This is to certify that the
dissertation entitled

GENOMIC ANALYSIS OF PATHOGEN EVOLUTION: VIRULENCE GENE ACQUISITION AND GENETIC EROSION IN *ESCHERICHIA COLI*

presented by

Adam Michael Nelson

has been accepted towards fulfillment
of the requirements for the

Ph.D degree in Genetics

Major Professor's Signature

05/29/08

Date

GENOMIC ANALYSIS OF PATHOGEN EVOLUTION: VIRULENCE GENE ACQUISITION AND GENETIC EROSION IN *ESCHERICHIA COLI*

By

Adam Michael Nelson

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Genetics

2008

ABSTRACT

GENOMIC ANALYSIS OF PATHOGEN EVOLUTION: VIRULENCE GENE ACQUISITION AND GENETIC EROSION IN *ESCHERICHIA COLI*

By

Adam Michael Nelson

Escherichia coli is a gram-negative, rod-shaped bacterium that lives naturally and commensally in the intestinal tract of humans and other mammals. However, some types of *E. coli* have acquired genetic elements encoding virulence factors that contribute to disease. Different disease phenotypes are caused by strains carrying a variable array of virulence factors. Many of these virulence factors are largely acquired through horizontal transfer. Pathogenic *E. coli* are grouped into at least twelve classes, or pathotypes, based on the type of disease they cause. Observing the acquisition of virulence elements as a means to study evolution has been previously used to determine the genetic ancestry of *E. coli* and other pathogenic microorganisms. Conversely, gene loss can also be important in enhancing pathogenicity by removing genes that encode proteins that may hinder increased virulence, or are no longer functional because they are mutated or otherwise incomplete. The research presented here is intended to enhance the current understanding of the roles of genetic acquisition and genetic loss to virulence changes in pathogenic *E. coli* by applying three paths of inquiry.

To address correlations between virulence profiles and disease incidence, 392 *E. coli* isolates from 115 pediatric patients were screened for virulence gene content using fluorescently labeled PCR amplicons in a capillary based sequencing system. Virulence profiles were compared to a phylogenetic framework to determine virulence distribution, and correlations between presence of specific genes and the incidence of disease in

patients. The screening of this pediatric population lead to the discovery of a variant of *E. coli* (sequence type 29), which was found very frequently (19%) in all of the samples examined. This population of ST-29 isolates was further characterized by PCR to determine the frequency of a panel of attachment-related loci, and by RFLP to determine the capsular polysaccharide type.

To track genetic erosion events, PCR-based screening of individual components in a pathogenicity island, called ETT2, showed examples of gene loss that rids possibly non-functional genetic material from the genome. When a strain acquires new genetic material by horizontal transfer, the new material must provide some fitness benefit or it will be under selection to be removed. Type III secretion systems are complicated structures that require a large number of genes to encode the proteins necessary for proper assembly. If critical components of the type III assembly are missing, the assembly will not function properly, and is likely to be under selective pressure for deletion. A streamlined genome may result in a more efficient pathogen. Here, the loss of all or portions of ETT2 is shown in a variety of pathogenic isolates of *E. coli*. At least six different deletion variants were discovered in the 57 strains examined.

These results help advance our understanding of evolution in *E. coli* by both acquisition and loss of virulence elements and further demonstrate the dynamic and diverse nature of the *E. coli* genome.

ACKNOWLEDGEMENTS

I would like to thank Thomas Whittam, my mentor, for his invaluable support and guidance, and for the opportunity to work on a project I really enjoyed during my time at Michigan State University. I also thank my guidance committee members, Drs. Michael Bagdasarian, Linda Mansfield and Vincent Young for their suggestions and support.

I have met so many people in our lab who were both friends and helpful co-workers that were crucial to the completion of this research. I wish to thank Galeb Abu-Ali, Dr. Teresa Bergholz, Sivapriya Kailasan, Sara Kienzle, Dr. David Lacher, Albert Lee, Dr. Shannon Manning, Lindsey Ouellette, Dr. Weihong Qi, Dr. James Riordan, Dr. Hans Steinsland, Dr. Cheryl Tarr, Dr. Seth Walk, and Dr. Lukas Wick.

I owe a great debt for the support of my family during my graduate career. I wish to thank my parents, Bruce and Sue Nelson, my brothers, Josh and Jordan, my wife, Suzanne, and my daughter Emily for their encouragement, patience and love.

TABLE OF CONTENTS

LIST OF TABLES..... viii

LIST OF FIGURES..... ix

KEY TO ABBREVIATIONS..... x

Chapter 1. Literature Review..... 1
- PATHOGENIC TYPES OF *E. coli*..... 2
- PARALLEL EVOLUTION OF VIRULENCE..... 4
- EVOLUTION BY HORIZONTAL TRANSFER OF GENETIC MATERIAL..... 5
- GENE LOSS AND THE EFFECT ON VIRULENCE – GENETIC EROSION OF *E. coli* TYPE III SECRETION SYSTEM 2..... 5
- DETERMINATION OF VIRULENCE PROFILES..... 8
- PURPOSE..... 9
- HYPOTHESES TO BE TESTED..... 9

Chapter 2. Genetic and virulence characterization of *Escherichia coli* isolated from pediatric patients in Seattle, WA..... 12
- SUMMARY..... 13
- INTRODUCTION..... 14
- MATERIALS AND METHODS..... 20
  - Bacterial Strains and DNA Isolation..... 20
  - PCR Primer Design..... 21
  - PCR of MLST Genes..... 24
  - Phylogenetic Analyses of Sequence Data..... 24
  - Virulence Genes and Protein Functions..... 25
  - MVGP Procedure Summary..... 27
  - PCR of MVGP Genes..... 28
  - Controls for MVGP..... 28
  - MVGP Data Confirmation..... 29
- RESULTS..... 30
  - Genetic Diversity of MLST Loci..... 30
  - Sequence Type Diversity and Clonal Groups..... 30
  - MVGP Analysis..... 33
  - Comparison of Virulence Gene Profiles and Phylogenetic Relationships..... 33
- DISCUSSION..... 39
  - Diversity of Virulence Gene Profiles..... 39
  - Frequency of ST-29 Isolates in this Sample Set..... 40
- ACKNOWLEDGEMENTS..... 42

Chapter 3. Characterization of a common meningitis-associated clone (ST-29) of *Escherichia coli* found in pediatric patients in Seattle, WA
SUMMARY 43
INTRODUCTION 44
MATERIALS AND METHODS 45
PCR Primer Design 49
Strain Growth Conditions and DNA Extractions 49
MVGP PCR 49
Controls 49
MLST PCR 49
Phylogenetic Analyses 50
PCR for Attachment Loci 50
Capsular Typing – PCR 50
Capsular Typing – Restriction Digestion 50
RESULTS 53
Attachment Genes 55
Capsular Typing 55
DISCUSSION 57
ACKNOWLEDGEMENTS 59

Chapter 4. Genetic erosion of *Escherichia coli* type III secretion system 2 (ETT2) in O157:H7 isolates 60
SUMMARY 61
INTRODUCTION 62
MATERIALS & METHODS 68
PCR Primer Design 68
Strain Growth Conditions and DNA Extractions 68
PCR 68
RESULTS 70
ETT2 in EHEC O157 71
ETT2 in NON-O157 STEC 71
ETT2 in O55:H7 71
DISCUSSION 79
ACKNOWLEDGEMENTS 82

Chapter 5. Summary and Synthesis 83
Future Considerations 85

REFERENCES 89

LIST OF TABLES

Table 1. Patients in the Seattle pediatric study grouped based on the diagnosis of microorganisms..... 19

Table 2. Oligonucleotide primers used in the Seattle pediatric study..... 22

Table 3. Virulence genes associated with diarrheal disease..... 26

Table 4. Sequence variation among alleles of seven MLST genes..... 31

Table 5. Summary of virulence gene detection frequency in sequence types and in patients from the Seattle pediatric population study..... 34

Table 6. Gene and protein data for attachment-related loci used in PCR assays of ST-29 isolates..... 52

Table 7. Comparison of genes from the *Salmonella* Pathogenicity Island 1 (SPI-1) of *Salmonella* enterica serovar Typhimurium and their corresponding homologous counterpart in ETT2..... 65

Table 8. Primer sequences used to amplify individual genes within ETT2..... 67

Table 9. Results from the PCR assay of individual ETT2 genes within a variety of *E. coli* serotypes and control strains. 73

LIST OF FIGURES

Figure 1. Fluorescent chromatogram output from the multilocus virulence gene profiling (MVGP) procedure..... 18

Figure 2. Phylogenetic tree showing the genetic relatedness of pediatric *E. coli* in Seattle, WA..... 32

Figure 3. Virulence type diversity in two representative patients..... 35

Figure 4. Non-metric dimensional scaling (NMDS) analysis of MVGP results of 94 patients based on the presence or absence of 29 known or putative virulence factors..... 38

Figure 5. Organization of the *kps* operon in *E. coli*..... 47

Figure 6. Representative banding pattern from a Sau96I restriction endonuclease digest of the amplified kpsCM region in *E. coli*..... 48

Figure 7. Frequency distribution of each gene from MVGP analysis of 392 isolates from 94 pediatric patients. 54

Figure 8. Organization of *ETT2* gene cluster in *E. coli*..... 66

Figure 9. Pairing of the phylogenetic analysis and ETT2 deletion profiles in many pathogenic types of *E. coli*..... 76

Figure 10. Map of the region of ETT2 that was examined by PCR, and the results of that screen. 77

Figure 11. Stepwise evolution model for the emergence of the modern pathogenic clone of EHEC O157:H7 from ancestral O55:H7 isolates.. 78

Figure 12. The EIP island from enteroaggregative strain 042..... 81

KEY TO ABBREVIATIONS:

***E. coli* Pathotypes**:

EPEC Enteropathogenic *E. coli* - cause watery diarrhea in infants, can form microcolonies and A/E phenotype. EPEC contain the LEE island and a large adherence-related plasmid.

EHEC Enterohemorrhagic *E. coli* - cause severe bloody diarrhea, HC, developing sometimes into HUS. EHEC encode Shiga toxins, a large virulence plasmid, and the LEE island.

ETEC Enterotoxigenic *E. coli* - cause travelers diarrhea, similar to cholera, but not as severe. ETEC can contain heat stable and/or heat labile toxins.

EIEC Enteroinvasive *E. coli* - can invade epithelial cells through proteins encoded on a virulence plasmid. EIEC have a similar invasive phenotype to *Shigella.*

STEC Shiga toxin-producing *E. coli* - any strain possessing one or both of the Shiga toxins (Stx1/Stx2). EHEC are a subset of STEC.

EAggEC Enteroaggregative *E. coli* - cause mild but prolonged watery diarrhea.

Phenotypes:

A/E Attaching/Effacing phenotype seen in EPEC and EHEC. Characterized by effacement of microvilli, intimate attachment, and pedestal formation due to actin polymerization. Characteristic of strains harboring the LEE island.

HC Hemorrhagic colitis - severe bloody diarrhea caused by EHEC.

HUS Hemolytic uremic syndrome - life-threatening disease caused by EHEC - characterized by acute renal failure, the destruction of platelets and destruction of red blood cells.

Virulence Factors:

stx1/stx2 Shiga toxin 1 or Shiga toxin 2 genes - both encoded on integrated bacteriophages in EHEC and STEC – encode cytotoxins which inhibit protein synthesis and cause characteristic HC and HUS symptoms.

KEY TO ABBREVIATIONS (cont'd):

ETT2 — *E. coli* type III secretion system 2 - cryptic island, widely distributed, but usually deleted, possibly because it no longer functions as a TTSS. ETT2 may encode structural proteins for type III secretion, but likely not effectors.

EIP — Largely uncharacterized island encoding possible effectors for ETT2. Found in EAggEC strain 042, but not in O157.

TTSS — Type III secretion system - contact dependent method for directly injecting bacterial proteins into the host cell by use of a needle-like complex that spans both the inner and outer bacterial membranes.

LEE — Locus of enterocyte effacement - pathogenicity island encoding both structural components and effector molecules for a functional type III secretion system in EHEC and EPEC.

Effector — The actual proteins injected into the host cell using a TTSS. Effectors are responsible for eliciting the pathological changes seen during A/E.

Strains of *E. coli*:

ECOR — *E. coli* reference collection - diverse set of strains of *E. coli* maintained by Dr. Thomas Whittam at Michigan State University.

K12 — Domesticated, non-pathogenic strain of *E. coli* lacking many virulence factors characteristic of pathogenic strains (LEE island, Shiga toxins, etc).

Molecular Techniques:

MLST — Multi locus sequence typing - a DNA sequencing technique using sequencing of 450-600 nucleotides of conserved genes to generate phylogenetic comparisons.

MVGP — Multi locus virulence gene profiling - typing system to examine 29 virulence genes & 1 positive control in uncharacterized *E. coli* isolates.

CEQ — Genetic analysis machine developed by Beckman Coulter for use with DNA sequencing (MLST) and virulence profiling (MVGP).

ORF — Open reading frame - DNA sequences containing a putative gene.

CHAPTER 1

LITERATURE REVIEW

Commensal strains of *E. coli* exist naturally in the gut of humans and other mammals without causing disease in healthy individuals. However, a small subset of *E. coli* have acquired additional genetic elements encoding the production of toxins and other virulence factors, such as invasion ability, secretion systems, and increased iron utilization. Pathogenicity islands are horizontally-acquired genetic elements (plasmids, phages, chromosomal islands) that increase pathogenicity in recipient strains. DNA fragments of various sizes containing genes or groups of genes can be transferred between strains or even between bacterial species, and are a force driving rapid evolution.

Pathogenic types of *E. coli*. At least 12 different classes, or pathotypes, of *E. coli* have been identified based on distinct disease phenotypes [1, 2], and often contain different pathogenicity islands encoding virulence factors responsible for conferring the different disease phenotypes. This allows grouping of strains based on similar disease-causing ability, rather than serological similarities based on typing of the somatic O-antigen and flagellar H antigens. So while two strains may share the same serotype (O103:H2, for example), they may not necessarily share the same virulence profile. Although the more closely related two strains are genetically, the more likely they will have the same array of virulence factors and cause the same disease. The major pathotypes of *E. coli* used in this work, and common virulence-related genes found within those pathotypes, will be discussed below.

In the gut, enteropathogenic *E. coli* (EPEC) strains have the ability to form microcolonies through localized adherence by genes encoded on a large virulence plasmid [3]. EPEC also contain numerous pathogenicity islands, including the EspC

island, which encodes a serine protease autotransporter toxin (*espC*) that aids in virulence, [4, 5] and the locus of enterocyte effacement (LEE) island which encodes a type III secretion system and the ability to form attaching-effacing (A/E) lesions on intestinal epithelial cells [6, 7]. A/E lesions are characterized by brush border microvilli destruction, intimate attachment to the host cell membrane, and the formation of a pedestal structure in the host cell through the polymerization of actin filaments [6, 8].

Enterohemorrhagic *E. coli* (EHEC) strains also encode LEE and can elicit A/E lesions as well. Unlike EPEC, they contain Shiga toxins (*stx1* and/or *stx2*) and the EHEC virulence plasmid. This plasmid contains hemolysin (*ehx*) for iron piracy, adhesion factors (*toxB, iha*) for attachment to the intestinal epithelia, iron utilization genes (*chuA*) to process iron, and a protease autotransporter (*espP*) to cause proteolytic damage and inhibit blood coagulation [9, 10].

Enterotoxigenic *E. coli* (ETEC) strains can encode two distinct toxins responsible for increased fluid accumulation in the gut lumen. The heat-stable (*estA*) toxin, which binds to and activates guanylate cyclase, causes the increase of chloride ion secretion resulting in the rapid accumulation of intestinal fluid. The heat-labile (*elt*) toxin causes a disease similar to cholera (heavy fluid and electrolyte accumulation in the gut), because it shares the same structure and mode of action as cholera toxin from *Vibrio cholerae* [11-13].

Enteroinvasive *E. coli* (EIEC) strains cause dysentery and have invasive phenotypes similar to *Shigella*, possessing a plasmid essential for invasion ability [14, 15]. Enteroaggregative *E. coli* (EAggEC) are able to adhere to HEp-2 cells in an aggregative or 'stacked brick' pattern of autoagglutination. These strains encode some

enterotoxins, including a plasmid-encoded serine autotransporter toxin (*pet*) [16, 17], *Shigella* enterotoxin 1 (*setIA*) [18-20], and the EAST heat-stable toxin (*astA*), which is also found in some EPEC strains [21]. EAggEC can cause mild, but prolonged watery diarrhea lacking fever or blood loss, however symptoms are frequently non-uniform [22-25].

Shiga-toxin producing *E. coli* (STEC) are strains grouped by one common factor; they contain one or both of the Shiga toxin genes (*stx1, stx2*) [26-28]. Since *stx1* and *stx2* are encoded on separate, mobile bacteriophages, STEC strains are often not closely related, but rather have only acquired and incorporated the *stx1* and/or *stx2*-containing phages into their genome.

Parallel evolution of virulence. Strains from each pathotype are often not evolutionarily related, but rather have acquired similar pathogenic ability through parallel acquisition of the same or similar virulence factors. Despite the harm caused to the host, selection favors increased virulence when it allows for better spread or transmission of the pathogen [29]. Often better transmission equals more severe disease symptoms.

The discovery of a new pathogenic variant of *E. coli* capable of causing widespread foodborne illness occurred in 1982 [30, 31]. Since then, this pathogenic variant, O157:H7, has been responsible for an estimated 74,000 infections each year in the U.S. [32], causing bloody diarrhea and hemolytic uremic syndrome, a life threatening illness characterized by acute renal failure, hemolytic anemia, and reduced ability to form blood clots through the destruction of platelets [28, 33]. Since O157:H7 strains also contain the LEE island, they have the ability to form the characteristic A/E lesions on epithelial cells, just as seen with EPEC.

Evolution by horizontal transfer of genetic material. The genome of *E. coli* is extremely heterogeneous, having undergone remarkable sequence divergence between the common pathotypes that cause disease. Horizontal transfer of genetic material is common, with striking variability between O157:H7 isolates and the commensal K12 strain. Nearly 1400 genes in O157:H7 are unique compared to K12, including many that encode virulence functions. The genome is nearly 20% larger in O157:H7 strains vs. K12 (5.5MB vs. 4.6MB) [34]. Other pathotypes are also highly divergent. For example, the recent genome sequencing of the uropathogenic strain CFT073 revealed only 39.2% of the total protein content in CFT073, the laboratory K12 strain MG1655, and EHEC O157:H7 strain EDL933 is conserved between all 3 strains. CFT073 is as different genetically from the commensal K12 strain as it is from EDL933 [35].

Gene loss and the effect on virulence - Genetic erosion in *E. coli* type III secretion system 2 (ETT2). The publication of genomic DNA sequences has facilitated rapid comparisons between strains to identify unique genes or pathogenicity islands. EPEC and EHEC encode a well-characterized TTSS called LEE, which is known to be functional [7, 36, 37]. Comparative genomic scans were used to discover a cryptic pathogenicity island named ETT2 (for *E. coli* Type III Secretion System 2) in O157:H7 strains [38]. ETT2 encodes a putative TTSS in select serotypes of *E. coli*, including O157:H7 [39]. Type III secretion systems need two basic sets of genes necessary for proper secretion. The structural genes encode proteins that assemble into a needle complex, with a basal component spanning both the inner and outer bacterial membrane

and a needle-like component that forms a molecular syringe that allows for direct transfer of bacterial proteins into the host cell [40]. It is thought that the structural genes for a second TTSS in *E. coli* O157 could be encoded in ETT2 [41].

However, the other crucial component needed for TTSS are the effector proteins, which are the injected proteins that travel through the type III needle complex into the host cell. Effectors then elicit pathological changes in the host [40]. Some effectors that work with ETT2 may be encoded on a separate pathogenicity island called EIP. Putative effectors with sequence homology to effectors that are secreted through two distinct TTSS in *Salmonella* (SPI-1 and SPI-2) are found in some strains of *E. coli* [41]. ETT2 shares sequence homology and genetic organization with the SPI-1 TTSS [39]. This suggests EIP-encoded effectors may work with the structural framework encoded by ETT2, since both loci may share a common origin from in *Salmonella*.

In LEE, effectors cause pathological changes, such as actin polymerization and pedestal formation. LEE has both the structural genes and the effectors encoded together in the same island. In contrast, ETT2 and EIP are separate islands and rarely found together [41].

The secretion framework is an energetically expensive structure to assemble, and likely provides little or no fitness benefit without the accompanying effectors necessary for pathogenic effects on the host cell. Genes encoding structural components are conserved throughout various gram-negative bacterial pathogens encoding TTSS. However, the effector molecules secreted into the host cell are often highly divergent and cause distinctly different effects [42, 43]. In EPEC and EHEC, these changes are characterized by massive actin polymerization causing structural rearrangement of the

eukaryotic cell surface into a pedestal shape that is essential for a close association between the *E. coli* and the eukaryotic cell. The intimate association is necessary for the onset of diarrheal disease characteristic of EPEC and EHEC infection [7, 36, 37]. It has not been demonstrated that ETT2 & EIP confer a similar phenotype. However, ETT2 & EIP may be important disease determinants when found together in the same strain. When they are not together, the numerous deletions observed in ETT2 may be an example of genetic refinement through erosion of factors not enhancing virulence.

Ren and colleagues showed a widespread distribution of individual genes or groups of genes in ETT2 [41]. This included parts of ETT2 in K12, which does not cause disease. Despite the diverse distribution, most strains examined had deletions that eliminated major sections of ETT2. Only a few strains contained all of the island, including all O157 isolates, EAggEC strain 042, and several strains from the *E. coli* Reference Collection (ECOR), a diverse set of 72 strains assembled from a variety of locations and hosts [44]. All strains showing deletions in ETT2, including K12, are apparently missing the EIP island [41].

ETT2 likely had a distant acquisition point and has been selectively eroded in strains where it was lacking EIP, and was possibly non-functional. A variety of deletions of ETT2 are evident in all strains that do not also encode EIP, except for O157 strains, which curiously have retained an intact set of structural TTSS components, despite containing several frameshift mutations [41]. This may represent the first stages of genetic erosion, preceding the larger deletions seen in K12 and other isolates. Only a few strains have been identified that contain both ETT2 and EIP. One example is an EAggEC strain called 042. To date, no strains have been identified as encoding only EIP but

without ETT2. Conversely, all strains with deletions in ETT2 are missing EIP, except for O157:H7 isolates, as previously mentioned.

