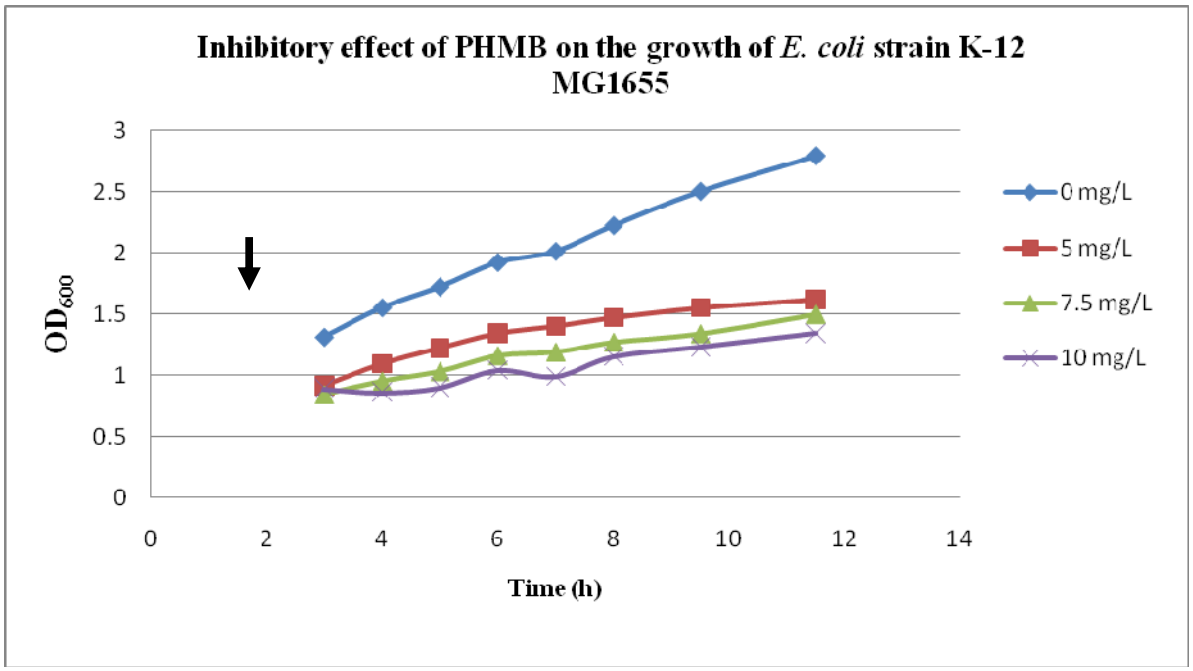
Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Alland, and T. S. Whittam. 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. Proc. Natl. Acad. Sci. U S A **105:**4868-4873.

APPENDIX B DETERMINATION OF MINIMUM INHIBITORY CONCENTRATION (MIC) OF POLYHEXAMETHYLENE BIGUANIDE (PHMB) FOR *E. COLI* O157:H7

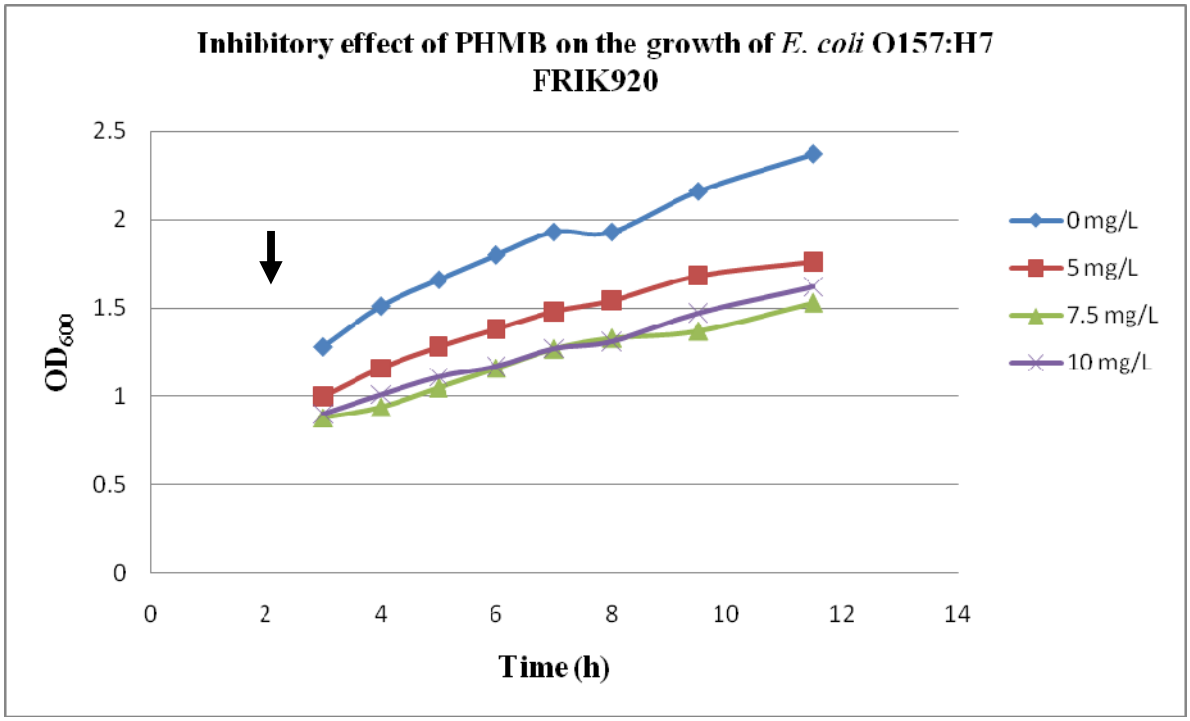
Objective: In a previous study, *rhsB* and *rhsD* were upregulated in *E. coli* strain K-12 MG1655 when it grew in Luria-Bertani (LB) broth supplemented with PHMB at MIC. Therefore, I was interested in whether PHMB at MIC would affect the expression levels of *rhs* genes in *E. coli* O157:H7. In order to test this hypothesis, I decided to first determine the MIC of PHMB for *E. coli* O157:H7. This preliminary data might facilitate future studies to elucidate the function(s) of *rhs* genes.

