THE GENOMIC ANALYSIS OF *ESCHERICHIA COLI* O157:H7 POPULATIONS

by

Jaehyoung Kim

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Food Science and Technology

Under the Supervision of Professor Andrew K. Benson

Lincoln, Nebraska

December, 2003

DISSERTATION TITLE

The Genomic Analysis of Escherichia coli 0157:H7 Populations

BY

Jaehyoung Kim

SUPERVISORY COMMITTEE:

Approved | Date

Signature

December 5, 2003

Dr. Andrew K. Benson
Typed Name

Signature

December 5, 2003

Dr. Robert Hutkins
Typed Name

Signature

December 5, 2003

Dr. Lloyd Bullerman
Typed Name

Signature

December 5, 2003

Dr. Raul Barletta
Typed Name

Signature

Typed Name

Signature

Typed Name

University of Nebraska | GRADUATE COLLEGE

GENOMIC ANALYSIS OF *ESCHERICHIA COLI* O157:H7 POPULATIONS

Jaehyoung Kim, Ph.D.

University of Nebraska, 2003

Advisor: Andrew K. Benson

E. coli O157:H7 comprises a geographically distributed population of highly related strains each carrying the same set of known virulence factors. These organisms are the most common cause of foodborne hemorrhagic colitis in the U.S., Canada, the U. K., and Japan. Despite the multi-locus similarity of *E. coli* O157:H7 strains, considerable genome diversity has been observed by high-resolution methods such as pulse-field gel electrophoresis. We developed a method termed octamer-based genome scanning (OBGS) to assess the genetic relatedness of *E. coli* O157:H7 isolates in detail. Our initial work by OBGS has shown that the populations can be divided into two lineages termed lineage I and lineage II. Descendants of these two lineages can be found in separated continents, suggesting that their divergence preceded geographic spread. Analysis of polymorphic loci shared by all lineage II strains further suggests that lineage II is a derived state of lineage I.

To understand the genetic basis for the divergence of the lineages, we have mapped lineage-specific polymorphisms by high-density OBGS analysis using a panel of 150 different primer combinations and a set of 40 representative lineage I and II

strains. Of 4,527 polymorphic OBGS products observed, 1,284 were lineage-specific, suggesting that they arose during divergence of the lineages. Cloning and DNA sequence analysis of 95 lineage-specific, polymorphic OBGS products identified a total of 60 independent loci. The polymorphisms included single nucleotide changes that create or destroy OBGS priming sites as well as insertions and deletions in various sizes. The largest functional category of genes carrying lineage-specific polymorphisms includes prophage genes, implying that one of the largest genetic differences between the lineages is prophage content. Two prophage-encoded, one plasmid-encoded, and four genomic polymorphisms map to genes which share homology with putative or known proteins associated with *E. coli* virulence. A total of 6 polymorphisms could confer lineage-specific abilities to survive in a certain niche. Such functional categories included central metabolism-, transport-, and regulation- associated polymorphisms. These finding support the hypothesis that some of polymorphisms could confer quantitative, lineage-specific differences in virulence or transmissibility.

ACKNOWLEDGEMENTS

I am deeply grateful to my committee members. First, I would like to thank Dr. Andrew K. Benson and Dr. Robert Hutkins for their exceptional supervision and support, as well as warm friendship. I would also thank my committee members Dr. Lloyd Bullerman and Dr. Raul Barletta for their valuable comments towards my dissertation. Especially, I would like to express my heartfelt thank to my major advisor, Dr. Andrew K. Benson for his creative and critical advices, support, guidance, and ultimate patience.

I would like to acknowledge the support that I received from the members of both the Benson and Hutkins laboratories. Especially I thank Joe Nietfieldt for assistance and critical advices for research. I also thank Zhijie Yang and Young-Jun Goh for assistance and friendship. Special thanks also go to Joy Kovar, Chao and Min Zhang, Anna Oomen, Kylie Bearnes, and Sara Brostrom for their support and help.

Especially, I would like to express my sincerest appreciation to my loving father and mother for their affection, support, and encouragement. I also would like to thank my dear father-in-law and mother-in-law for their support and affection. Special thanks go to my sisters, brother, brother-in-laws, nephew and niece. I also sincerely thank my aunts and aunts-in-laws.

I would like to give my special appreciation to my dear loving wife, Hee-Yoon Cho, for her sincere support, care, encouragement, and warmhearted affection.