Determination of virulence profiles. A large-scale comprehensive assessment of the prevalence of many virulence genes in clinical samples has not been done for pathogenic *E. coli*. Existing studies have only examined smaller numbers of specific strains. These studies were not randomized, nor did they examine more than a handful of virulence factors [45-47]. The genetic elements necessary for enhanced virulence in the various pathotypes are not fully understood, despite the importance of these pathogens. In addition, the effect of strain mixtures providing complementing virulence factors to cause disease is poorly understood. Multiple samples from the same patient may show the patient is infected with more than one distinct pathogenic clone of *E. coli*. Two techniques were used to address these issues: the identification of phylogenetic frameworks and the generation of virulence profiles in each strain.

Multi-locus sequence typing (MLST) compares the internal sequences of housekeeping genes (n=7), which are considered to be under stabilizing selection resulting in minimal variation in functionally essential proteins. Phylogenetic trees are constructed by statistical software that compares genetic relatedness of nucleotide sequences. This technique has been used successfully in a variety of microorganisms [48-51], including *E. coli*.

Multi-locus virulence gene profiling (MVGP) was been developed to rapidly characterize outbreak strains and new serotypes of pathogenic *E. coli,* and to discover novel virulence profiles in each pathotype. MVGP works by combining fluorescently

labeled PCR amplicons with capillary-based sequencing tools to generate chromatograms displaying virulence profiles for individual strains. Each profile identifies individual virulence factors as peaks of defined size and dye label, and provides a fingerprint of the virulence content of each strain. When using samples isolated from patients displaying disease symptoms of unknown etiology, MVGP can help identify virulence factors important for causing or contributing to disease.

Purpose. The primary objective of this research is to identify examples of evolution through gene loss and gene acquisition in a variety of genes associated with pathogenesis. In patients with unexplained cases of diarrheal disease, analysis of the *E. coli* collected from them was screened for a panel of known and putative virulence genes that may have contributed to illness. It is important to note that the presence of a virulence-related gene in an isolate of *E. coli* is not automatically correlated to either the expression of that gene or the ability of that strain to cause disease symptoms, but it at least makes that strain suspect, and worth investigating further.

The phenomenon of gene loss enhancing virulence or genetic fitness has been examined with the erosion of a cryptic pathogenicity island in *E. coli* called ETT2. This island is believed to contain the structural genes to assemble a type III secretion system. ETT2 is found in O157:H7 isolates, but has numerous deletions in many other serotypes of *E. coli*. The numerous deletions may be due to selection acting on a non-functional pathogenicity island to remove unnecessary genetic material and streamline the genome.

Hypotheses to be tested. We will test the following hypotheses from chapter 2:

1. There will be a correlation between the number of virulence genes per patient and an increase in disease duration. We will use chi-square tests to determine if there is a

correlation between virulence genes (0-5 genes) vs (6 or more) and the duration of illness (0-3 days), (4-7 days), and (8 or more days).

2. When patients are put into two categories, a diagnosis group consisting of all the possible diagnoses (*Salmonella*, rotavirus, etc) combined, there will be more virulence genes in those patients with a diagnosis than in patients without a diagnosis. We will also use chi-square to test this hypothesis using groups for genes (0-5) and (6 or more) per patient).

From chapter 3, we will test these hypotheses:

1. Because the clinical isolate RS218 is a ST-29 strain and has a K1 capsule type, we predict all or nearly all pediatric ST-29 isolates from Seattle will also have the K1 capsule type. A capsule typing method using restriction fragment length polymorphism will be used to determine capsule profiles.

2. We also predict the prevalence of ST-29 isolates is related to an enhanced ability to adhere to the intestine. Therefore we predict we will find some attachment-related genes common in extra-intestinal *E. coli* to be in very high frequency (near 100%) in the ST-29 strains. A panel of genes encoding proteins related to attachment will be screened by individual pairs of PCR primers.

From chapter 4, we will test these hypotheses:

1. The region of ETT2 we are examining will be found intact and undeleted in all strains from the ancestral stepwise lineage related to modern O157:H7 isolates. Strains that represent the stepwise changes from ancestral EPEC-like O55:H7 to the modern EHEC O157:H7 will be screened by PCR for 11 individual genes within a 17-kb region encoding the structural apparatus of the type III secretion system.

2. We predict other non-O157 isolates will have a variety of deletion profiles, including some newly described deletion types. In addition to strains from the O157 stepwise model, additional strains from other pathogroups will be examined and we predict new deletion profiles unseen in previous studies will be found.

CHAPTER 2

GENETIC AND VIRULENCE CHARACTERIZATION OF *Escherichia coli* ISOLATED FROM PEDIATRIC PATIENTS IN SEATTLE, WA

SUMMARY

A collection of 392 *E. coli* isolates were examined for virulence content using multi-locus virulence gene profiling. These strains were taken from children hospitalized with diarrheal disease of unknown cause. More than half of the *E. coli* strains were from patients where no identifiable pathogenic organism was isolated (*Campylobacter jejuni*, *Clostridium difficile*, rotavirus, *Shigella*, etc). The *E. coli* from these patients was examined for the presence of 29 known and putative virulence genes that serve as both general markers for all major pathotypes of *E. coli* and specific markers for certain pathogenicity islands or plasmids. These strains were also sequenced for 7 housekeeping genes to determine clonal relationships and generate phylogenetic trees. The virulence results from MVGP were paired with the phylogenetic framework generated by MLST to infer patterns of acquisition in specific virulence genes and to study their distribution across different clonal groups of *E. coli*.

INTRODUCTION

Escherichia coli are Gram-negative bacteria that include harmless, useful, and harmful organisms within this genospecies [1]. There are a variety of pathogenic variants of *E. coli* that can not only be group based on sequence similarities, but also by the characteristics of the diseases that they cause. Often virulence factors are frequently or exclusively found in specific pathotypes, and are necessary to cause the specific disease associated with each pathotype [2]. *E. coli* can frequently acquire new virulence abilities by horizontal transfer of genetic material [34, 35]. Despite the possible role of a number of these putative virulence loci in disease, many have not been extensively studied. Correlations between specific virulence profiles in clinical samples of *E. coli* and the incidence of disease in patients have not yet been established.

In addition to well-defined pathogenic clones, the effect of strain mixtures that provide complementing virulence factors is poorly understood. Analysis of multiple isolates from the same patient might show if a patient is infected with more than one distinct pathogenic clone of *E. coli*. Two techniques have been used to address these issues: the generation of phylogenetic frameworks by multi-locus sequence typing (MLST) [52-54], and the identification of virulence profiles in each strain by multi-locus virulence gene profiling (MVGP) [47, 53, 55, 56].

MLST is a procedure that involves the sequencing of genes under stabilizing selection to generate phylogenies that show the evolutionary history of the strains under examination. Here, we have used a panel of seven "housekeeping" genes encoding proteins that serve basic biological functions in the cell. Once the sequences are assembled, they are concatenated and analyzed for single nucleotide variations to allow

for the generation of phylogenies. Based on the pattern of alleles for each gene, an allele assignment can be made. When all of the allele assignment numbers are put together for each of the genes sequenced, a "barcode" of that particular strain is generated. The barcode represents a particular sequence type (ST), which is unique. So when two strains share the same sequence type, every nucleotide sequenced is exactly the same in all seven genes examined. MLST has been used repeatedly to characterize *E. coli* populations [56-58]. One major advantage of this system is that the data is portable and can easily be compared with results from other laboratories, provided the same primers were used to sequence the same genes in both labs.

MVGP uses similar reagents and equipment as MLST, but generates data in a different manner. It works by combining fluorescently labeled PCR amplicons with capillary-based sequencing tools to generate chromatograms displaying virulence profiles for individual strains. This technique allows rapid generation of the virulence profile of bacterial isolates. Each profile identifies individual virulence factors as peaks of defined size and dye label, and "fingerprints" the virulence content of each strain. When using isolates from patients with illnesses of unknown etiology, MVGP can help identify virulence factors that may be important for causing or contributing to disease. This technique also can be applied to any set of strains of *E. coli* and can be customized by adding or removing virulence loci, so newly identified virulence genes can be incorporated, if desired. MVGP has already been used to characterize a number of populations of *E. coli*, including clinical STEC isolates of serotype O121:H19 [47], *E. albertii,* which is closely related to *Shigella boydii* type 13 [59] , and *E. coli* serogroups O174 [60].

The MVGP procedure consists of the following steps: 1) collection of bacterial strains for study; 2) extraction of DNA; 3) PCR using one regular and one Beckman dye-labeled primer; 4) pooling of labeled post-PCR products together; 5) running the sample on a Beckman CEQ sequencing machine; and 6) analysis of the results to determine virulence profiles. The Beckman dyes are available in 3 colors: D2 (black), D3 (green), and D4 (blue). The red dye (D1) is reserved for a size standard and should not be used to label PCR products. Post-PCR amplified product can be pooled together (we used 14 and 16 genes in two sets). The data output is shown as a chromatogram with peaks of increased fluorescent intensity at a specific size representing each amplicon (Figure 1).

Here we used MLST and MVGP to examine 392 pediatric *E. coli* from patients with diarrheal disease of unknown cause. Samples were collected for this study at the University of Washington and Children's Hospital & Regional Medical Center in Seattle, WA. The *E. coli* isolates were collected from stool samples from patients from a number of categories based on the identification of other microorganisms (Table 1). Up to five isolates from each patient were taken for comparison to improve the statistical power of the results, and to determine if individual patients have mixtures of strains with distinct, and possibly synergistic, virulence profiles. So, for example, in patients diagnosed with specific pathogens such as *Salmonella*, five *E. coli* isolates were also collected from those patients for use in this study. Each isolate was confirmed to be *E. coli* by PCR testing for the presence of the *uidA* gene, which encodes beta-glucuronidase, an indicator of *E. coli*. To blind the study from any potential bias, all samples were identified only by a unique number. It was not until after typing was finished that epidemiological info (diagnosis, illness duration, fever, vomiting, and bloody diarrhea) for specific strains was revealed.

We hypothesized that there would be a correlation between the number of virulence genes per patient and an increase in disease duration. Also, when patients are put into two categories, a diagnosis group consisting of all the possible diagnoses (*Salmonella*, rotavirus, etc) there will be more virulence genes in those patients than in those patients without a diagnosis.

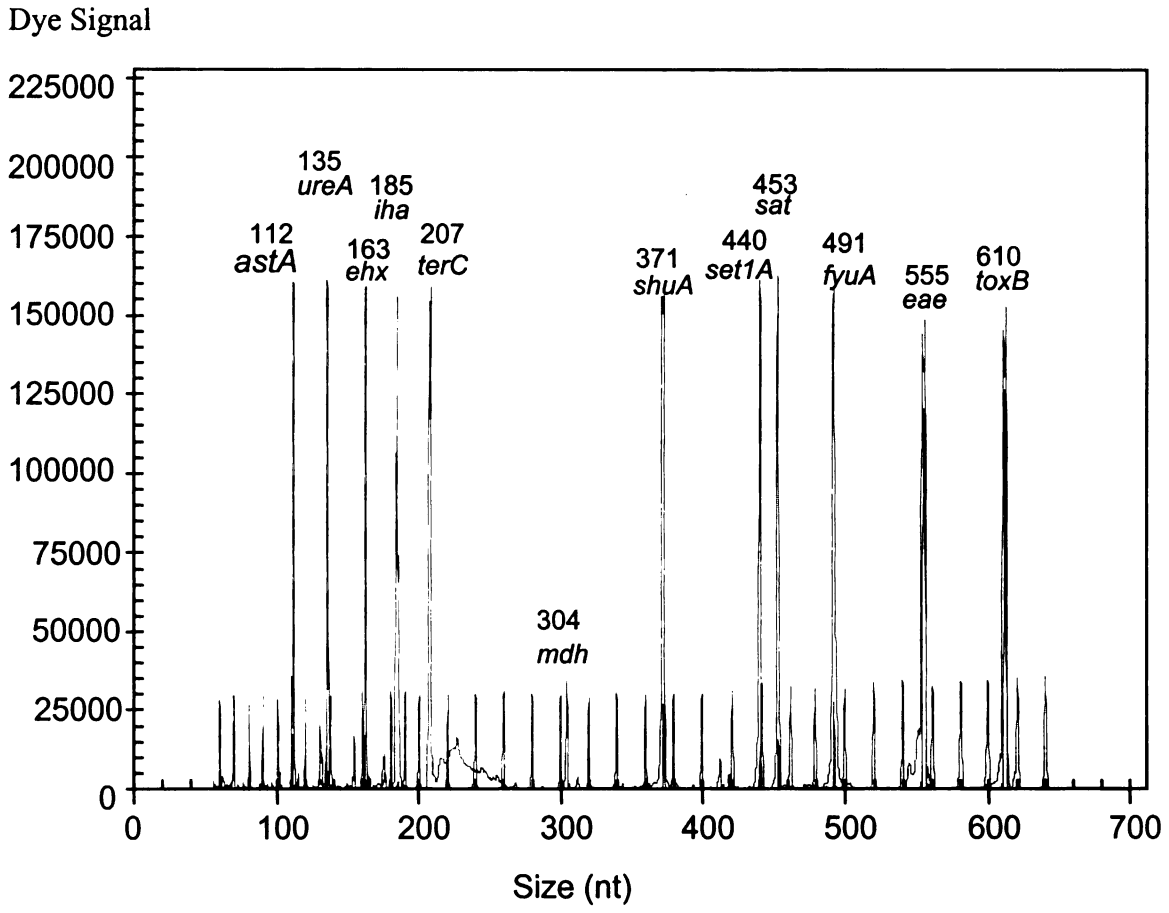


Figure 1. Fluorescent chromatogram output from the multilocus virulence gene profiling (MVGP) procedure. The Beckman CEQ genetic analysis system displays each dye-labeled PCR amplicon as a peak of fluorescence. Peaks are labeled with nucleotide size, and a locus tag that specifies the gene name.

Table 1. Patients in the Seattle pediatric study grouped based on the diagnosis of microorganisms

Diagnosis	Number of Patients	Number of Isolates	Average Illness Duration (in days)	Average Number of Stools	Vomiting	Fever[a]	Bloody Diarrhea	Severity[b]	Severity and Duration[c]
C. jejuni	3	13	5.3	1	1/3	3/3	1/2[a]	1.8	9.5
C. difficile	12	37	5.5	1.2	6/12	8/12	3/10[a]	1.5	8.3
STEC (O157)	5	23	2.5	1	4/5	3/5	3/5	2.0	5.0
STEC (non-O157)	5	24	6.3	1.2	2/5	1/5	4/5	1.4	8.8
Rotavirus	8	35	4.1	1.5	7/8	4/8	0/7[a]	1.4	5.7
Salmonella	4	15	1.8	1	2/4	4/4	3/4	2.3	4.1
Shigella	5	25	2	1	4/5	5/5	3/5	2.4	4.8
Negative (all)	51	219	5.1	1	32/49[a]	27/50[a]	9/50[a]	1.3	6.6
Negative (Group 1)	25	100	4.3	1	17/24	12/23[a]	2/24	1.3	5.6
Negative (Group 2)	22	99	6.2	1	14/21[a]	13/20[a]	5/22	1.5	9.3

[a] Number of responded patients

[b] Calculated as the sum of fractions from vomiting, fever and bloody diarrhea

[c] Index is calculated from the severity index multiplied by the duration in days

MATERIALS AND METHODS

Bacterial Strains and DNA Isolation. *E. coli* isolates were cultured from the stools of children with diarrhea who presented to a pediatric Emergency Department at the Children's Hospital and Regional Medical Center (CHRMC) in Seattle, WA. All patients who presented with diarrhea to the CHRMC emergency department during the period from November 1998 through October 2001 were considered eligible and appropriate for enrollment in the study. This population has also been reported in other studies [61-63]. Parents were given an information form that described the study and also informed them of its voluntary nature. If they agreed to enroll their child, they were instructed to complete a questionnaire in English, Russian, Spanish, Somali, or Vietnamese. Questions asked addressed illness history and patient demographic characteristics. If the patient was unable to provide stool during the visit, a swab specimen from the rectum was obtained, but only if the family consented. The procedures followed were performed with the approval of the CHRMC Institutional Review Board.

A panel of diagnostic tests was done to determine a possible cause of illness on all stools submitted in sufficient quantity. These including inoculating the specimen onto sheep's blood, MacConkey, sorbitol-MacConkey, Hektoen, and *Salmonella-Shigella–, Campylobacter-,* and *Yersinia*-selective (Prepared Media Laboratories) agars and inoculation into Selenite F (BBL; Becton Dickinson) and MacConkey (Binax) broths. *Campylobacter* plates were incubated under microaerophilic conditions at 42°C, while all other incubations were at 35°C. These techniques are standard for the isolation of *Aeromonas* species, *Campylobacter* species, *E. coli* O157:H7, *Pleisiomonas shigelloides,* and *Salmonella, Shigella, Vibrio,* and *Yersinia* species. Also, the MacConkey broth was

tested after overnight incubation by EIA (Meridian Biosciences) for the Shiga toxin antigen and *Clostridium difficile* toxin was tested using a cytotoxicity assay with cultured human diploid fibroblasts. Cytotoxicity was confirmed by neutralization with *Clostridium sordellii* antitoxin. We sought parasites using trichrome stains and formalin/ethyl acetate sedimentation. Fluorescence antibody testing was also used to detect *Giardia* and *Cryptosporidium* species (Techlab). On frozen specimens, we sought rotavirus, adenoviruses, and astrovirus using Rotaclone (Meridian Biosciences), Adenoclone (Meridian Biosciences), and astrovirus (IDEIA Astrovirus EIA kit, DAKO) EIAs, respectively.

All *E. coli* were grown on LB agar overnight at 37 °C from freezer stocks created with single-colony picks of each strain. Strains were inoculated into 10 ml of sterile LB broth for overnight growth at 37 °C, with moderate shaking. DNA isolation was performed using the PureGene (Minneapolis, MN) DNA isolation kit protocol for Gram-negative bacteria. Extracted DNA was quantified by a Nanodrop ND-1000 (Wilmington, DE), then diluted with sterile water to ~100 ng/ml. DNA samples were stored at –20 °C until use.

PCR Primer Design. DNA sequences of MVGP components were obtained from GenBank. Homologous genes were aligned, with PCR primers designed from the conserved regions of each gene (Table 2). All primers were synthesized by Integrated DNA Technologies (Coralville, IA) and were stored at a concentration of 100 mM in ddH_2O. Working concentrations of each primer were 1 mM for MVGP, and 20 mM for MLST.

Table 2. Oligonucleotide primers used in the Seattle pediatric study.

Gene	Name	5' - 3' sequence	Reference
astA	EAST-1P1	GGTCGCGAGTGACGGCTTTGT	[64]
	EAST-1P2	CCATCAACACAGTATATCCGA	[64]
bfpA	bfpA-F11	GTCTGCGTCTGATTCCAATA	This study
	bfpA-R1	TCAGCAGGAGTAATAGC	This study
cdtA	cdtA1-R2	TGCCGCTCTGACAGGTGGACTTA	This study
	cdtA1-F2	GCCTTTAAAAACGGGGTGATACA	This study
chuA	chuA-636F	TGAAACCGCGCCGAATGACGAGT	This study
	chuA-1171R	GGGTTCCGCCAAGCAGGGTAATC	This study
eae	eae-F626	ATTATGGAACGGCAGAGGTTAAT	This study
	eae-R1166	ATCCCCATCGTCACCAGAGG	This study
ehx	MFS1Fb	GTTTATTCTGGAGCAGGCTC	[65]
	MFS1R	CTCCACGTCACCATACATAT	[65]
elt	TW20	GGCGACAGATTATACCGTGC	[66]
	JW11	CGGTCTCTATATTCCCTGTT	[66]
espC	601F	GTTGGGGCTCGGACGACTTAT	This study
	1151R	CCGGCACCCTTGAATGTTAATT	This study
espP	2859F	CGCGCCAAAAGACACCAATGAA	This study
	3321R	CAGGCCAGCCCCCACAGACTT	This study
estA	JW14	ATTTTTMTTTCTGTATTRTCTT	[66]
	JW7	CACCCGGTACARGGCAGGATT	[66]
fyuA	fyuA-924F	GCAGCAGCAGCATTATTCG	[47]
	fyuA-RP	CGCAGTAGGCACGATGTTGTA	[67]
iha	iha-F1	ACGCAGCCGCCAGTGTT	This study
	iha-R1	CCATCAATCAGTATCAGCGTGTAA	This study
invG	invG-481F	TGACCTGGTCGTTAATGCTG	This study
	invG-572R	CGCCACGTAACTCATAAGTCC	This study
irp2	irp2-FP	AAGGATTCGCTGTTACCGGAC	[67]
	irp2-RP	TCGTCGGGCAGCGTTTCTTCT	[67]
mdh	269F	GGTATGGATCGTTCCGACCT	This study
	p10	GGCAGAATGGTAACACCAGAGT	This study
pet	F2227	GTTACGGCCAGCAGTTCCCTTTTC	This study
	R2596	AATTGCCGGTCACTTTCCAGAGC	This study
pic	921F	CGATGCCCCCGTAGACTTTGTTTC	This study
	1333R	TACCGTCTCCCCTTTTCAGTCCTC	This study
saa	saa-1442F	CGTGATGAACAGGCTATTGC	[68]
	saa-1522R	ATGGACATGCCTGTGGCAAC	[68]

Table 2 (cont'd)

sat	1083F	TGGTAGCGGTGGTATTATCTTTGA	This study
	1525R	CGGCTTCTTTCGTTGTATCTGAGT	This study
senA	F486	GGGGGATTTTGTCATTCAGC	This study
	R975	CATTCCTTCCGCAGTTAGTAGTTC	This study
sepA	1672F	GGAGCGCCGGGAGACCT	This study
	2093R	GCCGCATCGAGTTTCAGTTTTTC	This study
setIA	F46	ACGGTTTTCCCAGTCTTTCT	This study
	R481	TATATCCCCCTTTGGTGGTA	This study
shuA	F285	CATCGCGGCGTGCTGGTTCTTG	This study
	R657	CTCGTCATTCGGCGCGGTTTCAC	This study
sigA	30F	GCCCAGGGAAAAATGTATGTAGAT	This study
	434R	AAGACTGTCGCGGGTTTTTA	This study
spaP	SpaP-143F	GGACTTCAGCAAGTGCCATC	This study
	SpaP-275R	ACCACTCATGCCTGTCTCAA	This study
stx1	1A-251F	GGGATAGATCCAGAGGAAGG	This study
	1A-832R	CCGGACACATAGAAGGAAACTC	This study
stx2	2A-506F	CTGGCGTTAATGGAGTTCAG	This study
	2A-848R	CCTGTCGCCAGTTATCTGAC	This study
terC	terC-106F	TATGCACCGTGATGACAAGC	This study
	terC-275R	GGCGAACCAGGAGAAGATTG	This study
toxB	toxB-911F	ATACCTACCTGCTCTGGATTGA	This study
	toxB-1468R	TTCTTACCTGATCTGATGCAGC	[10]
ureA	ureA-109F	TAACTATCCCGAATCCGTGG	This study
	ureA-213R	GGGATCATTTCTGGTATGCCT	This study

PCR of MLST Genes. *E. coli* isolates were characterized using MLST for sequence and phylogenetic characterization of seven conserved housekeeping loci, and MVGP for virulence characterization using a panel of 29 known and putative virulence loci.