Method: *E. coli* O157:H7 strains Sakai and FRIK920 were used in this study. *E. coli* strain K-12 MG1655 was used as a positive control. Strains were first transferred from glycerol stock to Luria-Bertani (LB) broth, grew overnight, and then were transferred for two more times before they were inoculated into 15 mL LB broth. The inoculated 15 mL LB broth was then incubated under 37°C at 200 rpm. After 2 hours, PHMB at concentrations of 5.0 mg/L, 7.5 mg/L and 10 mg/L was added and mixed. OD_{600} was measured every hour.

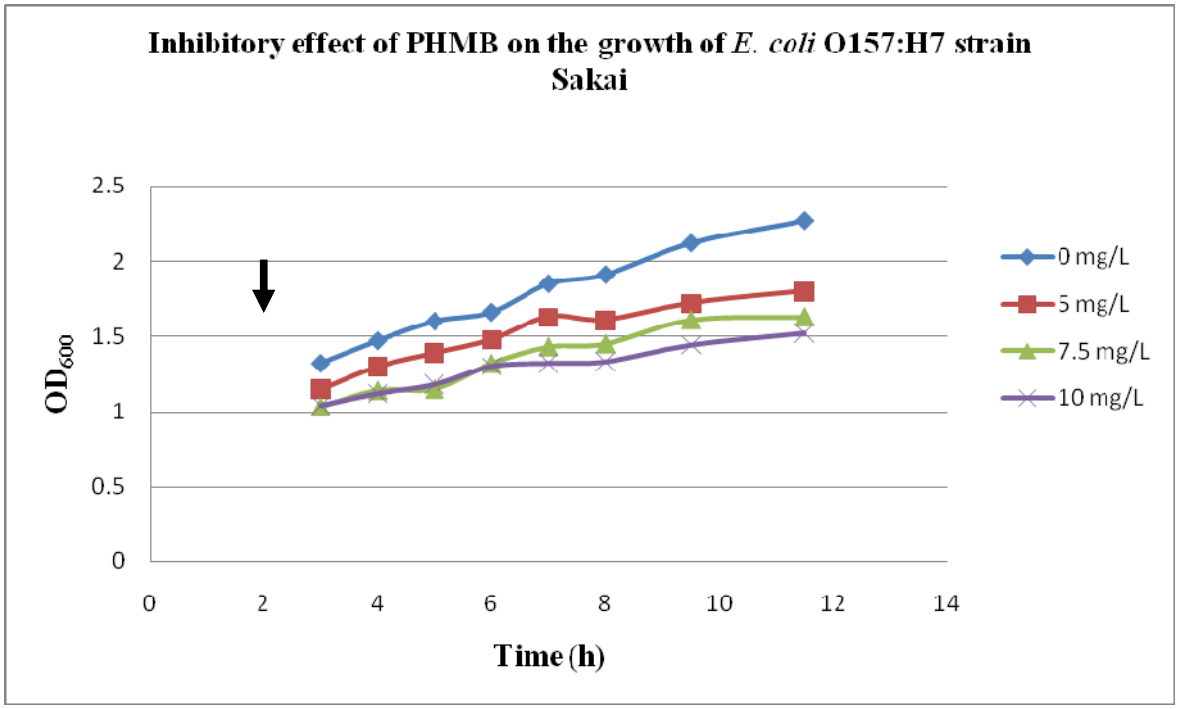
Results:



Appendix Figure 1. Inhibitory effect of PHMB on the growth of *E. coli* strain K-12 MG1655. PHMB at concentrations of 5.0 mg/L, 7.5 mg/L and 10 mg/L was added at Hour 2, as indicated by the black arrow.



Appendix Figure 2. Inhibitory effect of PHMB on the growth of *E. coli* O157:H7 strain FRIK920. PHMB at concentrations of 5.0 mg/L, 7.5 mg/L and 10 mg/L was added at Hour 2, as indicated by the black arrow.



Appendix Figure 3. Inhibitory effect of PHMB on the growth of *E. coli* O157:H7 strain Sakai. PHMB at concentrations of 5.0 mg/L, 7.5 mg/L and 10 mg/L was added at Hour 2, as indicated by the black arrow.

Conclusions:

This preliminary study demonstrates that PHMB inhibits the growth of *E. coli* O157:H7 strains Sakai and FRIK920 as well as *E. coli* strain K-12 MG1655. The inhibitory effect increases as the concentration of PHMB increases. However, replicates and statistical analysis are needed to support these conclusions.