Preface

This dissertation consists of Chapters 1, 2, 3, 4 and a conclusion section. Chapter 1 is a literature review of the research concerned with aspects of molecular and population genetics of *E. coli* O157:H7. Chapter 2 covers the identification of unique subpopulation of E. coli O157:H7 strains in cattle. This chapter has been unaltered from its original form submitted for publication in the Proceedings of the National Academy of Sciences of the United States of America (Kim et al. 1999. Proc. Natl. Acad. Sci. USA. 96: 13288-13293). Chapter 3 details ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of Enterohemorrhagic *E. coli* O157. This chapter also has been unaltered from its original form submitted for publication in the journal of the Bacteriology (Kim et al. 2001. J. Bacteriol. 183: 6885-6897). Chapter 4 covers the high-density analysis on the O157:H7 populations and characterizes the lineage-specific polymorphisms. This chapter will be submitted to Journal of Bacteriology. Finally, a conclusions section provides a brief overview of the major research findings contained within this dissertation

Table of Contents

List of Tables

Chapter 2

Chapter 3

Chapter 4

List of Figures

Chapter 1

Molecular and Population Genetics of *E. coli* O157:H7

Introduction

The recognition of *E. coli* O157:H7 as a pathogen and subsequent outbreaks of foodborne illness caused by this serotype has brought about significant alarm in the food industry. Although the number of cases of hemorrhagic colitis caused by enterohemorrhagic strain of *E. coli* (EHEC) are relatively low compared to other foodborne pathogens such as *Salmonella* and *Camphylobacter,* the devastating and potentially lethal sequelae such as hemorrhagic uremic syndrome (HUS) and thrombocytopenic purpura (TTP) that sometimes accompany EHEC infections has led to significant research efforts directed at understanding the prevalence, physiology, and pathology of the EHEC virotype of *E. coli.*

The O157:H7 serotype is a predominant serotype of EHEC isolated from patients in the United States (35); however, other EHEC serotypes have been also identified and are more prevalent among hemorrhagic colitis patients in other countries (12, 86). EHEC strains further comprise a subgroup of the diverse Shiga Toxin producing *E. coli* (STEC), all of which produce at least one of the two known Shiga toxins types. The EHEC are distinct from most STEC due to their unique attachment phenotype characterized by attachment and effacement (A/E) lesions (56). Population genetic analysis of EHEC and STEC strains has shown that EHEC strains comprise two divergent lineages, termed EHEC 1 and EHEC 2, that are only distantly related but apparently underwent similar pathways of virulence gene acquisition (64, 85,117). The EHEC 1 lineage is comprised solely of a geographically disseminated cluster of related genotypes bearing O157:H7 and O157:H- serotypes while the EHEC 2 lineage is serotypically and genotypically more diverse.

Studies using high resolution genome analyses such as Octamer-Based Genome Scanning (OBGS) on human and bovine isolates of *E. coli* O157:H7 have shown the existence of two distinct subpopulations (referred to as lineage I and lineage II) within the O157:H7 lineage. Descendants of both lineages can be detected in separated continents suggesting that their divergence is ancestral (49, 50). In a study of U.S. strains, human and cattle isolates are non-randomly distributed among the two lineages, suggesting that one of these lineages may be less virulent or may not be efficiently transmitted to humans from bovine sources.

The review will focus on the virulence traits and pathogenesis of *E. coli* O157:H7 including the characteristics of attachment, formation of attaching and effacing lesions, and Shiga toxins. The role of prophage and cryptic prophage found in the O157:H7 genome will be discussed, Finally, phylogeny of the O157:H7 will be reviewed.

Virulence Traits of Enterohemorrhagic *E. coli*

Hemorrhagic colitis is caused by a number of serotypes of Shiga toxin-producing *Escherichia coli* (STEC) (35). Among clinical STEC strains that have been isolated, a subset of enterohemorrhagic *E. coli* (EHEC) have been described that share sets of virulence genes. These genes encode factors for attachment to host cells, the elaboration of effector molecules, and production of two different types of Shiga toxins (56). The sets of virulence genes are found within the locus of enterocyte effacement (LEE) pathogenicity island, lambdoid bacteriophages, and a large virulence associated plasmid. Several virulence factors of EHEC infections include *eae*A gene encoding

intimin (20, 22), and other genes directly involved in producing attaching and effacing lesions in host cells. These genes are located in the Locus of Enterocyte Effacement (LEE) pathogenicity island (62). A gene encoding a hemolysin (93), a *Clostridium difficle*-like enterotoxin (3), and genes involved in adhesion (45, 22) are carried on the pO157 plasmid, and *stx*I and *stx*II genes encoding the type I and II Shiga toxins are carried on lysogenic bacteriophages (72, 75).