Aliquots, (1 μl) of each DNA sample were amplified in a 25-μl reaction mixture using the AmpliTaq Gold system (Applied Biosystems, Foster City, Calif.). Each reaction contained 2.5 μl 10X Gold buffer (150 mM Tris-HCl, pH 8.3, 500 mM KCl), 2.5 μl dNTPs (2 mM each dATP, dCTP, dGTP, and dTTP), 2.0 μl 25 mM $MgCl_2$, 0.5 μl of each primer (1mM), 1.5 units AmpliTaq Gold, and 15.7 μl sterile ddH_2O. Amplification in a Hybaid PCR Express thermocycler (Hybaid Limited, Middlesex, England) utilized an initial denaturing step at 94 °C for 10 min., followed by 35 cycles of 92°C for 1 min., 58 °C for 1 min., and 72 °C for 30 sec. A final step of 72 °C for 5 min. was used for final completion of any partially extended product.

The PCR products were purified using QIAquick PCR Purification Kit (QIAGen Inc., Valencia, CA). Purified PCR amplicons were sequenced using a Beckman CEQ 2000 XL DNA sequencer (Beckman, Fullerton, CA) according to the manufacturer's suggested protocol.

Phylogenetic Analyses of Sequence Data. Sequences were aligned with the ClustalW algorithm using the computer software MegAlign (Lasergene), and allelic sequences were determined. Neighbor-joining trees of the concatenated internal regions of 7 conserved housekeeping genes (*aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, and *uidA*) were constructed using p-distance of nucleotide substitution with the computer software

MEGA version 2.1 [69] and the inferred phylogenies were each tested with 500 bootstrap replications.

Virulence Genes and Protein Functions. Known and putative virulence genes used for MVGP screening are listed in Table 3. A summary of gene functions are described below and grouped by the major pathotype in which the genes are found. The gene *mdh*, encoding malate dehydrogenase, was used as a positive control for all *E. coli*. A brief description of the virulence genes listed in Table 3 is described below.

Two iron-related genes were used to identify enteroaggregative *E. coli*, *fyuA* [67, 70, 71] and *irp2* [67]. These are both markers for the High Pathogenicity Island [67], which is believed to have been horizontally acquired by *E. coli* from *Yersinia*. EAggEC also contain the serine protease autotransporter, *pet* in strain 042. This toxin is associated with mucosal damage, increased mucus release, exfoliation of cells, and development of crypt abscesses [16, 17].

Three markers were used for the pO157 plasmid: hemolysin (encoded by *ehx*) for iron piracy [72], EPEC secreted protein P (encoded by *espP*) [73-76] and an adherence enhancing protein [10] (encoded by *toxB*). In addition, strains were screened for an autoagglutinating adhesin (encoded by *saa*), a marker for a subset of LEE-negative STEC strains capable of causing severe gastrointestinal disease [68].

EHEC also have other genes encoding a variety of possible virulence functions, including toxins (*cdtA*) [77-80], and iron utilization (*chuA*) [81, 82]. EHEC also contain the TAI island (O-island 43). We screened for 3 markers from this island, *iha* [83], *terC* [84], and *ureA* [85, 86].

Table 3. Virulence genes associated with diarrheal disease

Major pathogens[a]	Gene	Protein	Location[b]
Control	*mdh*	Malate dehydrogenase	C
EAggEC	*fyuA*	Yersinabactin receptor FyuA	C
	irp2	Iron repressor HMWP2	C
	pet	Secreted autotransporter toxin	P
EHEC	*cdtA*	Cytolethal distending toxin	C
	chuA	Heme utilization	C
	iha	IrgA homologue adhesin	C
	terC	Tellurite resistance protein C	C
	ureA	Urease	C
	ehx	Enterohemolysin	P (pO157)
	saa	Autoagglutinating adhesin	P (pO157)
	espP	EPEC secreted protein P	P (pO157)
	toxB	Toxin B	P (pO157)
EHEC/EPEC	*eae*	Intimin	C
	spaP	Surface protective antigen P	C
EHEC/STEC	*stx1*	Shiga toxin 1	C
	stx2	Shiga toxin 2	C
EPEC	*astA*	EAST1 heat stable toxin	C, P
	bfpA	Bundlin	P (EAF)
	espC	Secreted protein C	C
ETEC	*elt*	Heat labile toxin (LT-A)	P
	estA	Heat stable toxin (STI)	P
S. dysenteriae	*shuA*	Heme receptor	C
S. flexneri	*pic*	Secreted autotransporter toxin	C
	sat	Secreted autotransporter toxin	C
	senA	Enterotoxin ShET2	P
	sepA	Secreted protein A	P
	set1a	Enterotoxin ShET1	C
	sigA	Secreted autotransporter toxin	C
Shigella sp.	*invG*	Transport	C

[a] EAggEC, entero-aggregative *E. coli*; EHEC, enterohemorrhagic *E. coli*; EPEC, enteropathogenic *E. coli;* ETEC, enterotoxigenic *E. coli*; STEC; Shiga toxin-producing *E. coli*; S., *Shigella*

[b] C, chromosome; P plasmid; pO157, plasmid found in *E. coli* O157:H7 strains; P (EAF), plasmid found in EPEC strains

EHEC and STEC also contain one or both Shiga toxins (encoded by *stx1* and *stx2*). The toxins' ability to stop peptide elongation during translation has been well characterized [87-89].

Three genes used in this study are common to both EHEC and EPEC: *eae* (encoding the adhesin intimin and serving as a marker for the LEE island), and two markers for O-island 115, *spaP* [90], and *invG* [91].

Three additional markers for EPEC used in this study include *astA*, the heat-stable enterotoxin EAST1 [64], *bfpA*, which encodes bundlin and is a marker for the EAF plasmid [92], and the serine protease autotransporter toxin gene *espC* [4, 5].

ETEC were identified by detection of two distinct toxins responsible for the disease symptoms in ETEC infection. ST, the heat-stable toxin (encoded by *estA*) causes the rapid accumulation of intestinal fluid [66], while LT, the heat-labile toxin (encoded by *elt*), is a secretory toxin with a function similar to that of cholera toxin [66, 93, 94].

Our marker for *S. dysenteriae* was the heme receptor gene *shuA* [95, 96], which is closely related to *chuA* from EHEC.

We used six markers for *S. flexneri*, including 4 serine protease autotransporter genes (*pic* [85, 97], *sat* [98, 99], *sepA* [100], and *sigA* [101, 102]), and enterotoxins ShET1 encoded by *set1a* [19, 20, 101] and ShET2 encoded by *senA* [24, 103].

MVGP Procedure Summary. Multi-locus virulence gene profiling involves PCR amplification of extracted DNA with a primer pair that has one regular (unlabeled) and one Beckman fluorescently-labeled primer with the dye attached to the 5' end of the primer. Up to 16 of the labeled amplicons are then pooled together, and then separated on

the CEQ Genetic Analysis System using capillary based sequencing analysis. Visual determination of virulence profiles resulted from analyses with pre-set parameters for each locus for size and dye label.

PCR of MVGP Genes. This technique allows rapid generation of virulence profile data using proprietary technology developed by Beckman Coulter (Fullerton, CA). We analyzed raw data with a collection of pre-set parameters, including data indicating locus tags specifing gene name, amplicon size (± 3 nucleotides), and dye color. A size marker estimates the size of unknown fragments within 1-2 nucleotides. Each profile identifies individual virulence factors as peaks of defined size and dye label, to provide a fingerprint of the virulence content of each strain.

PCR for virulence loci examined by MVGP used the same protocol as the MLST PCR except for the following changes: the 20 µl cocktail contained 2.0 µl of 10X Buffer, 2.0 µl dNTPs, 2.4 µl $MgCl_2$, 1.0 µl each of forward and reverse primer, and 6.4 µl sterile ddH_2O; cycling conditions included: 40 cycles of 92 °C for 30 sec., 50 °C for 30 sec., and 72 °C for 45 sec., with a final extension step at 72 °C for 7 min. 5 µl of DNA template was used per reaction. PCR primers and amplicon sizes for virulence genes are listed in Table 2.

Controls for MVGP. Positive controls consisted of the malate dehydrogenase gene (*mdh*) for all strains and a pooled positive reaction containing a mixture of DNA from all strains serving as positive controls for the other 29 loci (Table 3) in the study. Negative controls contained H_2O and DNA from *E. coli* K12, which was negative for

most genes. In some cases, two sets of primers with the same dye-label were multiplexed to save reagents and space. In that case, the H_2O per reaction was reduced from 6.4 to 4.4 ml. The *mdh* gene was used as the positive control strains for each gene in the MGVP set.

Each sample run contained both a positive control for each individual strain tested (mdh), but also a pooled positive control that contained positive labeled amplicons for each gene tested in that set.

MVGP Data Confirmation. Inconsistencies in the data were clarified by standard PCR using 1.5% agarose gel electrophoresis. Approximately 900 individual PCR reactions were performed to confirm cases where a gene was not uniformly positive in all strains of the same ST in a patient. PCR was used to also check inconsistencies involving results when one, but not both genes on a specific element were detected, such as *irp2* and *fyuA*, markers for the High Pathogenicity Island from *Shigella*. In all cases, every strain with variable results was re-tested to be as thorough as possible.

Images in this dissertation are presented in color.

RESULTS

Genetic Diversity of MLST Loci. All 392 strains from 94 patients in this study were sequenced by MLST for seven loci. The nucleotide diversities of these loci were between 4.2% (*lysP*) to 14.2% (*fadD*). The overall variability across all loci was 9.4% (353 variable sites in 3753 total nucleotides sequenced) (Table 4). The lowest and highest rates of synonymous substitutions, as measured by the Nei-Gojobori method of nucleotide substitution, were also found in *lysP* (3.00 ± 0.90) and *fadD* (9.74 ± 1.62) respectively. However, for non-synonymous substitutions, *lysP* also had the lowest rate (0.00 ± 0.00), but *uidA* had the highest rate (0.41 ± 0.12). The overall rates of synonymous and non-synonymous substitution were 5.69 ± 0.41 and 0.10 ± 0.02, respectively.

Sequence Type Diversity and Clonal Groups. A phylogenetic framework was assembled to display the diversity of different STs. The 392 strains contained a total of 67 unique STs (Figure 2). This sample set was widely distributed across the diversity of all known *E. coli* and included some strains representing specific classes of disease-associated *E. coli* that were highly overrepresented in the sample. For example, clonal group 35 (CG-35) contains strains associated with septicemia and meningitis, including clinical isolate RS218 [104, 105]. Our sample set contained 74 strains of ST-29, which belongs to CG-35. ST-29 is the sequence type of 19% of all strains in this study, and was found four times more frequently than the next most common type, ST-27 (n=18).

Table 4. Sequence variation among alleles of seven MLST genes. Shown here are the number of nucleotides sequenced per gene, the number of variable nucleotides, the number of parsimoniously informative sites (nucleotide variations in more than one strain), the number of alleles identified for that gene in this sample set, and the p-distance.

Locus[a]	No. sites	No. variable sites	No. PI	No. alleles	% p distance ± SE
aspC	513	45	25	33	1.20 ± 0.24
clpX	567	68	53	34	2.36 ± 0.31
fadD	492	70	57	36	2.70 ± 0.41
icdA	567	52	42	33	1.84 ± 0.29
lysP	477	20	14	20	0.84 ± 0.23
mdh	549	36	29	27	1.11 ± 0.22
uidA	588	62	42	39	1.83 ± 0.31
Total	3753	353	262	222	1.70 ± 0.12

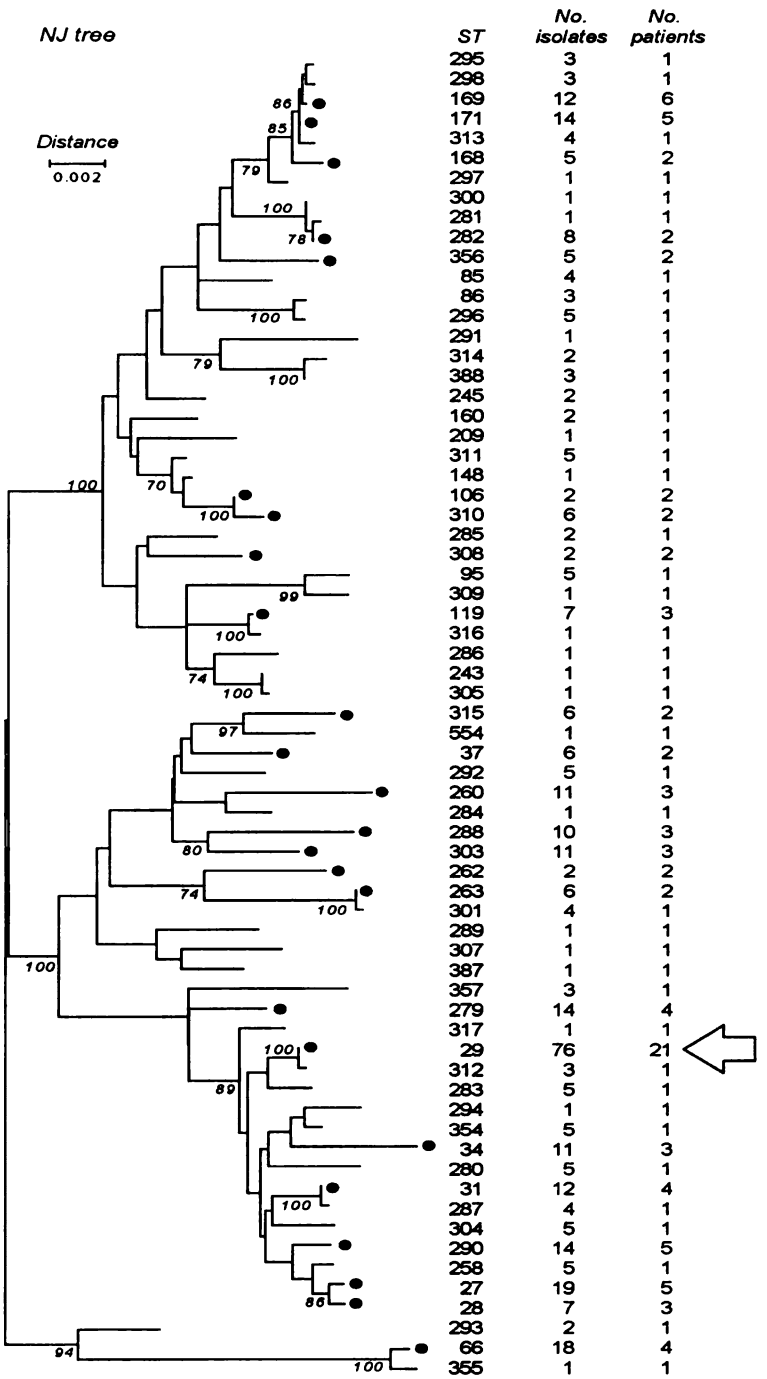


Figure 2. Neighbor-joining tree showing the genetic relatedness of 67 sequence types 392 *E. coli* from 94 patients. Each branch tip represents a unique sequence type (ST) based on multilocus sequence typing of 7 genes. Distance is measured by p distance with bootstrap support greater than 70% given by italics. The number of isolates and number of patients in which each ST was isolated are given in the right hand columns. The circles denote STs that were recovered from more then 1 patient ($n = 24$). The most frequently detected ST (ST-29) is indicated by the gray arrow.

The second and third most common clonal groups included CG-23 (ECOR A) with 34 isolates, and CG-38 (UTI-1 associated strains from ST-27 and ST-28), with 25 isolates. We found more than one representative for 24 of the 67 STs (35.8%).

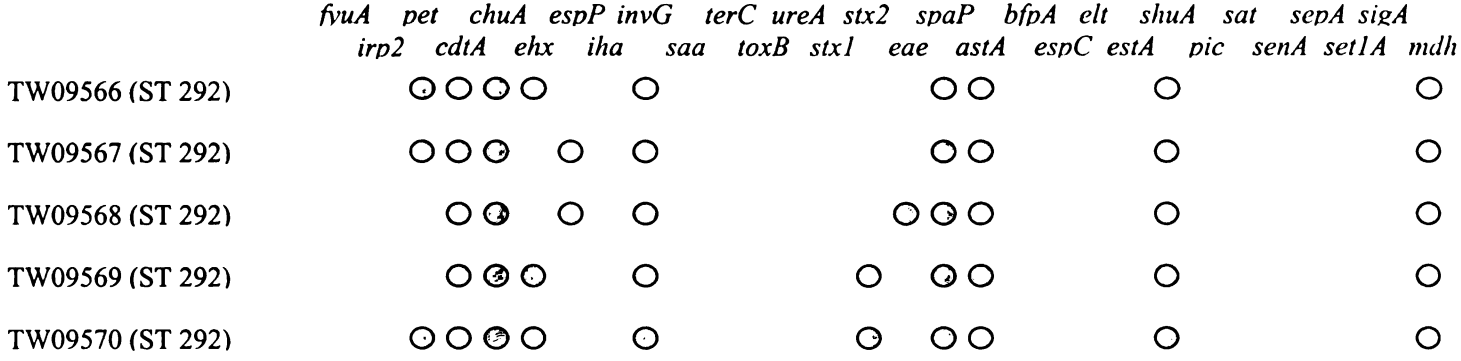
MVGP Analysis. A total of 392 isolates from 94 patients were examined via MGVP, with an average of 8.1 unique virulence genes detected per patient. The overall virulence type (VT) diversity is 71%, indicating that two strains chosen at random have a 71% chance of displaying different profiles in virulence genes. Ten patients had strains with zero diversity (all isolates with the same VT), whereas twenty patients had strains with 100% diversity (all isolates with different VTs). Genes encoding iron acquisition or utilization ability were the most frequently detected in this sample set (Table 5), including *shuA* (72.5% positive), *irp2* (72.0%), *fyuA* (68.4%), and *chuA* (65.5%). In contrast, other genes restricted to specific pathotypes were found less frequently, including *stx1* and *stx2* (6.0% detection each).

Comparison of Virulence Gene Profiles and Phylogenetic Relationships. Virulence profiles were compared with phylogenetic frameworks generated by MLST. Three examples of virulence diversity within individual patients include the following: Patient 1 had 5 individual isolates with the same ST (ST-29) and VT. Patients 34 and 35 (Figure 3) had isolates that showed variability in their profiles. Patient 34 contained 5 strains with the same ST (ST-292), but had 5 slightly different VTs. Patient 35 had strains with 4 different STs (STs 119, 272, 293, and 294) and 5 different VTs. Collectively, patient 35 contained strains with 17 of 29 virulence genes examined in this study (59%). This patient shows the

Table 5. Summary of virulence gene detection frequency in sequence types and in patients from the Seattle pediatric population study.

Gene	% STs (67)	% Patients (94)	Gene	% STs (67)	% Patients (94)
shuA	70.1	79.8	*set1a*	17.9	12.8
irp2	70.1	81.9	*pet*	16.4	23.4
fyuA	67.2	85.1	*sepA*	16.4	9.6
spaP	64.2	50.0	*espP*	11.9	11.7
chuA	62.7	74.5	*stx1*	11.9	12.8
terC	62.7	74.5	*toxB*	10.4	11.7
invG	59.7	47.9	*pic*	10.4	8.5
eae	44.8	40.4	*stx2*	7.5	10.6
iha	37.3	40.4	*espC*	7.5	6.4
ureA	35.8	25.5	*senA*	6.0	6.4
saa	32.8	30.9	*bfpA*	3.0	2.1
sat	28.4	29.8	*cdt*	3.0	2.1
ehx	22.4	17.0	*elt*	3.0	1.1
astA	20.9	11.7	*estA*	1.5	1.1
sigA	20.9	10.6			

Patient 34 – Diagnosis: *Salmonella* subgenus D1



Patient 35 – Diagnosis: *E. coli* O103:H2

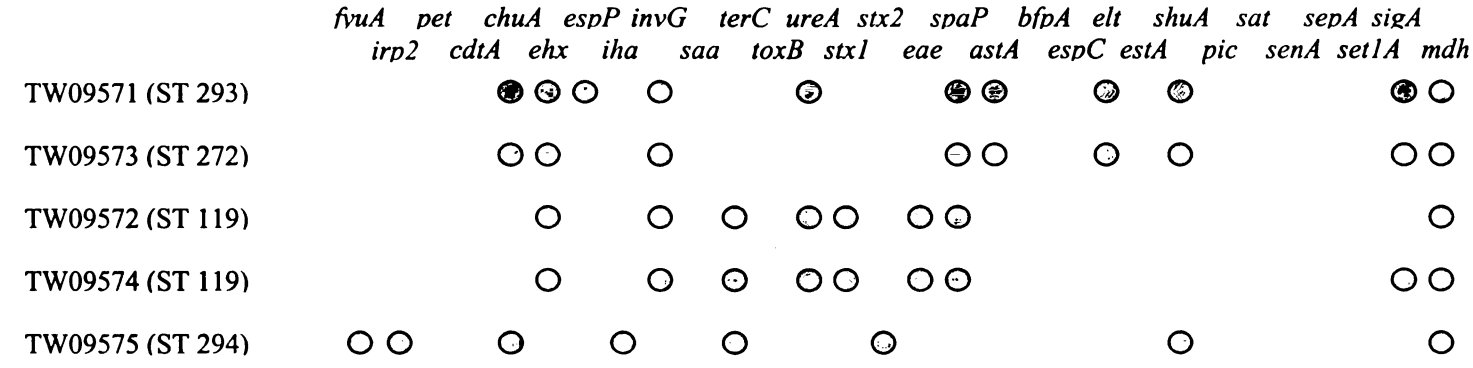


Figure 3. Virulence type diversity in two representative pediatric patients. The colored symbols represent a positive MVGP result for that particular gene. Negative samples are not shown. Patient diagnosis is also indicated.

greatest diversity of STs and VTs in this study, and may be an example of how different strains can contribute to unique disease by complementation of virulence ability.

A number of genes related to type III secretion (*eae, invG,* and *spaP*) and iron utilization (*chuA, shuA, fyuA,* and *irp2*) are seen in 30 or more different STs, and 38 or more patients (Table 5). Detection of autotransporter genes was highly variable.

Another analysis technique called non-metric multidimensional scaling (NMDS) analysis was used to place most patients into two clustered groups based on the virulence content of results from each patient, not per isolate. This allowed for distinctions based on the detection of virulence genes in any strain found in a patient, while keeping the individual patient as the unit of measurement, and not individual isolates. Since we used 29 virulence loci, NMDS combines these 29 dimensions into a two-dimensional image.

NMDS showed two distinct groups (Figure 4). Groups 1 and 2 contain 42 and 43 total patients respectively. All 10 STEC patients were within group 1, whereas 9 of 11 patients with *C. difficile* were in group 2. Group 1 also had 25 patients with no diagnosis and an average of 9.0 virulence genes per patient, while group 2 had 23 patients with no diagnosis and an average of 6.6 virulence genes.

Statistical analysis. Our first hypothesis was that a longer illness duration (0-3 days vs. 4-6 days vs. 7 or more days) was significantly correlated with more virulence genes found per patient. We accepted this hypothesis based on a chi-square test (Σ=7.5, p-value =0.02). The second hypothesis was that we would find a significant correlation between the number of virulence genes (0-5) vs. (6 or more) and diagnosis group. This was based on the idea that the E. coli from the no diagnosis group were not likely the cause of the illness seen in the patient and therefore would tend to have less virulence

genes. We rejected this hypothesis based on a chi-square test (Σ=2.07, p-value =0.150), so there is no significifcant difference between the number of virulence genes in the diagnosis and the no diagnosis groups.

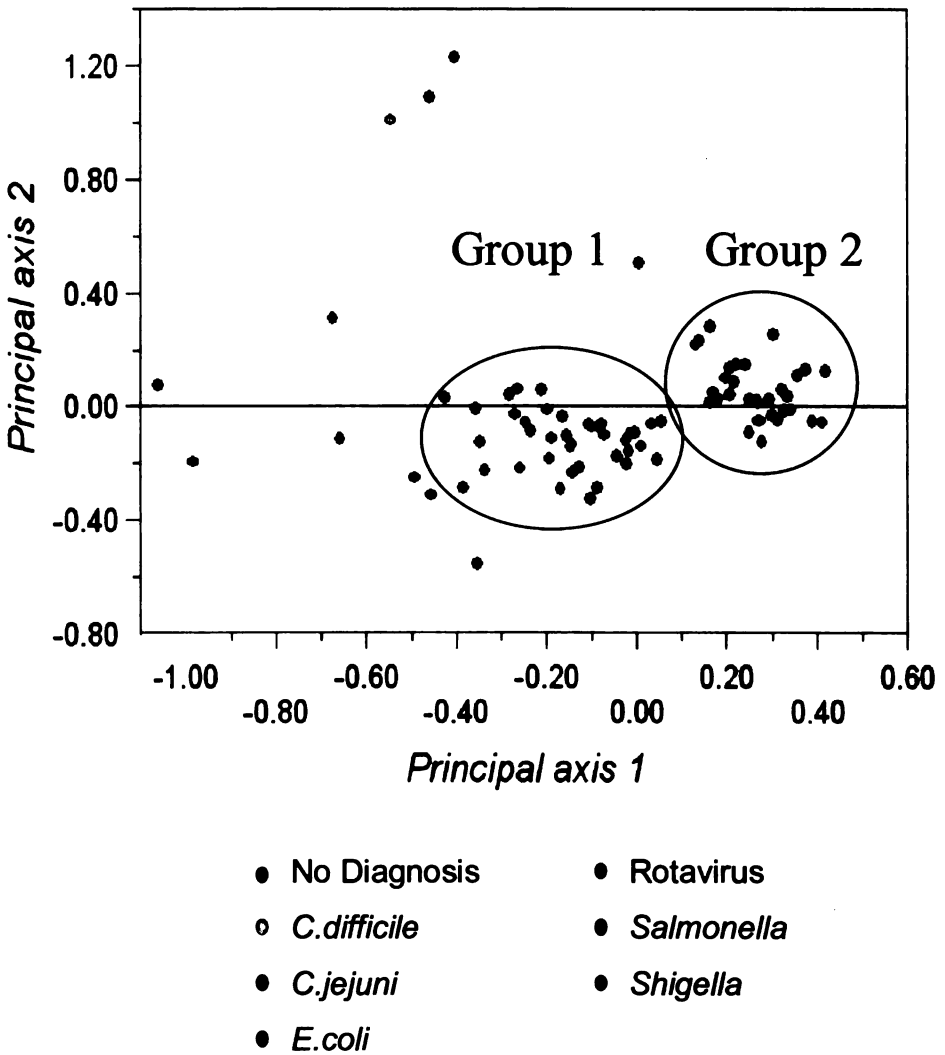


Figure 4. Non-metric dimensional scaling (NMDS) analysis of the MVGP results of 94 patients based on the presence or absence of 29 known or putative virulence factors. Two clusters are indicated by the circles, named group 1 and group 2.

DISCUSSION

Diversity of Virulence Gene Profiles. Virulence profiles in this study are highly variable. Individual patients often contain multiple STs with minor to major variation in virulence diversity. At least 70 STs were found from 392 isolates sequenced, including 45 newly discovered STs. Some individual patients contained strains with different STs and VTs. Even in patients with strains that all have the same ST, variability in virulence exists, such as was seem in the analysis of strains from patient 34 (Figure 3). Other patients had diversity in both STs and VT. The greatest example of ST and VT diversity was in patient 35, with four unique sequence types (STs 119, 272, 293 and 294) and five unique virulence profiles. This patient harbored *E. coli* with 17 of 29 (59%) virulence-related genes used in our study.

Genes encoding virulence factors responsible for iron acquisition and utilization (*shuA, fyuA, chuA,* and *irp2*) were the most prevalent in the samples examined. Although iron is important for the survival of *E. coli* [106, 107], the role of iron-related genes in pathogenesis is not entirely clear. Such widespread distribution in this sample set may indicate a role in survival, as a means for virulence enhancement, rather than overt pathogenic ability.

Some of the variation seen among isolates with the same ST may be a result of natural rates of gene flux. Because many of the virulence genes in this study are found on mobile elements, it is possible the plasmid or phage containing certain genes could be lost. There is also a possibility of mutations at the primer site affecting proper amplification and the affect of false-negative or false-positive results. Extensive rechecks

by standard PCR and gel electrophoresis helped clarify initially inconsistent results in the data set.

Frequency of ST-29 Isolates in this Sample Set. The detection of a high frequency of ST-29 isolates in this study was surprising. These may represent a group of strains with enhanced colonization ability. ST-29s appear to be largely lacking the genes encoding many virulence factors typically associated with disease. They do, however contain, an even higher frequency of iron-related genes than the non-ST-29 strains in this study.

The detection of such a high number of ST-29 isolates was unexpected. This sequence type is the same as the meningitis-associated isolate RS218. Nearly all (68/70) of the ST-29's in this set also contained the same capsule type (K1) as RS218 [105]. This isolate may represent a widespread meningitis associated clone that has been under-investigated. The reasons for the high frequency of detection may indicate this strain has an ability to cause disease that is yet uncharacterized, or it simply might colonize and survive better in the gut. Why a strain with potential extra-intestinal disease-causing ability is so frequent in the stool of children is also unclear. Since neonatal meningitis caused by strains like RS218 is present only in very young children [108], perhaps ST-29's can cause disease, but only when a child has an underdeveloped immune system. Perhaps, in older children, ST-29 isolates are not able to cause disease, but still persist in the gut through a mechanism of attachment or immune avoidance.

A total of 10 of 51 patients without a diagnosis (19.6%) had ST-29 isolated from their stool. However, ST-29's were isolated from 5 of 12 (41.7%) of patients diagnosed

with *C. difficile* and 3 of 5 (60%) of patients diagnosed with *Shigella flexneri*. No ST-29 isolates were found in patients with a diagnosis of *C. jejuni*, rotavirus, *Salmonella paratyphi* B, *Salmonella* subgenus 1 (*Salmonella* D) or *E. coli* serotypes O103:H2, O111:H8, O111:NM, or O118:H16. The presence of *C. jejuni*, rotavirus, *Salmonella*, etc, may change the local environment of the gut in ways that inhibit the survival of ST-29 isolates.

Virulence profiling of *E. coli* has been used previously to detect a large number of mostly attachment-related loci in strains from the *E. coli* Reference collection (ECOR) [109]. We hypothesized ST-29 isolates would contain a gene or genes found in nearly every strain that is associated with attachment in extra-intestinal variants of *E. coli*. We further investigated ST-29 strains by examining a select number of attachment-related loci used by Johnson, et al. [109], which were not included in our original MVGP gene set. This work is detailed in a subsequent study.

It is also unclear if ST-29 isolates have a differential geographic distribution. The Seattle, Washington state area has a high incidence of illness due to *E. coli* O157:H7 [110-114], but also has the research and laboratory facilities for efficient detection of this serotype. The closely related strains of ST-29 may in fact exist in large numbers worldwide, but the low detection frequency may be due to a lack of appropriate surveillance.

ACKNOWLEDGEMENTS

I would like to thank Lindsey Ouellette for her tireless work sequence typing the strains used in this set, as well as Drs. Eileen Klein and Phillip Tarr for collecting and sending us the strains examined here. Thanks are also due to Albert Lee for his help with virulence gene data confirmation and to Drs. David Lacher and Cheryl Tarr for technical assistance.

This research was supported by funds from the Centers for Disease Control and Prevention (Cooperative Agreement CCU015040) and NIAID, NIH, DHHS, under NIH research contract # N01-AI-30058 (TSW).

CHAPTER 3

CHARACTERIZATION OF A COMMON MENINGITIS-ASSOCIATED CLONE (ST-29) FOUND IN PEDIATRIC PATIENTS IN SEATTLE, WA

SUMMARY

The intrinsic virulence of individual *Escherichia coli* strains is determined in part by the ability to encode toxins, enhanced nutrient utilization, attachment, or other abilities important for disease. However the distribution and relationship of virulence factors to disease-causing ability is not entirely understood. As seen in Chapter 2, we reported the sequence types and virulence profiles of 392 *E. coli* isolates from pediatric patients with diarrheal disease of unknown cause. In this strain set, a high prevalence (19%) of isolates of a single sequence type (ST-29) were detected by MLST analysis. ST-29 belongs to a larger clonal group that includes STs of extraintestinal pathogenic *E. coli* (ExPEC) including the meningitis-associated strain RS218. We hypothesized that ST-29 strains have a fitness advantage related to adherence to the intestine. To address this hypothesis we assayed for the presence of additional genes associated with ExPEC adherence (*aafA, aggR, bmaE, fimH, iutA, focG, ibeA, nfaA, papAH, sfaA, sfaS*) in 70 ST-29 strains using PCR-based assays. MVGP screening showed iron acquisition and iron utilization genes were the most frequently detected, including *shuA* (97%), *irp2* (97%), *fyuA* (90%), and *chuA* (90%). Toxin genes like *stx1* (10%), and *stx2* (3%) were found less frequently. The prevalence of attachment factors was mixed, with some detected frequently, *fimH* (97%) and *nfaA* (90%), whereas others were low or absent (*aggR, focG,* and *aafA* were not detected in any ST-29 strain). An RFLP-based capsule typing method showed 68 of 70 (97%) ST-29 isolates to be K1, the same type as RS218. ST-29 warrants further study because it appears to mark a common *E. coli* intestinal clone, related to ExPEC, which lacks virulence genes associated with diarrheagenic *E. coli*.

INTRODUCTION

Escherichia coli is a gram negative bacterial resident of the intestinal flora within a healthy human host. However, some strains of *E. coli* have acquired the ability to cause a variety of illness, ranging from mild diarrhea to a severe bloody diarrhea that can be life threatening. Certain strains are also associated with disease outside the intestine, including meningitis, septicemia, urinary tract infections, and kidney damage [1, 115]. The different pathogenic abilities are seen because some strains have acquired genes or pathogenicity islands that allow for the production of toxins and other factors necessary for disease onset [1, 2]. *E. coli* is capable of frequent acquisitions of new virulence ability by horizontal transfer of genetic material. One way to group *E. coli* is based on the type of disease they cause. Strains of the same disease class are not necessarily evolutionarily related, but may have acquired the same virulence factors encoded in their genome to produce similar disease phenotypes.

We are interested in determining the prevalence of the genes encoding many common virulence factors in specific populations. In Chapter 2, we examined 392 *E. coli* isolates from children hospitalized in Seattle, WA with unexplained cases of diarrheal disease. These strains were characterized by two methods: multi-locus sequence typing (MLST), and multi-locus virulence gene profiling (MVGP). DNA sequencing revealed a predominant sequence type (ST-29) found in 18.9% of strains (n=74).

MLST detected 67 unique sequence types in this population that were distributed throughout the diversity of *E. coli*. However, ST-29 was discovered at an exceptionally high frequency compared to other sequence types. ST-29 isolates do not contain many of the classical virulence factors associated with enteric disease. We hypothesized ST-29

isolates may be more frequently detected because they are better able to attach to the intestinal tract. The majority of ST-29 isolates, 60 (81%) were examined for the presence of 11 attachment associated genes by PCR. In addition, 70 (95%) of the ST-29 strains were capsular typed using an RFLP-based method (Figure 5) to determine the specific allele of their polysaccharide capsule. The banding pattern of the ~10,000-12,000 base pair kpsC-kpsM amplified region after digestion with the Sau96I restriction endonuclease is seen in Figure 6. Determination of allele type is easily made based on the distinct banding pattern for each allele standard.

ST-29 is important for further study because it is the same sequence type as RS218, a strain implicated in extra-intestinal infections, including septicemia and meningitis [116, 117]. We also propose that ST-29 marks a clone that has a role in intestinal pathogenesis that has yet to be determined.

We wish to test the following hypotheses regarding ST-29 isolates examined here.

1. Because the clinical isolate RS218 is a ST-29 strain and has a K1 capsule type, we predict all or nearly all pediatric ST-29 isolates from Seattle will also have the K1 capsule type. A capsule typing method using restriction fragment length polymorphism will be used to determine capsule profiles.

2. We also predict the prevalence of ST-29 isolates is related to an enhanced ability to adhere to the intestine. Therefore we predict we will find some attachment-related genes common in extra-intestinal *E. coli* to be in very high frequency (near 100%) in the ST-29 strains. Individual pairs of PCR primers will screen a panel of genes encoding proteins related to attachment.

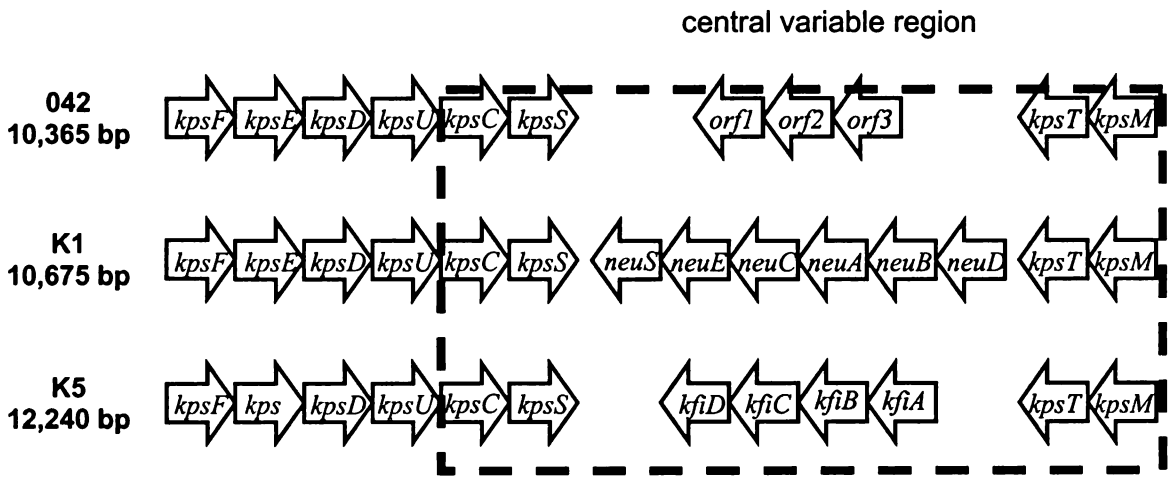


Figure 5. Organization of the kps operon in *E. coli*. Primers were designed in the conserved flanking regions (*kpsC* and *kpsM* genes) and amplified a central variable region that differs greatly between strains of *E. coli*. The area amplified is indicated by the boxed region.

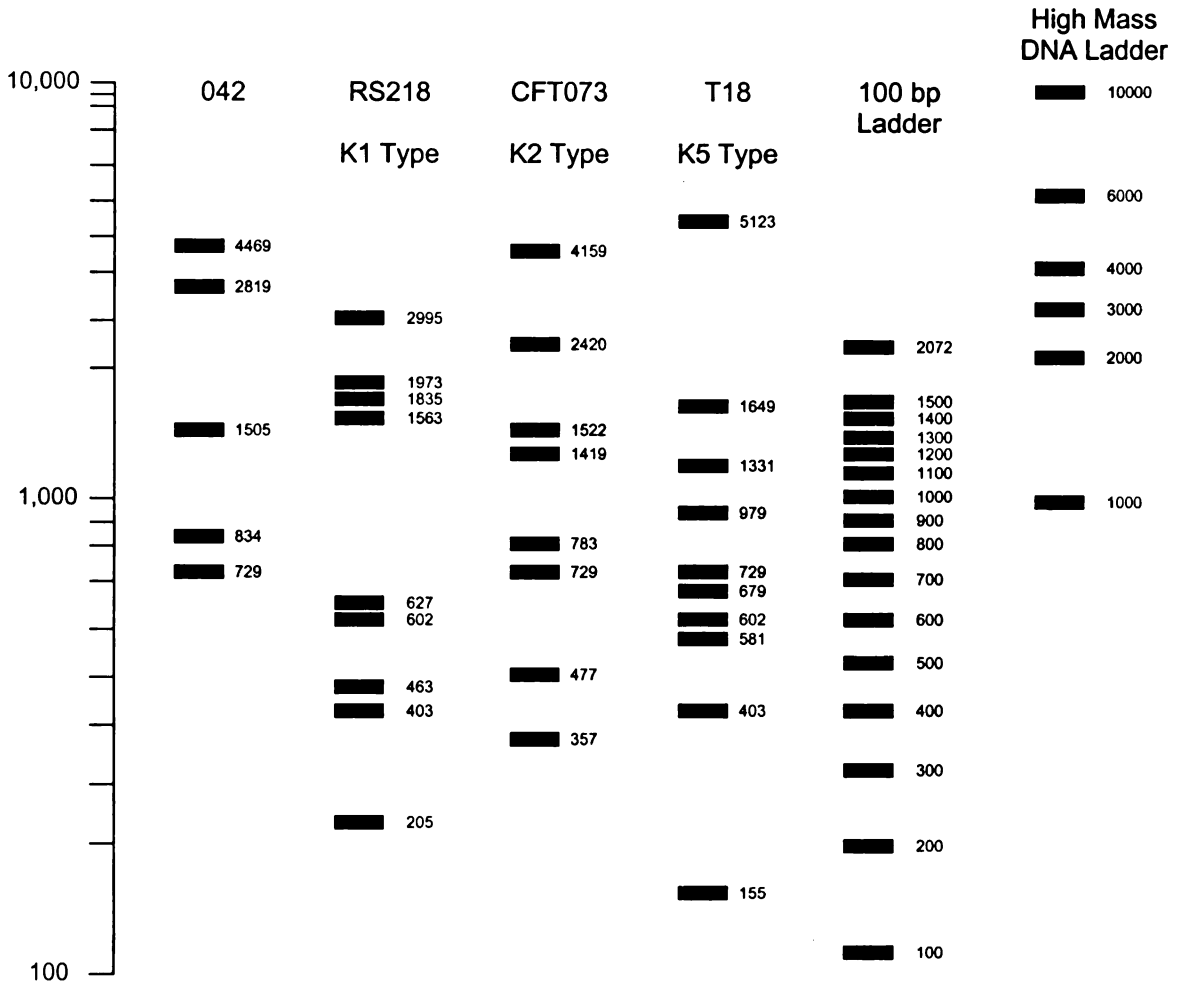


Figure 6. Representative banding pattern from a Sau96I digest of the amplified *kpsCM* region in *E. coli*. Shown here are the specific types used as standards when comparing the ST-29 isolates used in this study. 97% of the 70 ST-29s examined by RFLP were K1, the same as the RS218 standard pattern.

MATERIALS & METHODS

MVGP PCR primer design: DNA sequences of MVGP components were obtained from GenBank. PCR primers were designed from the conserved regions of each gene sequence available. All primers (Table 2) were synthesized by Integrated DNA Technologies (Coralville, IA.) and were stored at a concentration of 100 mM in ddH2O. Working concentrations of each primer were 1 mM for MVGP, and 20 mM for MLST.

Strain growth conditions and DNA extractions: All *E. coli* strains were grown on LB agar plates overnight at 37 °C from freezer stocks created with single-colony picks of each strain. Strains were inoculated into 10 ml of sterile LB broth for overnight growth at 37 °C, with moderate shaking. DNA isolation was performed according to the PureGene (Minneapolis, MN) DNA isolation kit protocol for gram-negative bacteria. After extraction, DNA was quantified by a Nanodrop ND-1000 (Wilmington, DE), and then diluted with water to ~100 ng/ml.

MVGP PCR: PCR of virulence-related loci were performed as described in Chapter 2.

Controls: Positive controls consisted of the malate dehydrogenase gene (*mdh*) for all strains, and a pooled positive reaction containing a mixture of DNA from all strains serving as positive controls for the other 29 genes in the study. Negative controls contained H_2O and DNA from K12, which was negative for most genes. In some cases, two sets of primers with the same dye-label were multiplexed to save reagents and space. In that case, the H_2O per reaction was reduced from 6.4 µl to 4.4 µl.

MLST PCR: MLST PCR, purification, and sequencing were performed as described in Chapter 2.

Phylogenetic analyses: Phylogenetic analyses of sequencing data was performed as described in Chapter 2.

PCR for attachment loci: 5 µl of DNA template was used with 2.5 µl of 10X Buffer, 2.5 µl dNTPs, 2.0 µl $MgCl_2$, 1.0 µl each of forward and reverse primer, 1.5 units AmpliTaq Gold, and 15.7 µl sterile ddH_2O. Cycling conditions included: 10 minute initial soak at 94 °C, 35 cycles of 92 °C for 1 min., 55-60 °C for 30 sec., and 72 °C for 30 sec., with a final extension step at 72 °C for 5 min. Specific annealing temperatures varied, depending on the gene (Table 6).

Capsular typing - PCR: Primers were used to amplify a region (~11,000 bp) in the kps operon (Figure 5). PCR was performed using Takara LA Taq polymerase (1.5U), Takara 10x buffer (2.5 µl), Takara 2.5 mM dNTPs (4.0 µl), 1.5 µl of each primer (10 mM), and 14.2 µl sterile ddH_2O. Primer sequences are: Forward: kpsC-F558 5' – AGCCGAAATTTGGGTGAAGGTG – 3' and Reverse: kpsM-R120 5' – YGCGCATTTGCTGATACTGTTGG – 3'. Amplicons (5 µl) were visualized on 1.5% agarose gels with 6 ml ethidium bromide stain to confirm successful amplification prior to digestion. Controls included strains RS218 (K1 capsule standard), CFT073 (K2 standard), T18 (K5 standard), and 042 (042 standard), along with a negative control of water.