Pathogenesis

EHEC pathogenesis is a consequence of specialized attachment systems and the synthesis of exotoxins, known as Shiga toxins (also called Shiga-like toxins or Verotoxins). Following ingestion, the histopathological sequence of events begins by formation of microcolonies at sites in the large intestine, followed by subsequent formation of attaching and effacing (A/E) lesions, a hallmark of both EHEC and EPEC pathogenesis (68; 52). Within these lesions, bacteria become tightly bound to host cells via specialized host-derived pedestal structures. Microvilli are displaced from regions adjacent to the bound bacteria and local cytoskeletal rearrangements lead to formation of pedestals at the site of bacterial attachment (29). Although many details of the A/E process have been defined, the mechanism of diarrhea remains elusive for both EPEC and EHEC. Loss of absortive microvilli in the small intestine (70), changes in ion secretion (11), breeches of integrity of tight junctions between enterocytes (80), and changes in signal transduction (81) have all been hypothesized to play a role in the genesis of diarrhea in EPEC. The mechanism in EHEC remains unknown.

Though EPEC and EHEC share the A/E characteristic, vascular damage and secondary complications of hemolytic uremic syndrome (HUS) and thrombotic thrombocytopenic purpura (TTP) have been associated only with EHEC and specifically with the ability of EHEC populations to synthesize Shiga toxins (47). These toxins, which are secreted by the bacteria colonizing the colon, pass through the intestinal wall via transcytosis (2) and subsequently spread hemotogenously to target tissues. Ultimately, the toxins preferentially attack renal tissue due to their high affinity for receptors that are found predominantly in renal tissue . After internalization by receptor-mediated endocytosis, the active subunit of the toxin catalytically inactivates ribosomes by cleaving a residue on the 28 S ribosomal RNA (90). Cell death within the glomeruli ultimately results in tissue damage which can be further complicated ensuing inflammatory response (40).

Initial Phase Attachment

The process of A/E lesion formation in EPEC appears to comprise two stages, an initial phase of diffuse adherence followed by a secondary phase of high-affinity binding, formation of organized pedestal structures, and displacement of microvilli. Diffuse adherence in EPEC in the small intestine is mediated by the bundle-forming pilus, which facilitates formation of highly structured microcolonies on the host cell surface (21, 29). The *bfp* genes, which collectively encode the bundle forming pilli, are present on the large virulence plasmid found in EPEC strains.

In EHEC and other A/E populations of *E. coli* the *bfp* genes are absent, suggesting that multiple mechanisms of diffuse adherence have emerged among

different populations of A/E pathogens (70). The precise mechanism of diffuse adherence in O157:H7 and other EHEC remains unknown, although several candidate loci have been identified. At least one of the candidate loci is encoded on pO157, the large virulence plasmid carried by O157:H7 and other EHEC strains, suggesting the phenotype may be plasmid mediated, as in EPEC, but nonetheless by a distinct machinery. Initial reports described a fimbrial mechanism in O157:H7 that was presumably encoded by pO157 since the phenotype was lost in plasmid-cured strains (45). DNA sequence analysis of pO157 did not reveal any fimbrial genes (7, 58) although it remains formally possible that a plasmid-encoded protein could mediate expression or function of chromosomally encoded fimbrial genes. Indeed fimbrial gene clusters are present on the O157:H7 chromosome (37, 79), however their function remains to be determined. Other studies did not detect fimbriae in O157:H7 strains bearing pO157, but nonetheless showed dependence of adherence phenotypes on the pO157 plasmid (105). Conflicting data from Fratamico et al (30) indicated that pO157 was not required for attachment, implying that the diffuse adherence phenotype could be mediated by multiple systems in a strain- and environment-dependent fashion.

Recent studies by Tatsuno et al. (100) provide compelling evidence that the *toxB* locus of pO157 is necessary for diffuse adherence, since attachment of pO157-cured derivatives was shown to be defective and the phenotype was restored by introduction of a mini-pO157 derivative consisting only of the pO157 origin and the *toxB* gene. The gene *toxB* shares significant homology with the chromosomal *eaf-1* gene from O111 EHEC, and Tn-*phoA* insertions into *eaf*-1 reduce adherence in CHO cells (73). In addition to a function in adherence, ToxB may also have cytotoxicity activity since it

shares substantial similarity to the lymphotoxin LifA from O126 EPEC (51). ToxB and Eaf-1 also share N-terminal similarity with the ToxA and ToxB proteins of *Clostridium dificile* (7, 58, 73), further supporting a possible cytotoxicity activity.