Capsular typing – digest: Amplicons were digested using restriction enzyme Sau96I. Each 30 µl reaction contained 3.0 µl of 10x buffer, 1.0 U of enzyme (0.2 µl), 16.8 µl sterile ddH_2O, and 10 µl template. Reactions were digested overnight at 37 °C then separated on a 1.5% agarose gel stained with 12 µl ethidium bromide. Ladders included 100 bp standard (2 µl), and a high DNA mass ladder (4 µl). Gels were loaded with 15 µl

of digested PCR product, and run at 80 volts of constant current for about 2.5 hours to allow for sufficient separation. An example of representative banding patterns for four different capsule types is seen in Figure 6.

Table 6. Gene and protein data for attachment-related loci used in PCR assays of ST-29 isolates. Gene names, primer names, primer sequences, protein functions, and percent positive detection are shown above, as well as the strain used as a positive control.

Gene	Primer	Amp Size	Ta (°C)	Primer Sequence (5'->3')	Protein Function	% Pos Detection	Control	Reference
aafA	805F	300	60	TCGCGGGAGGACTTTCACC	aggregative adherence fimbriae (AAF/II subunit)	0	042	This Study
aafA	1104R			CGCCGCTCGTTATACCATCTG			042	This Study
aggR	557F	474	55	ACACTGTTGCCCGCATTATTTC	transcriptional activator	0	042	This Study
aggR	1030R			CTTGCCCCCTTATTTCCAGAGA			042	This Study
bmaE	79F	507	53	ATGGCGCTAACTTGCCATGCTG	blood group M-specific adhesin	55	RS218	[109]
bmaE	583R			AGGGGGACATATAGCCCCCTTC			RS218	[109]
fimH	1814F	508	55	TGCAGAACGGATAAGCCGTGG	D-mannose-specific adhestin, type 1 fimbrae	93	CFT073	[109]
fimH	2278R			GCAGTCACCTGCCCTCCGGTA			CFT073	[109]
focG	635F	360	54	CAGCACAGGCAGTGGATACGA	pilus tip molecule, F1C fimbrae	0	042	[109]
focG	993R			GAATGTCGCCTGCCCATTGCT			042	[109]
ibeA	261F	170	57	AGGCAGGTGTGCGCCGCGTAC	invasion of brain epithelieum	20	RS218	[109]
ibeA	439R			TGGTGCTCCGGCAAACCATGC			RS218	[109]
iutA	851F	300	55	GGCTGGACATCATGGGAACTGG	ferric aerobactin receptor (iron uptake/transport)	32	CFT073	[109]
iutA	1152R			CGTCGGGAACGGGTAGAATCG			CFT073	[109]
nfaA	84F	559	50	GCTTACTGATTCTGGGATGGA	non-fimbrial adhesin I	90	CFT073	[109]
nfaA	640R			CGGTGGCCGAGTCATATGCCA			CFT073	[109]
papAH	1796F	720	55	ATGGCAGTGGTGTCTTTTGGTG	PapA - major subunit of P fimbrae	42	RS218	[109]
papAH	2516R			CGTCCCACCATACGTGCTCTTC			RS218	[109]
sfaA	sfa F	410	53	CTCCGGAGAACTGGGTGCATCTTAC	major subunit of S fimbriae	20	RS218	[109]
sfaA	sfa R			CGGAGGAGTAATTACAAACCTGGCA			RS218	[109]
sfaS	675F	240	55	GTGGATACGACGATTACTGTG	tip of S fimbrae	20	RS218	[109]
sfaS	916R			CCGCCAGCATTCCCTGTATTC			RS218	[109]

RESULTS

The ST-29 isolates used in this study were from 21 unrelated, unassociated pediatric patients in the Seattle area. This represented 18.9% of all strains previously collected for the larger paired sequence and virulence typing experiment described in Chapter 2. ST-29s contained very few or no genes typical of most diarrheagenic *E. coli* (*stx1/2, eae, toxB, bfpA*), although these strains do have a higher frequency of iron acquisition and utilization genes than non-ST-29 strains (Figure 7)

The only genes with a higher frequency of detection in ST-29s isolates were: *shuA* (97% vs. 67% in non-ST-29's), *irp2* (97% vs. 66%), *fyuA* (91% vs. 63%), *chuA* (91% vs. 60%), *terC* (65% vs. 42%), *pet* (31% vs. 12%), *saa* (23% vs. 15%), and *stx1* (8% vs. 5%).

Strain 991830484B was the only ST-29 isolate with the K2 capsule profile and was negative for the most frequently detected genes, *shuA* and *irp2*. It was also negative for *fyuA*, *chuA*, and *terC*, and was the only strain positive for *invG* and *spaP* (the only positive results of these two genes in all of the ST-29 isolates). This isolate was taken from a child without a definitive diagnosis for the bloody diarrhea symptoms that lead to hospitalization. All tests for *Campylobacter jejuni*, *Clostridium difficile*, *Shigella*, *Salmonella*, rotavirus and both O157 and non-O157 STEC as described in Chapter 2 were negative. The child had no fever or vomiting prior to admission. Strain 991830484B was the only ST-29 isolate from that patient; the other four isolates were ST-303.

Another strain, 2320612D, taken from a patient exhibiting both vomiting and fever prior to admission, was also negative for the panel of tests described previously.

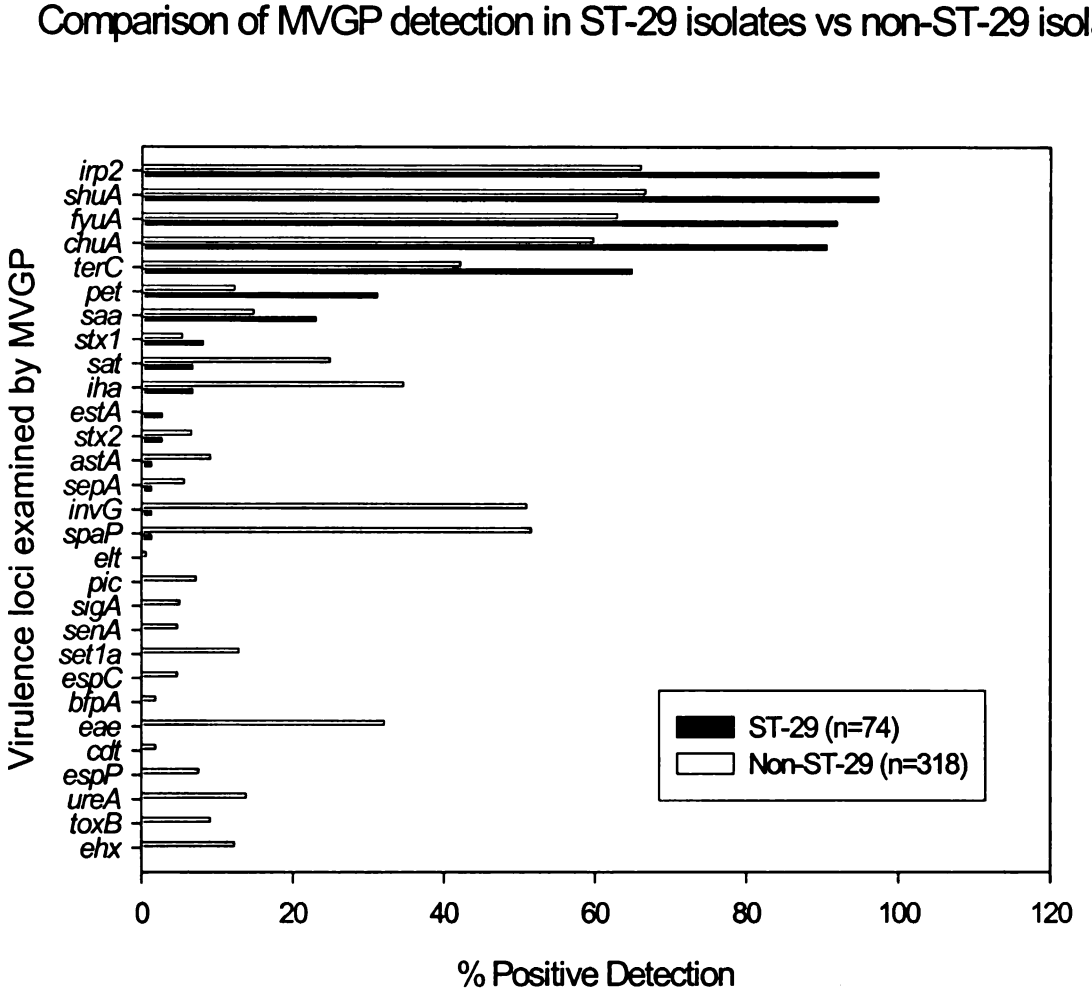


Figure 7. Frequency distribution of each gene from MVGP analysis of 392 isolates from 94 pediatric patients. Bars indicate the percent of isolates in which each gene was found. The graph compares the frequency of virulence gene detection by MVGP in ST-29 isolates (black) vs. all non-ST-29 isolates (grey).

This strain was positive for both *stx1* and *stx2*, but data on bloody diarrhea was unavailable because the stool was not examined by the parent.

Attachment Genes: Many genes were detected at low frequencies (0-20%), but some were found in nearly all ST-29 strains examined (Table 6). Aside from the attachment-related loci used in the MVGP screening, these additional genes included an aggregative adherence fimbrae (*aafA*), a transcriptional activator (*aggR*), a blood group M-specific adhesin (*bmaE*), a gene encoding an invasion of brain epithelium (*ibeA*), the major subunit of S fimbrae (*sfaA*), and *sfaS*, encoding the tip of the S fimbrae, which may play a role in human brain microvascular endothelial cells [109]. Of the 60 ST-29 strains we examined, only *bmaE* showed reasonably frequent detection by PCR (55%), whereas all the others were detected in 20% or less of all isolates tested. Other attachment-related loci may be important for ST-29 isolates, aside from those examined here; including the possibility of loci yet to be characterized. However, the prevalence may also be unrelated to attachment-aided survival.

Aside from n*faA* with 90% detection, there does not appear to be any gene detected frequently that is unique to extra-intestinal pathogenic *E. coli* (ExPEC). The non-fimbrial adhesin encoded by the *nfa* operon was found in at least 3 strains associated with extra-intestinal infection, CFT073 [35], 536 [118], and RS218, as well as EAggEC strain 042. We predicted there would be one or more genes very frequently detected (near 100%) in the ST-29 sample set. Therefore, based on the 93% detection of *fimH*, and the 90% detection of *nfaA*, we accept this hypothesis.

Capsular Typing: 69 of 70 ST-29 strains had successful amplification of the kps region (Figure 5). Of these, all had K1 capsule according to RFLP digestion, except strain

991830484B (TW09832), which had the K2 capsule profile. The high frequency of K1 capsule type in these ST-29s is consistent with the capsule type of strain RS218 [116], which is also a K1/ST-29 isolate (Figure 6). K2 is the capsule type is seen in the urinary tract pathogenic strain CFT073 [35, 119-121]

We hypothesized all or nearly all of the ST-29 isolates will have the same capsule type as RS218. 68 of 70 ST-29 strains did share the K1 capsule type with RS218, so we fail to reject this hypothesis.

DISCUSSION

In this diverse set of *E. coli* isolated from children, ST-29s were detected at a high frequency – 19%. We hypothesized that this ST marks a widespread clone with an enhanced ability to colonize the intestine. This hypothesis may explain why these strains are found in such high numbers in patients with unattributed bloody diarrhea. It is also possible ST-29 isolates may be more frequent in certain populations than previously realized, but lack of surveillance may be the reason for this infrequent identification.

The detection of *nfaA* in 90% of the strains warrants further investigation into the role of this adhesin, especially considering it is found in ExPEC. The other attachment-related loci we examined that were found at low frequencies in ST-29 isolates does not necessarily mean that these strains do not use increased attachment as a means for survival. It is likely that additional loci that we did not examine, or loci that has yet to be characterized, may be critical for attachment in these strains. Despite all of these strains sharing the same sequence type, they still showed variability in both the virulence profiles and the attachment-related genes. This is especially true with *bmaE* (55%), *papAH* (42%), and *iutA* (32%), which seem to have significant variability in their presence or absence within this set. It is unclear why this is the case.

We also performed MLST on *E. coli* isolates taken from patients with Crohn's disease and colon cancer in England [122] and found those to also be ST-29. These strains may play a role in disease by using a mechanism of pathogenesis or mode of invasion that has yet to be characterized. It should be reiterated that while these ST-29 isolates contain genetic and virulence profile similarities to ExPEC, they were isolated from stool samples and not from gut contents in close association with the intestine. It is

unclear if any ST-29 isolates have been discovered in proximity to the intestine, but it has been shown that many virulence factors important for extra-intestinal pathogenesis are also important for binding to intestinal cells [123]. It is entirely possible these strains may be more common than once believed. Since they do not contain many of the 'classical' virulence factors or cause disease of a specific pathotype, then they may be overlooked in a clinical setting. Therefore, the lack of both intestinal and extraintestinal detection of ST-29 isolates may be due to a lack of surveillance utilizing appropriate sampling.

ACKNOWLEDGEMENTS

Albert Lee, who did many of the PCR reactions involving the attachment-related loci, was tremendously helpful in the generation of this data. Lindsey Ouellette did the MLST work on the ST-29s strains from the Seattle pediatric population and the strains from patients in England with Crohn's disease and colon cancer.

This research was supported by funds from the Centers for Disease Control and Prevention (Cooperative Agreement CCU015040) and NIAID, NIH, DHHS, under NIH research contract # N01-AI-30058 (TSW).

CHAPTER 4

GENETIC EROSION OF *E. Coli* TYPE III SECRETION SYSTEM 2 (ETT2) IN O157:H7 ISOLATES

SUMMARY

Recently, a cryptic gene cluster encoding a secondary type III secretion system (ETT2) was identified in enterohemorrhagic *Escherichia coli* O157:H7 strains. This locus contains genes displaying a high degree of similarity to the *inv-spa-prg* locus from the SPI-1 island in *Salmonella*. A 17-kb central region of ETT2 is thought to encode most of the putative proteins involved in the formation of the needle complex. We determined the acquisition point of this 17-kb region of ETT2 using an existing model for the stepwise evolution of modern strains of *E. coli* O157:H7 from ancestral O55:H7 strains. The gene cluster appears to be quite variable; truncated portions are also found in other Shiga toxin-producing *E. coli*, including serogroups O26 and O111. Strains were tested using gene-specific PCR primers for 11 of 17 open reading frames in ETT2, including the following genes: *eprK*, *eprH*, *epaS*, *epaP*, *eivJ*, *eivI*, *eivC*, *eivA*, *eivE*, *eivG*, and *eivF*. All twenty-two representatives from four electrophoretic types of O157:H7 descended from this lineage contained the entire ETT2 gene cluster. More ancestral O55:H7 strains that are *stx2* negative contain only a small terminal portion of ETT2. There are at least 3 different deletion types in the strains representing the rise of modern O157:H7, but all of those are found in the older O55:H7 strains before the somatic antigen shift to O157. Non-O157 isolates examined had other, distinct types of deletion patterns. It appears this island has an ancient point of acquisition, but has been undergoing different forms of genetic erosion to remove portions from the genome over time.

INTRODUCTION

Pathogenic *Escherichia coli* are responsible for numerous outbreaks of food-borne disease. Enterohemorrhagic *E. coli* (EHEC) strains of serotype O157:H7 and other non-O157 shiga toxin-producing *E. coli* (STEC) are the most common causes of a number of diseases, including bloody and non-bloody diarrhea, and hemolytic uremic syndrome. Variations in the severity of disease are due in part to the presence and absence of specific virulence factors [1, 115].

The mechanism of delivery of toxins and other effector molecules via a type III secretion system (TTSS) in *E. coli* has been extensively characterized [124-127]. The locus for enterocyte effacement (LEE) island encodes a functional TTSS in EHEC and enteropathogenic *E. coli* (EPEC). LEE also encodes the effectors intimin, an adhesin, and Tir, the receptor for intimin, which are responsible for the close association of the bacterial cell to the host epithelium [36]. The presence of LEE in isolates of *E. coli* confers the ability to form close associations of the bacterium and host cells leading to actin polymerization and cytoskeletal rearrangements characteristic of EPEC infections [128].

Recent genome sequence analysis has identified another island in *E. coli* that presumably also encodes a TTSS. This secondary type III system has been named ETT2 for *E. coli* type III secretion system 2, and is found in addition to the first TTSS encoded by LEE. Previous work [39] has shown the island contains open reading frames with a high similarity to some parts of a TTSS encoded by *Salmonella* called SPI-1 (*Salmonella* Pathogenicity Island 1). ETT2 is located at 64.5 min on the *E. coli* chromosome (O-island 121) and was presumably acquired by horizontal transfer because it has a lower G+C

content than the rest of the chromosome (36.9% vs. 50.8%) [39]. ETT2 is widely distributed, including in both O157:H7 and K12, but is usually found with many deletions [41].

SPI-1 in *Salmonella* is required for invasion of mammalian cells. The phenotype was first identified during a search for genes responsible for mediating bacterial entry into non-phagocytic cells [42, 129]. Encoded within SPI-1 are genes responsible for the formation of a needle complex capable of delivery of effector molecules, which are also encoded within SPI-1 and elsewhere in the genome of *Salmonella* [129]. The genetic structure of portions of ETT2 and SPI-1 are highly conserved [39]. A comparison of SPI-1 genes and ETT2 genes is shown in Table 7.

Previous work [39, 41] has identified the distribution of regions of ETT2 in various pathovars of *E. coli*. While some O157:H7 isolates contain the entire island, non-O157 STEC contain a truncated portion of the island that is missing the *eiv* gene cluster [39]. This study has developed a finer scale map of the components of ETT2 by screening for individual genes within the island via PCR. Parts of ETT2 are found in several O157 strains, however, the central region analyzed here is not present in the genome of K12, EPEC strains like E2348/69, or the uropathogen CFT073 [41]. A major portion of ETT2 is also missing in the O111:NM strain B171. The only known non-O157 isolate that appears to contain the entire ETT2 gene cluster is the enteroaggregative strain 042 [41] (Figure 8).

Primers were designed to detect 11 ORFs in a 17-kb central region of ETT2 that contains homology to the needle complex genes of SPI-1 (Table 8).

In this study, we report the acquisition point of the ETT2 island in an existing stepwise model of the evolution of the modern O157:H7 pathogenic clone as well as the distribution of individual components within a set of 57 isolates (mostly O157 and non-O157 STEC).

Hypotheses to be tested. 1. The region of ETT2 we are examining will be found intact and undeleted in all strains from the ancestral stepwise lineage related to modern O157:H7 isolates. Strains that represent the stepwise changes from ancestral EPEC-like O55:H7 to the modern EHEC O157:H7 will be screened by PCR for 11 individual genes within a 17-kb region encoding the structural apparatus of the type III secretion system.

2. We predict other non-O157 isolates will have a variety of deletion profiles, including some newly described deletion types. In addition to strains from the O157 stepwise model, additional strains from other pathogroups will be examined and we predict new deletion profiles unseen in previous studies will be found.

Table 7. Comparison of genes from the *Salmonella* Pathogenicity Island 1 (SPI-1) of *Salmonella* enterica serovar Typhimurium and their corresponding homologous counterpart in ETT2. Protein functions are listed only for SPI-1 homologues; ETT2 genes are thought to play a similar role due to sequence conservation. Amplicon sizes (for those ETT2 genes tested in this study) are included.

TTSS Genes		ETT2 Gene Size (bp)	Amplicon Size	Protein Function	Reference
ETT2	SPI-1				
eprK	*prgK*	735	387	required for invasion and secretion phenotypes - forms part of the base of the needle complex	[130, 131]
eprJ	*prgJ*	333	-	required for invasion and secretion phenotypes - may play a role in formation of needle complex	[130]
eprI	*prgI*	240	-	required for invasion and secretion phenotypes - forms needle in needle complex	[130]
eprH	*prgH*	735	416	required for invasion and secretion phenotypes - forms part of the base of the needle complex	[130, 131]
epaS	*spaS*	793	475	required component of export apparatus - probable inner membrane platform protein	[90]
epaR	*spaR*	786	-	required for protein secretion and *Salmonella* access to cultured epithelial cells	[90]
epaQ	*spaQ*	261	-	required for protein secretion and *Salmonella* access to cultured epithelial cells	[90]
epaP	*spaP*	666	366	required for protein secretion and *Salmonella* access to cultured epithelial cells	[90]
epaO	*spaO*	987	-	required for protein secretion and *Salmonella* access to cultured epithelial cells	[90]
eivJ	*invJ*	618	210	controls the length of the needle substructure of the needle complex	[90, 131-133]
eivI	*invI*	336	201	may function as a chaperone for InvJ secretion	[131-133]
eivC	*invC*	1320	355	ATPase activity - required for invasion of nonphagocytic cells	[134]
-	*invB*	-	-	chaperone - binds to SopA to stabilize it - also binds to SopE, SopE2	[135]
eivA	*invA*	1261	340	required for invasion of nonphagocytic cells - may form a channel in the membrane	[136, 137]
eivE	*invE*	1146	357	controls effector translocation - interacts with SipBC protein complex	[138-140]
eivG	*invG*	904	380	forms part of the base of the needle complex	[129, 131]
eivF	*invF*	750	277	transcriptional regulatory expression - required for efficient entry into host cells	[141]
-	*invH*	-	-	outer membrane lipoprotein required for proper localization of invG	[142]

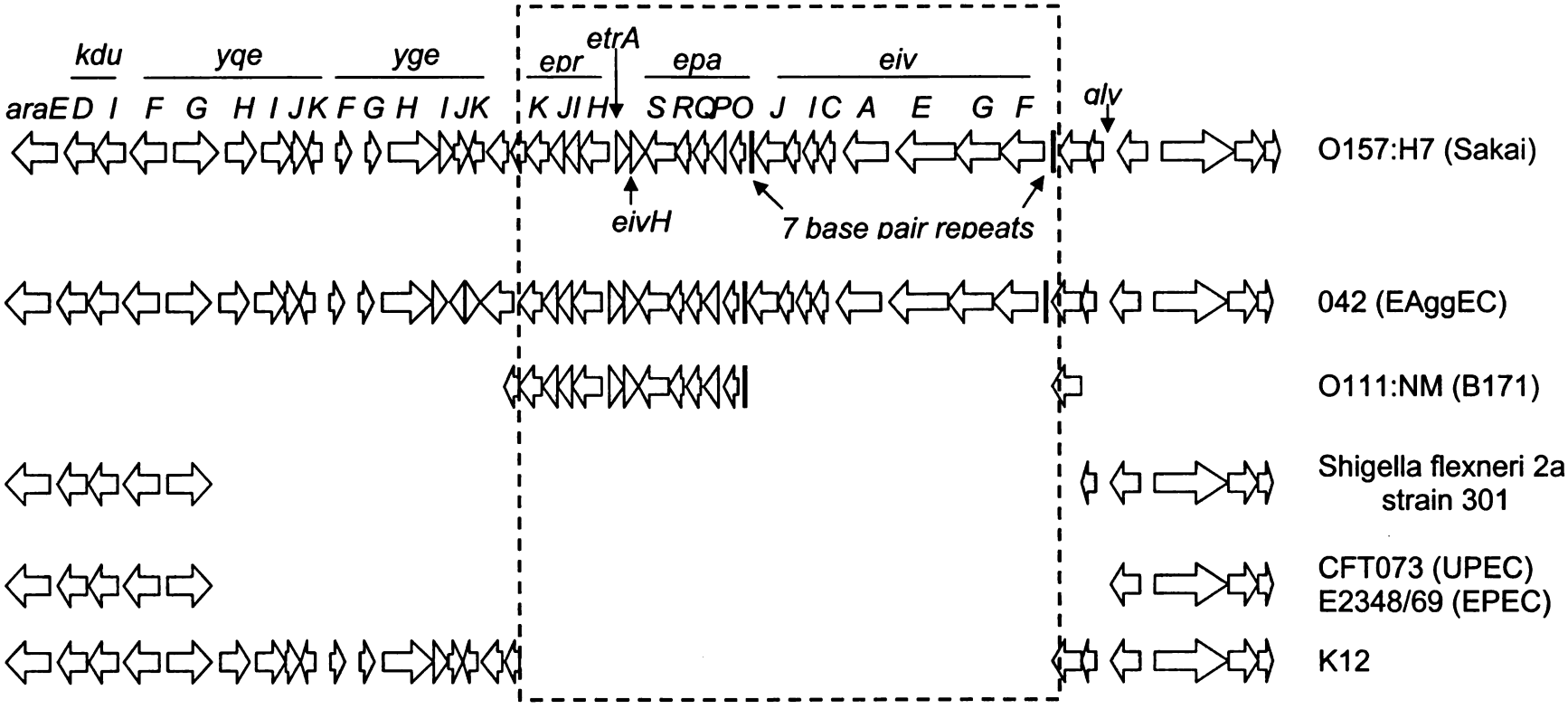


Figure 8. Organization of ETT2 gene cluster in *E. coli*. (A) - Large deletions of the structural genes (indicated by the dashed line) are seen in non-O157 STEC, such as B171, an O111:NM isolate, as well as K-12. *Shigella flexneri* 2a isolates, Uropathogenic (UPEC) strain CFT073 and EPEC E2348/69 are missing all of ETT2.

Table 8. Primer sequences used to amplify individual genes within ETT2. All primers were designed specifically for this study.

Gene	Primer	Sequence 5' -> 3'	Size (bp)
eprK	144F	TCAAGGCAAAACTGGATTC	387
	530R	TGTACGCTTGCAAAACTATTA	
eprH	60F	TGATCCGCAGCTGAGACATA	416
	475R	CAATAGCCTGACTTTCCAATACAT	
epaS	54F	AAAAGGGCAAATTCTAAAAAGTA	475
	528R	CTCTCCCCAATCTGATAATAAAG	
epaP	193F	ATGCCGGTTGGGAAGGAA	366
	558R	TACCGGACTCATCATCATCATACC	
eivJ	187F	TACCGCGATGGAGACAAAATAA	210
	396R	TCCCCAACGCTGAAACTGA	
eivI	44F	TACTCGGCGAAGTTGTATTT	201
	244R	TTTTTAGTCGGTCGTATTTTTCTC	
eivC	689F	GGTGTAATGCCGCCCTGATG	355
	1052R	ATTGCCGGATAGTGACCTTGACC	
eivA	88F	ATTCCATTGCCCACCTATTT	340
	410R	AGCGGGCAGCAACTTCAG	
eivE	416F	ATGTCCGGCTAAACGCTGAAG	357
	772R	CTGCCGACTGAAGAGACAATAG	
eivG	417F	GGACGGCAATGGTACTTTCTAT	380
	796R	AACGCGCTCCTGCTGTCG	
eivF	270F	TGAATCAACGGGGAGTGTGG	277
	546R	ACGGCGAAAATGTGAATACGATA	

MATERIALS & METHODS

PCR Primer Design. DNA sequences of ETT2 components *eprK, eprH, epaS, epaP, eivJ, eivI, eivC, eivA, eivE, eivG*, and *eivF* were obtained from GenBank. Homologous genes were aligned, with PCR primers designed from the conserved regions of each gene. All primers were synthesized by Integrated DNA Technologies (Coralville, IA.) and were stored at a concentration of 100 mM in ddH_2O. Working concentrations of each primer were 10 mM.

Strain Growth Conditions & DNA Extractions. All *E. coli* strains were grown overnight at 37 °C in 10 ml of sterile LB broth with moderate shaking, prior to PCR. DNA isolation was performed by suspending a single large colony of cells in 200 ml of sterile TE before heating at 95 °C for 10 min and centrifuging at 13,200 rpm for 10 min. All colonies were grown on LB agar plates overnight at 37 °C from freezer stocks created with single-colony picks of each strain.

PCR. Aliquots, (5 µl) of each DNA sample, were amplified in a 25-µl reaction mixture using the AmpliTaq Gold system (Applied Biosystems, Foster City, Calif.). Each reaction contained 2.5 µl 10X Gold buffer (150 mM Tris-HCl, pH 8.3, 500 mM KCl), 2.5 µl dNTPs (2 mM each dATP, dCTP, dGTP, and dTTP), 4.0 µl 25 mM $MgCl_2$, 1.25 µl of each 10 mM primer, 1.5 units AmpliTaq Gold, and 8.2 µl sterile ddH_2O. Amplification in a Hybaid PCR Express thermocycler (Hybaid Limited, Middlesex, England) utilized an initial denaturing step at 94 °C for 10 min., followed by 40 cycles of 92 °C for 1 min., 50 °C for 45 sec., and 72 °C for 45 sec. A final step of 72 °C for 5 min. was used for completion of any partially extended product. The positive control strain for all ETT2 genes was Sakai, an *E. coli* O157:H7 strain that the primers were designed

from sequence available on GenBank (accession number NC_002128). Negative controls included the laboratory strain K-12, an EPEC 1 strain E2348/69 and one reaction that contained no template DNA. Visualization of PCR products (5 μl) was achieved on ethidium bromide-stained 1.5% agarose gels via illumination with UV light.

Primer sequences and amplicon sizes for the individual ETT2 genes examined in this study are shown in Table 8.

RESULTS

ETT2 in EHEC O157. Phylogenetic analysis involving MLST of 7 housekeeping genes in a set of pathogenic strains allowed for the determination of the evolutionary history of the 57 strains tested (Table 9). The MLST loci sequenced here were the same loci applied to the Seattle pediatric population in Chapter 2 and the ST-29 isolates in Chapter 3.

Patterns of ETT2 gene detection were then overlaid onto the phylogenetic framework developed via MLST (Figure 9). This combination allowed for the discovery of possible points of acquisition of these virulence factors in evolutionary time.

The detection of 11 genes distributed across a 17-kb segment of ETT2 (Fig. 10) provides evidence for the existence of an intact island in all of the O157:H7/H- strains examined in this study. These finding agree with previous work [38, 39] on the distribution of ETT2 in these EHEC isolates. While Makino et al. characterized the presence of gene groups or regions of ETT2, this study characterizes ETT2, and the associated erosion of many genes within it, on a finer scale.

It is unclear if this secretion system is able to function in the cell. Because of the level of genetic attrition seen in this island, it seems unlikely that it is able to encode a fully functional needle complex [41]. However, some level of genetic expression or control apparently still remains within some genes in ETT2 [143, 144]. So even large-scale deletions seen in some strains my not entirely abolish expression of the remaining components of ETT2.

We predicted ETT2 would remain intact in all of the strains representing the stepwise evolution of modern O157. This hypothesis must be rejected based on the

deletion profiles observed in three different O55:H7 isolates. ETT2 is retained, in part, in these strains, and is intact in all known O157:H7 strains, but some isolates are undergoing erosion of this island by a variety of different deletions.

ETT2 in non-O157 STEC. As seen in previous work [38, 39], Shiga toxin-encoding non-O157 isolates had a truncated copy of ETT2 that was lacking the *eiv* gene cluster, containing 7 genes. All O26, O111 and other non-O157 serotypes contained a truncation at or near the *epa-eiv* gene cluster joint (Figure 10F). Although no experiments have been published on transformation of this truncated ETT2 island into another *E. coli* strain to see if it is functional, the lack of many critical genes needed for needle formation and proper assembly makes the possibility of functional type III secretion using ETT2 unlikely. Complementation experiments to restore a full copy of the island also have not been attempted. However, residual expression of some genes may still be possible, as was seen in O157:H7 [144].

We predicted to see new deletion types in non-O157 isolates. We accept this hypothesis due to the discovery of 3 new deletion profiles (types 4-6). Other existing deletions seen in strains like K-12 and CFT073 were also confirmed by this work.

ETT2 in O55:H7. A different truncation of ETT2 was seen when examining representatives of the O55:H7 serotypes. While other serotypes seem to be consistent in the presence or absence of ETT2, there appears to be some variation within O55:H7 strains. A stepwise model for the evolution of modern O157:H7 from ancestral O55:H7 isolates has been described previously [145]. Here, this model has been modified to show an ancient point of acquisition of ETT2 in modern O157:H7, and indicate a variety of deletions (Figure 11). Two of three isolates on the stepwise model after the acquisition of

the Stx2 phage (5905 and MDCH-10) were positive for all individual ETT2 genes tested, and presumably have an intact island (Figure 10B). These were the only O55:H7 strains examined in the study to be positive for all ETT2 genes examined. A third *stx2*+ O55:H7 strain, 97-3256, contained a unique type of deletion not seen in any other strain (Figure 10C). This strain was negative for *eprK*, *eprH* and *epaS*. However, this strain was positive for *epaP*, but may be missing the entire *epr* gene cluster and part of the adjoining *epa* cluster.

All four isolates representing more ancestral O55:H7 strains from before the acquisition of the Stx2 phage showed two distinct types of deletions of ETT2. Both DEC 5d and TB182A had a major portion of the *eiv* gene cluster, as well as all of the *epr* and *epa* genes that were not detectable via PCR (Figures 9-10). Microarray analysis involving

ORF comparisons in the genetic content of O157:H7 strain Sakai, K-12 and O55:H7 isolates DEC 5d and TB182A have confirmed this deletion in ETT2 [146]. This deletion also extends to 8 ORFs upstream of the *epa* genes in ETT2. The other two O55:H7 isolates on this stepwise model (DEC 5e and LTO55-62) contained another unique type of deletion (Figure 10E). All genes examined by PCR were positive, except for *epaS*. This may represent either sequence diversion in the primer site, or partial or total loss of the gene.

Table 9. Results of PCR screen of individual ETT2 genes from a set of pathogenic *E. coli* isolates. Positive results are indicated by a (+), negative results are blank. Strains are sorted by clonal groups and serotypes.

Strain	O	H	ST	Clonal Group	Group	*eprK*	*eprH*	*epaS*	*epaP*	*eivJ*	*eivI*	*eivC*	*eivA*	*eivE*	*eivG*	*eivF*
EDL-933	157	7	69	EHEC 1	**O157 Stepwise Set**	+	+	+	+	+	+	+	+	+	+	+
Sakai	157	7	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
93-111	157	7	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
G5101	157	7	65	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
OK-1	157	7	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
1628	157	7	65	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
1659	157	7	65	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
FDA413	157	NM	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
86-24	157	7	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
493/89	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
3112/93	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
0108(1)	157	NM	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
0108(2)	157	NM	66	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
210/89	157	-	75	EHEC 1	**German O157's**	+	+	+	+	+	+	+	+	+	+	+
514/91	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
5412/89	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
3748/89	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
4326/93	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
2664/91	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
919/90	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
1782/88	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
658/91	157	-	75	EHEC 1		+	+	+	+	+	+	+	+	+	+	+

Table 9 (cont'd)

Strain	O	H	ST	Clonal Group	Group	*eprK*	*eprH*	*epaS*	*epaP*	*eivJ*	*eivI*	*eivC*	*eivA*	*eivE*	*eivG*	*eivF*
5905	55	7	73	EHEC 1	**Other**	+	+	+	+	+	+	+	+	+	+	+
MDCH-10	55	7	73	EHEC 1	**O55's**	+	+	+	+	+	+	+	+	+	+	+
97-3256	55	7	73	EHEC 1					+	+	+	+	+	+	+	+
TB182A	55	7	73	EHEC 1									+	+	+	+
DEC 5a	55	7	73	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
DEC 5b	55	7	73	EHEC 1		+	+	+	+	+	+	+	+	+	+	+
DEC 5c	55	7	73	EHEC 1									+	+	+	+
DEC 5d	55	7	73	EHEC 1									+	+	+	+
DEC 5e	55	7	74	EHEC 1		+	+		+	+	+	+	+	+	+	+
LTO55-62	55	7	73	EHEC 1		+	+		+	+	+	+	+	+	+	+
LTO55-43	55	7	214	Unassigned		+	+	+	+	+	+	+	+	+	+	+
2666-74	26	[2]	106	EHEC 2	**O26**	+	+	+	+							
TB285A	26	2	106	EHEC 2	**Set**	+	+	+	+							
395-2	26	NM	106	EHEC 2		+	+	+	+							
413/89	26	-	106	EHEC 2		+	+	+	+							
DA-10	26	NM	106	EHEC 2		+	+	+	+							
DEC 8B	111	8	106	EHEC 2	**O111**	+	+	+	+							
3215-99	111	8	106	EHEC 2	**Set**	+	+	+	+							
CL-37	111	8	109	EHEC 2		+	+	+	+							
3007-85	111	NM	106	EHEC 2		+	+	+	+							
412/55	111	NM	106	EHEC 2		+	+	+	+							

Table 9 (cont'd)

Strain	O	H	ST	Clonal Group	Group	*eprK*	*eprH*	*epaS*	*epaP*	*eivJ*	*eivI*	*eivC*	*eivA*	*eivE*	*eivG*	*eivF*
75-83	145	NM	79	O145	**Other**	+	+	+	+							
B2F1	91	21	89	EHEC 2	**STEC**	+	+	+	+							
5482	118	16	106	EHEC 2		+	+	+	+							
90-1787	174	[21]	135	STEC 2		+	+	+	+							
CL-3	113	21	136	STEC 2		+	+	+	+							
MT#80	103	2	119	EPEC 2		+	+	+	+							
6:38	103	N	119	EPEC 2		+	+	+	+							
G5506	104	21	123	O104:H21		+	+									
DA-5	121	NM	182	STEC 3												
MT#2	121	19	182	STEC 3												
2747-71	F2a	NM	51	*S. flexneri*												
3823-69	D1	ND	82	*S dysenteriae*												
E2348/69	127	6	18	EPEC 1	**Controls**											
K-12	R		173	ECOR A												

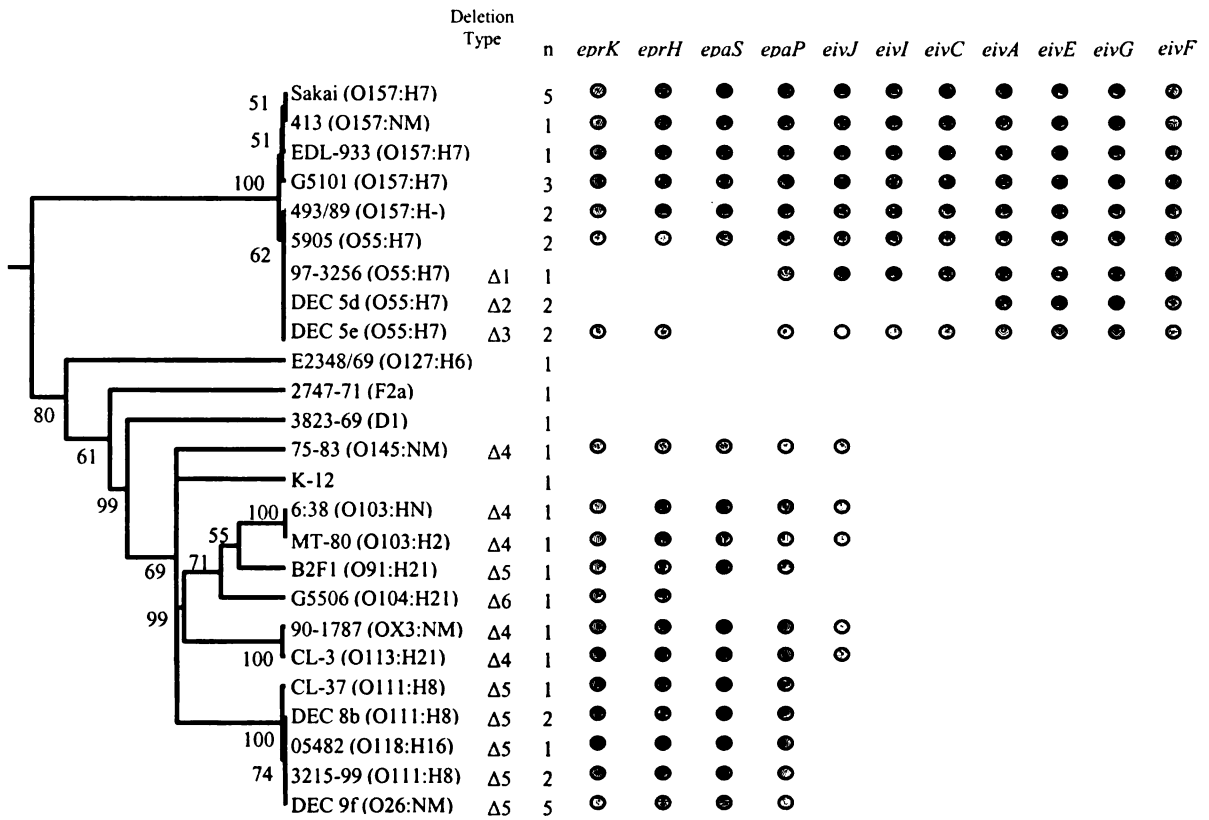


Figure 9. This neighbor-joining tree was constructed using the Kimura 2-parameter model of nucleotide substitution and shows the relationships between strains using 3732 nucleotide sites from 7 housekeeping genes (*aspC, clpX, fadD, icdA, lysP, mdh*, and *uidA*). Values for 500 bootstrap replications are given at the internal nodes. Gray filled symbols indicate positive PCR results for multiple representatives (n) of each sequence type tested in this study. Also indicated are the six distinct deletion types of this region.

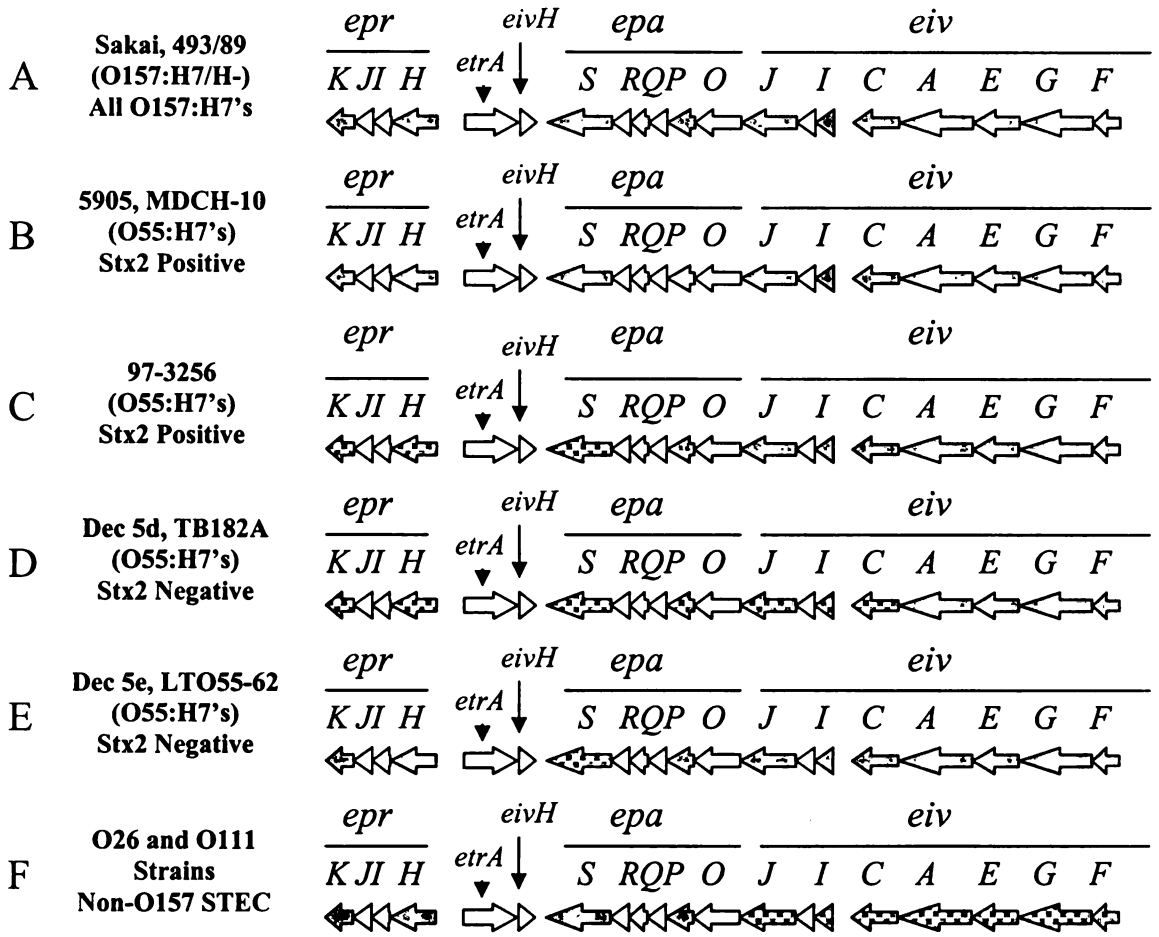


Figure 10. Results of PCR screen of individual genes in ETT2 showing different types of deletion patterns. All solid gray shaded arrows represent genes with positive PCR detection results. Genes negative for PCR screening are shaded in a grey checkerboard pattern, while genes not tested are shown in white. (A) Profile of positive results for O157:H7 and O157:H- isolates. All O157's were positive for all genes tested. (B) Two of three Stx2 positive O55:H7 strains were also positive for all genes tested, just like O157:H7 isolates. (C) A single stx2 positive O55:H7 strain had a distinct deletion type, missing the *epr* gene cluster and part of the *epa* cluster. (D) Stx2 negative O55:H7's results showing a different truncation than seen in non-O157 STEC strains, which is missing a majority of the genes examined. (E) Some Stx2-negative O55:H7 isolated showed the same profile as O157 strains, except for a single deletion in *epaS*. (B) The truncation seen in previous work with ETT2 in non-O157 STEC isolates was also found in this study.