The *toxB*⁻ effect on the attachment phenotype may be mediated by its action on expression of EspA, rather than by the direct function of ToxB as a receptor. EspA expression is substantially decreased in the *toxB*- background (100). This finding is also congruent with studies of attachment in O26 EHEC showing that EspA was necessary for early stages in attachment, prior to pedestal formation (23). EspA is better known for its function in formation of the type III secretion tube required to mediate the high affinity binding phase of A/E/ lesion formation. Thus, EspA could play distinct roles by mediating early stages of diffuse adherence and subsequently serving as a conduit for secretion of effector proteins required for high affinity binding.

In addition to plasmid-encoded genes, chromosomal genes from O157:H7 have been shown to confer diffuse adherence characteristics when cloned in K-12 strains of *E. coli*. Recently, Torres et al (104) identified a fimbrial operon, *lpf* ABCC'DE in EHEC that are closely related to the long polar (LP) fimbrial operon (*lpf*) of *Salmonella enterica* serovar *typhimurium*. They proposed that LP fimbriae participate in the interaction of *E. coli* O157:H7 with eukaryotic cells by assisting in microcoloby formation. A nonfimbriated *E. coli* K-12 strain transformed with *lpf* genes had increased adherence to tissue culture cells, and isogenic O157:H7 lpf mutants showed slight reduction in adherence to tissue culture. Another gene responsible for conferring the phenotype shares significant similarity to the iron regulated protein IrgA of *Vibrio cholerae* and was subsequently named Iha (Iron regulated protein Homologue Adhesin)

(99). The *iha* gene is part of an island, termed the tellurite adherence island (TAI), which encodes tellurite resistance in addition to the adhesin (99). Genome sequence analysis of the U.S. O157:H7 strain EDL 933 (79) and the Japanese O157:H7 Sakai strain (37) revealed that the TAI island is actually part of a larger 88 kb island comprising three functional segments encoding a urease system, tellurite resistance, and the Iha adhesin. The tellurite resistance region is similar to genes encoded by *Serratia marcessans* (99) while the urease cluster is similar to a urease gene cluster of *Klebsiella aerogenes* (38). The pattern of homology suggests the region is a mosaic of functional segments acquired on independent occasions. Indeed, each of the functional regions of the island is separated by stretches of DNA encoding multiple insertion sequences and putative transposases. Whether the functional mosaic was amassed before entry into the O157:H7 lineage or built up in a stepwise fashion within O157:H7 remains to be determined.

Comparison of the EDL933 and Sakai genomes shows that the entire 88 kb island is duplicated in EDL933. In both genomes, a copy of the island exists in the *serW* tRNA, but EDL933W also contains an additional identical copy located 370 kb downstream in the *serX* tRNA locus. The simplest model is that the ancestral island was duplicated and subsequently transposed in EDL933 and that the TAI found at *serW* reflects the ancestral state.

The TAI island is found among O157:H7 strains that are non-sorbitol fermenting and ß-glucuronidase negative. Populations of sorbitol-fermenting, ß-glucuronidase positive O157:H- EHEC (SFO157 EHEC) strains that are common to Germany and central Europe, however, lack this island (99) and are believed to have diverged early

during the ontogeny of the O157:H7 lineage (28). In addition to their unique sorbitol and ß-glucuronidase phenotypes, the virulence plasmid in SFO157EHEC is also distinct from the pO157 (referred to as pSFO157) and it encodes a unique fimbrial system that facilitates mannose sensitive hemeagglutination (6). Although the contribution of this fimbrial system and other systems described above to the diffuse adherence phenotype remains to be fully elucidated, it seems plausible that independent mechanisms for diffuse adherence have been acquired on different occasions during radiation of divergent subpopulations within the O157:H7 EHEC 1 lineage.

A/E Lesion Formation and LEE Pathogenicity Island

After an initial phase of diffuse adherence, the characteristic A/E lesion begins to appear marked by the emergence of pedestals located immediately below the bacterium on the host cell surface and the associated displacement of neighboring microvilli. The known genes required for A/E formation are encoded within a pathogenicity island known as the locus of enterocyte effacement (LEE). LEE has been found in various pathogenic types of *E. coli* as well as *Citrobacter rodentum*, and DNA sequence analysis of the islands from these organisms demonstrate that it is a highly conserved element (16, 62, 78, 123). Genes encoding components of the type III secretion system (*esc* genes) comprise a large proportion of the island. Several of the secreted effector proteins (EspA, EspB, EspD, and EspF) are encoded at one end of the island, and the adhesin Intimin and its Translocated Intimin Receptor (Tir) are encoded in the central portion by the *eaeA* and *tir* genes, respectively.