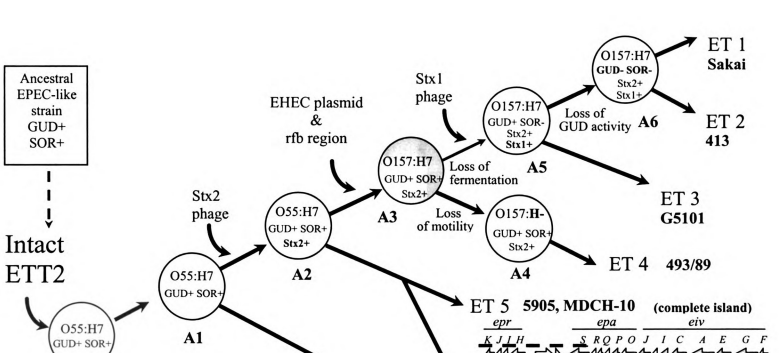


Figure 11. Stepwise evolution model for the emergence of the modern pathogenic clone of EHEC O157:H7 (ET1) from ancestral O55:H7 isolates (A1). The acquisition point of the ETT2 island is shown on the model as being ancient, given its widespread distribution in many other types of *E. coli*, including K12 and non-O157 STEC. The gray shaded Gud+, Sor+, Stx2+ O157:H7 at position A3 represents a type not yet seen any modern isolate. Only the descendant, the non-motile version (493/89), has been isolated.

DISCUSSION

Recent analysis of the genetic composition of ETT2 has shown a number of small deletions, frameshift mutations and premature stop codons [41]. This previous work, and the data presented here demonstrate a degree of genetic erosion in relatives of modern O157:H7 isolates, which contain various deletions. Despite this level of genetic attrition, some residual expression of ETT2 genes has been detected, which apparently plays a role in enhancing virulence [143, 144]. Since ETT2 may serve as a backup copy of a functional TTSS in *E. coli*, it is possible these isolates had, at one time, an entire ETT2 island then lost a major portion, along with a number of ORFs upstream, as a result of a deletion.

This work demonstrates the effect of genetic erosion on an apparently non-functional secretion system in strains representing the genetic ancestry of modern O157:H7. This erosion may be ongoing in a number of O55:H7 isolates. It is unclear why O157:H7 has retained an apparently intact ETT2 island, but this element may play a role in regulating the expression of other genes, rather than producing its own secretion system.

The type III secretion system encoded by LEE contains a number of effector molecules that are secreted through the needle apparatus [36]. Initial genomic analysis of ETT2 indicates this island may lack any apparent effectors [41]. This may explain, in part, why ETT2 is undergoing selective pressure to be deleted. Because it is complicated and requires a number of genes, the type III secretion machinery is energetically-expensive to assemble. Without a collection of effector molecules to be secreted into host cells, any needle apparatus that may have been encoded by ETT2 would have allowed for

the assembly of a complicated structure without any function. This is roughly akin to a house or other building without any occupants - it was expensive to build, but is now without function because of a lack of occupants.

The search for potential effector molecules for ETT2 is still underway. With the LEE-encoded TTSS, there have been a number of effectors identified that are encoded outside of the LEE island, but use LEE for secretion [147, 148]. Previous work with ETT2 identified potential effector molecules in an island dubbed EIP [41]. This island has curiously only been found in the enteroaggregative *E. coli* strain 042, which also contains a complete ETT2 (Figure 12). It is unclear why O157:H7 strains have an intact ETT2, but lack EIP. The potential effectors encoded in EIP, or other yet-uncharacterized effectors elsewhere may in fact work in conjunction with ETT2 to form a functional TTSS unit in some strains, although this has not yet been demonstrated.

Despite the possibility of non-ETT2 encoded effectors, none have yet been found in any strain of *E. coli* except 042. It is possible only strain 042 and a small number of other strains acquired EIP, leaving most strains with ETT2 without any functional effectors. This could be one reason for the frequent cases of genetic erosion outlined in this work and elsewhere. Obviously a cell copying large regions of DNA spanning 20-40 kb that encode no functional system or tools for survival is wasting valuable energy and resources. There must be strong selective pressures on these types of regions for either conversion to a functional form by recombination or most likely removal by deletion. While acquisition of genetic material can be a means of evolution, the removal of non-functional DNA may also help streamline the genome and enhance fitness.

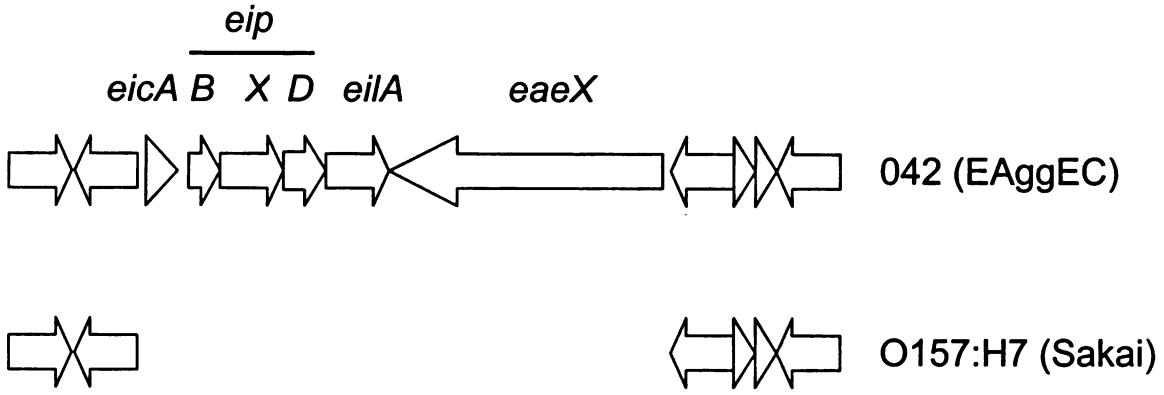


Figure 12. The EIP island from Enteroaggregative (EAggEC) strain 042, encodes possible ETT2 effectors, including an intimin homologue (*eaeX*). EIP has been identified only in a handful of strains, and is absent in all O157:H7 strains, including Sakai.

ACKNOWLEDGEMENTS

Special thanks to Dr. David Lacher for assistance in the design of the ETT2 primers, and for initial optimization advice. Thanks are also due to Lindsey Ouellette for MLST sequencing of the O157:H7 isolates used in this study.

This work was supported by a USDA National Needs Fellowship, NIH, and the Microbial Research Unit at Michigan State University.

CHAPTER 5

SUMMARY AND SYNTHESIS

The genetic diversity of *Escherichia coli* is dynamic, complex, and driven by a variety of external forces that help shape its evolution. The overall purpose of the research presented here is to characterize a small portion of that dynamic genome, to allow for the more efficient tracking and characterizing of unknown *E. coli* isolates, and to show in greater detail how gene flux is both a process of horizontal acquisition of mobile elements, and occasionally a loss of genetic material.

The need for rapid characterization of isolates becomes important with increasing threats to food safety, and ever-larger outbreaks of foodborne disease. MVGP is one technique that can be used to give an initial "fingerprint" of an isolate's virulence gene content, and characterize that strain to a particular pathotype within the diversity of *E. coli*. As shown in Chapter 2, this technique is also useful for characterizing strains isolated from a clinical setting. Because any gene can be added or removed easily, the set of virulence loci examined can be tailored exactly to specific strain sets, or to specific and investigative groups. The reagents and machinery used for MLST are similar or identical to what is necessary to perform MVGP, so labs already doing phylogenetic analysis can adapt their existing reagents and laboratory equipment to virulence profiling without much of a learning curve. Also, while the initial order of Beckman dye-labeled primers is costly, the amount used with each reaction is so small, so a primer may last for hundreds or even thousands of reactions.

Aside from the practical aspects of MVGP for use as a clinical tool, the work presented here also illustrates why it is important to screen for *E. coli* from specific groups of individuals. The microbiota of the gut is extremely diverse and *E. coli* plays only a small component of the flora under most circumstances. However, it is an

important component that can often cause or contribute to disease. It also appears to be highly variable with respect to virulence gene content, partially due to the mobile nature of those elements. Determining the basal level of virulence diversity within patients under different circumstances is important to understand not only the diversity of individual virulence genes, but also to link, statistically and epidemiologically, some of these virulence loci to specific clinical symptoms.

Many of the virulence loci examined here are listed as putative, and given the risks posed by many of the pathotypes in which they are found, they are likely to remain so indefinitely. The level of regulation and approvals needed for human trials, especially involving potentially pathogenic organisms is understandably complex. There are also obvious ethical considerations, as many of the virulence genes included here may produce proteins that could significantly affect human health. Despite these necessary restrictions, the molecular characterization of virulence-related loci is important. One of the first steps towards characterization is the determination of the distribution of these loci in both the general population and a variety of specific sub-populations, including those with a variety of diseases.

Future considerations. There are a number of future projects that could help expand the work started here. One includes a more detailed look at ST-29 loci, including microarray analysis to compare multiple isolates for genomic content, and also for expression differences. This may help identify any loci that could play a role in survival through enhanced attachment or immune system avoidance. Animal models could be

used to ascertain hypotheses that ST-29 strains do indeed have enhanced survival, because as of now this is only an epidemiological association.

Recent work with ETT2 has shown that despite many examples of gene loss within this region, and no apparent ETT2-specific effectors identified, some genes within this island are still actively transcribed [144]. Curiously, ETT2 may have an antagonistic effect, on the LEE-encoded TTSS in O157:H7. Strains with mutational inhibition of ETT2 loci *etrA* and *eivF* had both "greatly increased secretion of proteins by the LEE" and "increased adhesin to human epithelial cells" [144]. Exactly why ETT2 appears to be encoding proteins that inhibit LEE-related gene expression is still unclear. ETT2 would be another good candidate for expression microarray analysis to further characterize this phenomenon. It is possible that other loci within ETT2 are expressed, either in different strains, aside from O157's, or under different conditions. This expression may have a dramatic effect on other loci, as seen with *etrA* and *eivF* in ETT2 [144].

A further characterization of the genes encoded within the EIP island may also shed some more light on this situation. In vivo studies involving a fluorescent actin staining (FAS) assay [149] could show if ETT2 does in fact function as a type III secretion system. The assay works by detecting the polymerized actin filaments characteristic of the close adherence of strains attaching to epithelial cells. This polymerization is a hallmark of the attaching/effacing phenotype of EPEC and EHEC, and indicates type III effectors have entered the host cell to induce physiological changes [6, 125, 150].

An ideal candidate for this would be strain 042 since it has both an intact ETT2 and EIP and does not contain the other type III island encoded by LEE. Using a strain

without LEE is critical, so the actin polymerization phenotype could be attributed to ETT2, and not LEE.

Finally, MVGP could easily be either expanded or adapted for use on other microbial pathogens. Simply by selecting a set of PCR primers for specific genes of interest, a labeled pair can be ordered and incorporated into a collection to generate a profile. This technique may be especially useful in screening for virulence profiles of other species with high genomic diversity that are not well characterized regarding virulence genes. Organisms with frequent horizontal gene exchange would likely benefit most from virulence profiling because this technique allows for the tracking of numerous virulence loci across the diversity of the species, as it has with *E. coli*.

Evolution in all types of *E. coli* appears to be a dynamic process, with many documented cases of horizontal acquisition of bacteriophages [27, 151], plasmids [152-154], and even large chromosomal pathogenicity islands [155, 156]. Even amongst strains of the same serotype, such as the O157:H7's, increased virulence is seen between distinct groups, or clades [157]. The O157:H7 strains responsible for the recent outbreaks on contaminated spinach [158, 159] and lettuce [158, 160] in late 2006 had enhanced virulence compared to O157:H7 isolates from past outbreaks (Sakai from 1996 or EDL-933 from 1982). Virulence was measured by the rate of hospitalization and development of HUS. Both of these markers were significantly higher in the spinach and lettuce outbreak strains, which belong to clade 8, than previous O157:H7 outbreak strains Sakai (clade 1) and EDL-933 (clade 3) [157]. This is evidence for ongoing evolution that appears to be favoring enhanced virulence, possibly as a means to spread more effectively. It is also possible that other serotypes of *E. coli* that are not frequently

detected in the United States may become more prevalent here in the future. Geographical preferences appear to exist in the distribution of certain serotypes, but this may change given today's global environment.

In the future, it seems reasonable to expect new outbreaks of highly virulent serotypes of *E. coli*, like O157:H7, to result in elevated levels of HUS, hospitalization, and death in affected individuals. Therefore, the urgency for rapid and coordinated identification of outbreaks should grow in the future, whether applied to food safety for the consumer and industry or to bioterrorism.

REFERENCES

1. Kaper, J.B., J.P. Nataro, and H.L. Mobley, *Pathogenic Escherichia coli.* Nat Rev Microbiol, 2004. **2**(2): p. 123-40.

2. Donnenberg, M.S. and T.S. Whittam, *Pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic Escherichia coli.* J Clin Invest, 2001. **107**(5): p. 539-48.

3. Tobe, T., et al., *Complete DNA sequence and structural analysis of the enteropathogenic Escherichia coli adherence factor plasmid.* Infect Immun, 1999. **67**(10): p. 5455-62.

4. Navarro-Garcia, F., et al., *The serine protease motif of EspC from enteropathogenic Escherichia coli produces epithelial damage by a mechanism different from that of Pet toxin from enteroaggregative E. coli.* Infect Immun, 2004. **72**(6): p. 3609-21.

5. Mellies, J.L., et al., *espC pathogenicity island of enteropathogenic Escherichia coli encodes an enterotoxin.* Infect Immun, 2001. **69**(1): p. 315-24.

6. Donnenberg, M.S., J.B. Kaper, and B.B. Finlay, *Interactions between enteropathogenic Escherichia coli and host epithelial cells.* Trends Microbiol, 1997. **5**(3): p. 109-14.

7. Donnenberg, M.S., L.C. Lai, and K.A. Taylor, *The locus of enterocyte effacement pathogenicity island of enteropathogenic Escherichia coli encodes secretion functions and remnants of transposons at its extreme right end.* Gene, 1997. **184**(1): p. 107-14.

8. Cleary, T.G., *The role of Shiga-toxin-producing Escherichia coli in hemorrhagic colitis and hemolytic uremic syndrome.* Semin Pediatr Infect Dis, 2004. **15**(4): p. 260-5.

9. Burland, V., et al., *The complete DNA sequence and analysis of the large virulence plasmid of Escherichia coli O157:H7.* Nucleic Acids Res, 1998. **26**(18): p. 4196-204.

10. Tatsuno, I., et al., *toxB gene on pO157 of enterohemorrhagic Escherichia coli O157:H7 is required for full epithelial cell adherence phenotype.* Infect Immun, 2001. **69**(11): p. 6660-9.

11. Merritt, E.A., et al., *Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide.* Protein Sci, 1994. **3**(2): p. 166-75.

12. Lencer, W.I., et al., *Targeting of cholera toxin and Escherichia coli heat labile toxin in polarized epithelia: role of COOH-terminal KDEL.* J Cell Biol, 1995. **131**(4): p. 951-62.

13. Rodighiero, C., et al., *Structural basis for the differential toxicity of cholera toxin and Escherichia coli heat-labile enterotoxin. Construction of hybrid toxins identifies the A2-domain as the determinant of differential toxicity.* J Biol Chem, 1999. **274**(7): p. 3962-9.

14. Sansonetti, P.J., et al., *Molecular comparison of virulence plasmids in Shigella and enteroinvasive Escherichia coli.* Ann Microbiol (Paris), 1983. **134A**(3): p. 295-318.

15. Harris, J.R., et al., *High-molecular-weight plasmid correlates with Escherichia coli enteroinvasiveness.* Infect Immun, 1982. **37**(3): p. 1295-8.

16. Navarro-Garcia, F., et al., *Cytoskeletal effects induced by pet, the serine protease enterotoxin of enteroaggregative Escherichia coli.* Infect Immun, 1999. **67**(5): p. 2184-92.

17. Eslava, C., et al., *Pet, an autotransporter enterotoxin from enteroaggregative Escherichia coli.* Infect Immun, 1998. **66**(7): p. 3155-63.

18. Vila, J., et al., *Enteroaggregative escherichia coli virulence factors in traveler's diarrhea strains.* J Infect Dis, 2000. **182**(6): p. 1780-3.

19. Fasano, A., et al., *Effect of shigella enterotoxin 1 (ShET1) on rabbit intestine in vitro and in vivo.* Gut, 1997. **40**(4): p. 505-11.

20. Fasano, A., et al., *Shigella enterotoxin 1: an enterotoxin of Shigella flexneri 2a active in rabbit small intestine in vivo and in vitro.* J Clin Invest, 1995. **95**(6): p. 2853-61.

21. Savarino, S.J., et al., *Enteroaggregative Escherichia coli heat-stable enterotoxin is not restricted to enteroaggregative E. coli.* J Infect Dis, 1996. **173**(4): p. 1019-22.

22. Itoh, Y., et al., *Laboratory investigation of enteroaggregative Escherichia coli O untypeable:H10 associated with a massive outbreak of gastrointestinal illness.* J Clin Microbiol, 1997. **35**(10): p. 2546-50.

23. Cobeljic, M., et al., *Enteroaggregative Escherichia coli associated with an outbreak of diarrhoea in a neonatal nursery ward.* Epidemiol Infect, 1996. **117**(1): p. 11-6.

24. Nataro, J.P., et al., *Heterogeneity of enteroaggregative Escherichia coli virulence demonstrated in volunteers.* J Infect Dis, 1995. **171**(2): p. 465-8.

25. Mathewson, J.J., et al., *Pathogenicity of enteroadherent Escherichia coli in adult volunteers.* J Infect Dis, 1986. **154**(3): p. 524-7.

26. Herold, S., H. Karch, and H. Schmidt, *Shiga toxin-encoding bacteriophages--genomes in motion.* Int J Med Microbiol, 2004. **294**(2-3): p. 115-21.

27. Schmidt, H., *Shiga-toxin-converting bacteriophages.* Res Microbiol, 2001. **152**(8): p. 687-95.

28. O'Brien, A.D., et al., *Shiga toxin: biochemistry, genetics, mode of action, and role in pathogenesis.* Curr Top Microbiol Immunol, 1992. **180**: p. 65-94.

29. Ewald, P.W., *Guarding against the most dangerous emerging pathogens.* Emerg Infect Dis, 1996. **2**(4): p. 245-57.

30. Wells, J.G., et al., *Laboratory investigation of hemorrhagic colitis outbreaks associated with a rare Escherichia coli serotype.* J Clin Microbiol, 1983. **18**(3): p. 512-20.

31. Riley, L.W., et al., *Hemorrhagic colitis associated with a rare Escherichia coli serotype.* N Engl J Med, 1983. **308**(12): p. 681-5.

32. Mead, P.S., et al., *Food-related illness and death in the United States.* Emerg Infect Dis, 1999. **5**(5): p. 607-25.

33. Karmali, M.A., *Infection by Shiga toxin-producing Escherichia coli: an overview.* Mol Biotechnol, 2004. **26**(2): p. 117-22.

34. Perna, N.T., et al., *Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.* Nature, 2001. **409**(6819): p. 529-33.

35. Welch, R.A., et al., *Extensive mosaic structure revealed by the complete genome sequence of uropathogenic Escherichia coli.* Proc Natl Acad Sci U S A, 2002. **99**(26): p. 17020-4.

36. Elliott, S.J., et al., *The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic Escherichia coli E2348/69.* Mol Microbiol, 1998. **28**(1): p. 1-4.

37. Sperandio, V., et al., *Characterization of the locus of enterocyte effacement (LEE) in different enteropathogenic Escherichia coli (EPEC) and Shiga-toxin producing Escherichia coli (STEC) serotypes.* FEMS Microbiol Lett, 1998. **164**(1): p. 133-9.

38. Hartleib, S., et al., *Prevalence of the new, SPI1-like, pathogenicity island ETT2 among Escherichia coli.* Int J Med Microbiol, 2003. **292**(7-8): p. 487-93.

39. Makino, S., et al., *Distribution of the secondary type III secretion system locus found in enterohemorrhagic Escherichia coli O157:H7 isolates among Shiga toxin-producing E. coli strains.* J Clin Microbiol, 2003. **41**(6): p. 2341-7.

40. Galan, J.E. and H. Wolf-Watz, *Protein delivery into eukaryotic cells by type III secretion machines.* Nature, 2006. **444**(7119): p. 567-73.

41. Ren, C.P., et al., *The ETT2 gene cluster, encoding a second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition.* J Bacteriol, 2004. **186**(11): p. 3547-60.

42. Galan, J.E. and A. Collmer, *Type III secretion machines: bacterial devices for protein delivery into host cells.* Science, 1999. **284**(5418): p. 1322-8.

43. Pallen, M.J., R.R. Chaudhuri, and I.R. Henderson, *Genomic analysis of secretion systems.* Curr Opin Microbiol, 2003. **6**(5): p. 519-27.

44. Ochman, H. and R.K. Selander, *Standard reference strains of Escherichia coli from natural populations.* J Bacteriol, 1984. **157**(2): p. 690-3.

45. Welinder-Olsson, C., et al., *Genetic profiling of enterohemorrhagic Escherichia coli strains in relation to clonality and clinical signs of infection.* J Clin Microbiol, 2002. **40**(3): p. 959-64.

46. Obata-Yasuoka, M., et al., *Vaginal Escherichia coli share common virulence factor profiles, serotypes and phylogeny with other extraintestinal E. coli.* Microbiology, 2002. **148**(Pt 9): p. 2745-52.

47. Tarr, C.L., et al., *Molecular characterization of a serotype O121:H19 clone, a distinct Shiga toxin-producing clone of pathogenic Escherichia coli.* Infect Immun, 2002. **70**(12): p. 6853-9.

48. Feil, E.J., et al., *Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data.* Genetics, 2000. **154**(4): p. 1439-50.

49. Kotetishvili, M., et al., *Multilocus sequence typing for characterization of clinical and environmental salmonella strains.* J Clin Microbiol, 2002. **40**(5): p. 1626-35.

50. Nicolas, P., et al., *The 1998 Senegal epidemic of meningitis was due to the clonal expansion of A:4:P1.9, clone III-1, sequence type 5 Neisseria meningitidis strains.* J Clin Microbiol, 2000. **38**(1): p. 198-200.

51. Zhou, J., M.C. Enright, and B.G. Spratt, *Identification of the major Spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing.* J Clin Microbiol, 2000. **38**(3): p. 977-86.

52. Kriz, P., J. Kalmusova, and J. Felsberg, *Multilocus sequence typing of Neisseria meningitidis directly from cerebrospinal fluid.* Epidemiol Infect, 2002. **128**(2): p. 157-60.

53. Urwin, R. and M.C. Maiden, *Multi-locus sequence typing: a tool for global epidemiology.* Trends Microbiol, 2003. **11**(10): p. 479-87.

54. Bizzarri, M.F., A. Prabhakar, and A.H. Bishop, *Multiple-Locus Sequence Typing Analysis of Bacillus thuringiensis Recovered from the Phylloplane of Clover (Trifolium hybridum) in Vegetative Form.* Microb Ecol, 2007.

55. Tarr, C.L. and T.S. Whittam, *Molecular evolution of the intimin gene in O111 clones of pathogenic Escherichia coli.* J Bacteriol, 2002. **184**(2): p. 479-87.

56. Lacher, D.W., et al., *Molecular evolution of typical enteropathogenic Escherichia coli: clonal analysis by multilocus sequence typing and virulence gene allelic profiling.* J Bacteriol, 2007. **189**(2): p. 342-50.

57. Walk, S.T., et al., *Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches.* Environ Microbiol, 2007. **9**(9): p. 2274-88.

58. Johnson, J.R., et al., *Molecular comparison of extraintestinal Escherichia coli isolates of the same electrophoretic lineages from humans and domestic animals.* J Infect Dis, 2001. **183**(1): p. 154-9.

59. Hyma, K.E., et al., *Evolutionary genetics of a new pathogenic Escherichia species: Escherichia albertii and related Shigella boydii strains.* J Bacteriol, 2005. **187**(2): p. 619-28.

60. Tarr, C.L., et al., *Molecular characterization reveals similar virulence gene content in unrelated clonal groups of Escherichia coli of serogroup O174 (OX3).* J Bacteriol, 2008. **190**(4): p. 1344-9.

61. Stapp, J.R., et al., *Comparison of Escherichia coli O157:H7 antigen detection in stool and broth cultures to that in sorbitol-MacConkey agar stool cultures.* J Clin Microbiol, 2000. **38**(9): p. 3404-6.

62. Klein, E.J., et al., *Diarrhea etiology in a Children's Hospital Emergency Department: a prospective cohort study.* Clin Infect Dis, 2006. **43**(7): p. 807-13.

63. Klein, E.J., et al., *Shiga toxin-producing Escherichia coli in children with diarrhea: a prospective point-of-care study.* J Pediatr, 2002. **141**(2): p. 172-7.

64. Monteiro-Neto, V., et al., *Virulence properties of Escherichia coli O111:H12 strains.* FEMS Microbiol Lett, 1997. **146**(1): p. 123-8.

65. Feng, P., S.D. Weagant, and S.R. Monday, *Genetic analysis for virulence factors in Escherichia coli O104:H21 that was implicated in an outbreak of hemorrhagic colitis.* J Clin Microbiol, 2001. **39**(1): p. 24-8.

66. Stacy-Phipps, S., J.J. Mecca, and J.B. Weiss, *Multiplex PCR assay and simple preparation method for stool specimens detect enterotoxigenic Escherichia coli DNA during course of infection.* J Clin Microbiol, 1995. **33**(5): p. 1054-9.

67. Schubert, S., et al., *Prevalence of the "high-pathogenicity island" of Yersinia species among Escherichia coli strains that are pathogenic to humans.* Infect Immun, 1998. **66**(2): p. 480-5.

68. Paton, A.W. and J.C. Paton, *Direct detection and characterization of Shiga toxigenic Escherichia coli by multiplex PCR for stx1, stx2, eae, ehxA, and saa.* J Clin Microbiol, 2002. **40**(1): p. 271-4.

69. Kumar, S., et al., *MEGA2: molecular evolutionary genetics analysis software.* Bioinformatics, 2001. **17**(12): p. 1244-5.

70. Okeke, I.N., et al., *Molecular epidemiology of the iron utilization genes of enteroaggregative Escherichia coli.* J Clin Microbiol, 2004. **42**(1): p. 36-44.

71. Karch, H., et al., *A genomic island, termed high-pathogenicity island, is present in certain non-O157 Shiga toxin-producing Escherichia coli clonal lineages.* Infect Immun, 1999. **67**(11): p. 5994-6001.

72. Oswald, E., et al., *Typing of intimin genes in human and animal enterohemorrhagic and enteropathogenic Escherichia coli: characterization of a new intimin variant.* Infect Immun, 2000. **68**(1): p. 64-71.

73. Brunder, W., et al., *The large plasmids of Shiga-toxin-producing Escherichia coli (STEC) are highly variable genetic elements.* Microbiology, 1999. **145 (Pt 5)**: p. 1005-14.

74. Velarde, J.J. and J.P. Nataro, *Hydrophobic residues of the autotransporter EspP linker domain are important for outer membrane translocation of its passenger.* J Biol Chem, 2004. **279**(30): p. 31495-504.

75. McNally, A., et al., *Differences in levels of secreted locus of enterocyte effacement proteins between human disease-associated and bovine Escherichia coli O157.* Infect Immun, 2001. **69**(8): p. 5107-14.

76. Brunder, W., H. Schmidt, and H. Karch, *EspP, a novel extracellular serine protease of enterohaemorrhagic Escherichia coli O157:H7 cleaves human coagulation factor V.* Mol Microbiol, 1997. **24**(4): p. 767-78.

77. Bielaszewska, M., et al., *Characterization of cytolethal distending toxin genes and expression in shiga toxin-producing Escherichia coli strains of non-O157 serogroups.* Infect Immun, 2004. **72**(3): p. 1812-6.

78. Bielaszewska, M., et al., *Cytolethal distending toxin from Shiga toxin-producing Escherichia coli O157 causes irreversible G2/M arrest, inhibition of proliferation, and death of human endothelial cells.* Infect Immun, 2005. **73**(1): p. 552-62.

79. Nesic, D., Y. Hsu, and C.E. Stebbins, *Assembly and function of a bacterial genotoxin.* Nature, 2004. **429**(6990): p. 429-33.

80. Pickett, C.L., et al., *Patterns of variations in Escherichia coli strains that produce cytolethal distending toxin.* Infect Immun, 2004. **72**(2): p. 684-90.

81. Torres, A.G. and S.M. Payne, *Haem iron-transport system in enterohaemorrhagic Escherichia coli O157:H7.* Mol Microbiol, 1997. **23**(4): p. 825-33.

82. Nagy, G., et al., *Expression of hemin receptor molecule ChuA is influenced by RfaH in uropathogenic Escherichia coli strain 536.* Infect Immun, 2001. **69**(3): p. 1924-8.

83. Tarr, P.I., et al., *Iha: a novel Escherichia coli O157:H7 adherence-conferring molecule encoded on a recently acquired chromosomal island of conserved structure.* Infect Immun, 2000. **68**(3): p. 1400-7.

84. Taylor, D.E., et al., *Genomic variability of O islands encoding tellurite resistance in enterohemorrhagic Escherichia coli O157:H7 isolates.* J Bacteriol, 2002. **184**(17): p. 4690-8.

85. Heimer, S.R., et al., *Urease of enterohemorrhagic Escherichia coli: evidence for regulation by fur and a trans-acting factor.* Infect Immun, 2002. **70**(2): p. 1027-31.

86. D'Orazio, S.E. and C.M. Collins, *The plasmid-encoded urease gene cluster of the family Enterobacteriaceae is positively regulated by UreR, a member of the AraC family of transcriptional activators.* J Bacteriol, 1993. **175**(11): p. 3459-67.

87. Wang, G., C.G. Clark, and F.G. Rodgers, *Detection in Escherichia coli of the genes encoding the major virulence factors, the genes defining the O157:H7 serotype, and components of the type 2 Shiga toxin family by multiplex PCR.* J Clin Microbiol, 2002. **40**(10): p. 3613-9.

88. De Baets, L., et al., *Genetic typing of shiga toxin 2 variants of Escherichia coli by PCR-restriction fragment length polymorphism analysis.* Appl Environ Microbiol, 2004. **70**(10): p. 6309-14.

89. Eklund, M., K. Leino, and A. Siitonen, *Clinical Escherichia coli strains carrying stx genes: stx variants and stx-positive virulence profiles.* J Clin Microbiol, 2002. **40**(12): p. 4585-93.

90. Collazo, C.M. and J.E. Galan, *Requirement for exported proteins in secretion through the invasion-associated type III system of Salmonella typhimurium.* Infect Immun, 1996. **64**(9): p. 3524-31.

91. Crago, A.M. and V. Koronakis, *Salmonella InvG forms a ring-like multimer that requires the InvH lipoprotein for outer membrane localization.* Mol Microbiol, 1998. **30**(1): p. 47-56.

92. Hwang, J., et al., *Structural and topographical studies of the type IV bundle-forming pilus assembly complex of enteropathogenic Escherichia coli.* J Bacteriol, 2003. **185**(22): p. 6695-701.

93. Green, B.A., et al., *Evidence that a new enterotoxin of Escherichia coli which activates adenylate cyclase in eucaryotic target cells is not plasmid mediated.* Infect Immun, 1983. **41**(1): p. 383-90.

94. Yamamoto, T. and T. Yokota, *Release of heat-labile enterotoxin subunits by Escherichia coli.* J Bacteriol, 1982. **150**(3): p. 1482-4.

95. Wyckoff, E.E., et al., *Structure of the Shigella dysenteriae haem transport locus and its phylogenetic distribution in enteric bacteria.* Mol Microbiol, 1998. **28**(6): p. 1139-52.

96. Mills, M. and S.M. Payne, *Identification of shuA, the gene encoding the heme receptor of Shigella dysenteriae, and analysis of invasion and intracellular multiplication of a shuA mutant.* Infect Immun, 1997. **65**(12): p. 5358-63.

97. Henderson, I.R., et al., *Characterization of pic, a secreted protease of Shigella flexneri and enteroaggregative Escherichia coli.* Infect Immun, 1999. **67**(11): p. 5587-96.

98. Guyer, D.M., et al., *Identification of sat, an autotransporter toxin produced by uropathogenic Escherichia coli.* Mol Microbiol, 2000. **38**(1): p. 53-66.

99. Guyer, D.M., et al., *Sat, the secreted autotransporter toxin of uropathogenic Escherichia coli, is a vacuolating cytotoxin for bladder and kidney epithelial cells.* Infect Immun, 2002. **70**(8): p. 4539-46.

100. Dutta, P.R., et al., *Functional comparison of serine protease autotransporters of enterobacteriaceae.* Infect Immun, 2002. **70**(12): p. 7105-13.

101. Al-Hasani, K., et al., *Distribution and structural variation of the she pathogenicity island in enteric bacterial pathogens.* J Med Microbiol, 2001. **50**(9): p. 780-6.

102. Al-Hasani, K., et al., *The sigA gene which is borne on the she pathogenicity island of Shigella flexneri 2a encodes an exported cytopathic protease involved in intestinal fluid accumulation.* Infect Immun, 2000. **68**(5): p. 2457-63.

103. Venkatesan, M.M., et al., *Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri.* Infect Immun, 2001. **69**(5): p. 3271-85.

104. Rode, C.K., et al., *Type-specific contributions to chromosome size differences in Escherichia coli.* Infect Immun, 1999. **67**(1): p. 230-6.

105. Huang, S.H., et al., *Escherichia coli invasion of brain microvascular endothelial cells in vitro and in vivo: molecular cloning and characterization of invasion gene ibe10.* Infect Immun, 1995. **63**(11): p. 4470-5.

106. Litwin, C.M. and S.B. Calderwood, *Role of iron in regulation of virulence genes.* Clin Microbiol Rev, 1993. **6**(2): p. 137-49.

107. Grandjean, D., et al., *Iron uptake is essential for Escherichia coli survival in drinking water.* Lett Appl Microbiol, 2006. **43**(1): p. 111-7.

108. Xie, Y., et al., *Identification and characterization of Escherichia coli RS218-derived islands in the pathogenesis of E. coli meningitis.* J Infect Dis, 2006. **194**(3): p. 358-64.

109. Johnson, J.R., et al., *Phylogenetic distribution of extraintestinal virulence-associated traits in Escherichia coli.* J Infect Dis, 2001. **183**(1): p. 78-88.

110. Ostroff, S.M., et al., *A statewide outbreak of Escherichia coli O157:H7 infections in Washington State.* Am J Epidemiol, 1990. **132**(2): p. 239-47.

111. Tarr, P.I., et al., *Escherichia coli O157:H7 and the hemolytic uremic syndrome: importance of early cultures in establishing the etiology.* J Infect Dis, 1990. **162**(2): p. 553-6.

112. Tarr, P.I., N.T. Tran, and R.A. Wilson, *Escherichia coli O157:H7 in retail ground beef in Seattle: results of a one-year prospective study.* J Food Prot, 1999. **62**(2): p. 133-9.

113. Haack, J.P., et al., *Escherichia coli O157 exposure in Wyoming and Seattle: serologic evidence of rural risk.* Emerg Infect Dis, 2003. **9**(10): p. 1226-31.

114. Denno, D.M., et al., *Explaining unexplained diarrhea and associating risks and infections.* Anim Health Res Rev, 2007. **8**(1): p. 69-80.

115. Nataro, J.P. and J.B. Kaper, *Diarrheagenic Escherichia coli.* Clin Microbiol Rev, 1998. **11**(1): p. 142-201.

116. Achtman, M., et al., *Six widespread bacterial clones among Escherichia coli K1 isolates.* Infect Immun, 1983. **39**(1): p. 315-35.

117. Bloch, C.A. and C.K. Rode, *Pathogenicity island evaluation in Escherichia coli K1 by crossing with laboratory strain K-12.* Infect Immun, 1996. **64**(8): p. 3218-23.

118. Hochhut, B., et al., *Role of pathogenicity island-associated integrases in the genome plasticity of uropathogenic Escherichia coli strain 536.* Mol Microbiol, 2006. **61**(3): p. 584-95.

119. Johnson, J.R., et al., *Phylogenetic and pathotypic similarities between Escherichia coli isolates from urinary tract infections in dogs and extraintestinal infections in humans.* J Infect Dis, 2001. **183**(6): p. 897-906.

120. Johnson, J.R., et al., *Virulence genotypes and phylogenetic background of Escherichia coli serogroup O6 isolates from humans, dogs, and cats.* J Clin Microbiol, 2008. **46**(2): p. 417-22.

121. Oelschlaeger, T.A., U. Dobrindt, and J. Hacker, *Virulence factors of uropathogens.* Curr Opin Urol, 2002. **12**(1): p. 33-8.

122. Martin, H.M., et al., *Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer.* Gastroenterology, 2004. **127**(1): p. 80-93.

123. Adlerberth, I., et al., *Adhesins of Escherichia coli associated with extra-intestinal pathogenicity confer binding to colonic epithelial cells.* Microb Pathog, 1995. **18**(6): p. 373-85.

124. Jores, J., L. Rumer, and L.H. Wieler, *Impact of the locus of enterocyte effacement pathogenicity island on the evolution of pathogenic Escherichia coli.* Int J Med Microbiol, 2004. **294**(2-3): p. 103-13.

125. Mota, L.J. and G.R. Cornelis, *The bacterial injection kit: type III secretion systems.* Ann Med, 2005. **37**(4): p. 234-49.

126. Clarke, S.C., et al., *Virulence of enteropathogenic Escherichia coli, a global pathogen.* Clin Microbiol Rev, 2003. **16**(3): p. 365-78.

127. Zaharik, M.L., et al., *Delivery of dangerous goods: type III secretion in enteric pathogens.* Int J Med Microbiol, 2002. **291**(8): p. 593-603.

128. Elliott, S.J., J. Yu, and J.B. Kaper, *The cloned locus of enterocyte effacement from enterohemorrhagic Escherichia coli O157:H7 is unable to confer the attaching and effacing phenotype upon E. coli K-12.* Infect Immun, 1999. **67**(8): p. 4260-3.

129. Galan, J.E., *Salmonella interactions with host cells: type III secretion at work.* Annu Rev Cell Dev Biol, 2001. **17**: p. 53-86.

130. Klein, J.R., T.F. Fahlen, and B.D. Jones, *Transcriptional organization and function of invasion genes within Salmonella enterica serovar Typhimurium pathogenicity island 1, including the prgH, prgI, prgJ, prgK, orgA, orgB, and orgC genes.* Infect Immun, 2000. **68**(6): p. 3368-76.

131. Kubori, T., et al., *Molecular characterization and assembly of the needle complex of the Salmonella typhimurium type III protein secretion system.* Proc Natl Acad Sci U S A, 2000. **97**(18): p. 10225-30.

132. Collazo, C.M., M.K. Zierler, and J.E. Galan, *Functional analysis of the Salmonella typhimurium invasion genes invI and invJ and identification of a target of the protein secretion apparatus encoded in the inv locus.* Mol Microbiol, 1995. **15**(1): p. 25-38.

133. Russmann, H., et al., *Molecular and functional analysis of the type III secretion signal of the Salmonella enterica InvJ protein.* Mol Microbiol, 2002. **46**(3): p. 769-79.

134. Eichelberg, K., C.C. Ginocchio, and J.E. Galan, *Molecular and functional characterization of the Salmonella typhimurium invasion genes invB and invC: homology of InvC to the F0F1 ATPase family of proteins.* J Bacteriol, 1994. **176**(15): p. 4501-10.

135. Ehrbar, K., et al., *InvB is required for type III-dependent secretion of SopA in Salmonella enterica serovar Typhimurium.* J Bacteriol, 2004. **186**(4): p. 1215-9.

136. Ginocchio, C.C. and J.E. Galan, *Functional conservation among members of the Salmonella typhimurium InvA family of proteins.* Infect Immun, 1995. **63**(2): p. 729-32.

137. Galan, J.E. and R. Curtiss, 3rd, *Distribution of the invA, -B, -C, and -D genes of Salmonella typhimurium among other Salmonella serovars: invA mutants of Salmonella typhi are deficient for entry into mammalian cells.* Infect Immun, 1991. **59**(9): p. 2901-8.

138. Ginocchio, C.C., et al., *Contact with epithelial cells induces the formation of surface appendages on Salmonella typhimurium.* Cell, 1994. **76**(4): p. 717-24.

139. Ginocchio, C., J. Pace, and J.E. Galan, *Identification and molecular characterization of a Salmonella typhimurium gene involved in triggering the internalization of salmonellae into cultured epithelial cells.* Proc Natl Acad Sci U S A, 1992. **89**(13): p. 5976-80.

140. Kubori, T. and J.E. Galan, *Salmonella type III secretion-associated protein InvE controls translocation of effector proteins into host cells.* J Bacteriol, 2002. **184**(17): p. 4699-708.

141. Eichelberg, K. and J.E. Galan, *Differential regulation of Salmonella typhimurium type III secreted proteins by pathogenicity island 1 (SPI-1)-encoded transcriptional activators InvF and hilA.* Infect Immun, 1999. **67**(8): p. 4099-105.

142. Daefler, S. and M. Russel, *The Salmonella typhimurium InvH protein is an outer membrane lipoprotein required for the proper localization of InvG.* Mol Microbiol, 1998. **28**(6): p. 1367-80.

143. Ideses, D., et al., *A degenerate type III secretion system from septicemic Escherichia coli contributes to pathogenesis.* J Bacteriol, 2005. **187**(23): p. 8164-71.

144. Zhang, L., et al., *Regulators encoded in the Escherichia coli type III secretion system 2 gene cluster influence expression of genes within the locus for enterocyte effacement in enterohemorrhagic E. coli O157:H7.* Infect Immun, 2004. **72**(12): p. 7282-93.

145. Feng, P.C., et al., *Genetic diversity among clonal lineages within Escherichia coli O157:H7 stepwise evolutionary model.* Emerg Infect Dis, 2007. **13**(11): p. 1701-6.

146. Wick, L.M., et al., *Evolution of genomic content in the stepwise emergence of Escherichia coli O157:H7.* J Bacteriol, 2005. **187**(5): p. 1783-91.

147. Garcia-Angulo, V.A., et al., *Regulation of expression and secretion of NleH, a new non-locus of enterocyte effacement-encoded effector in Citrobacter rodentium.* J Bacteriol, 2008. **190**(7): p. 2388-99.

148. Gruenheid, S., et al., *Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic Escherichia coli O157:H7.* Mol Microbiol, 2004. **51**(5): p. 1233-49.

149. Shariff, M., N. Bhandari, and S. Saini, *Fluorescent-actin staining (FAS) test: a novel assay to detect enteropathogenic E. coli.* Indian J Pediatr, 1994. **61**(6): p. 737-8.

150. Jarvis, K.G., et al., *Enteropathogenic Escherichia coli contains a putative type III secretion system necessary for the export of proteins involved in attaching and effacing lesion formation.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7996-8000.

151. Shaikh, N. and P.I. Tarr, *Escherichia coli O157:H7 Shiga toxin-encoding bacteriophages: integrations, excisions, truncations, and evolutionary implications.* J Bacteriol, 2003. **185**(12): p. 3596-605.

152. Hopkins, K.L., R.H. Davies, and E.J. Threlfall, *Mechanisms of quinolone resistance in Escherichia coli and Salmonella: recent developments.* Int J Antimicrob Agents, 2005. **25**(5): p. 358-73.

153. Johnson, T.J., et al., *Plasmid replicon typing of commensal and pathogenic Escherichia coli isolates.* Appl Environ Microbiol, 2007. **73**(6): p. 1976-83.

154. Maeda, S., et al., *Horizontal transfer of nonconjugative plasmids in a colony biofilm of Escherichia coli.* FEMS Microbiol Lett, 2006. **255**(1): p. 115-20.

155. Mills, D.M., V. Bajaj, and C.A. Lee, *A 40 kb chromosomal fragment encoding Salmonella typhimurium invasion genes is absent from the corresponding region of the Escherichia coli K-12 chromosome.* Mol Microbiol, 1995. **15**(4): p. 749-59.

156. Perna, N.T., et al., *Molecular evolution of a pathogenicity island from enterohemorrhagic Escherichia coli O157:H7.* Infect Immun, 1998. **66**(8): p. 3810-7.

157. Manning, S.D., et al., *Variation in virulence among clades of Escherichia coli O157:H7 associated with disease outbreaks.* Proc Natl Acad Sci U S A, 2008. **105**(12): p. 4868-73.

158. Cooley, M., et al., *Incidence and Tracking of Escherichia coli O157:H7 in a Major Produce Production Region in California.* PLoS ONE, 2007. **2**(11): p. e1159.

159. Jay, M.T., et al., *Escherichia coli O157:H7 in feral swine near spinach fields and cattle, central California coast.* Emerg Infect Dis, 2007. **13**(12): p. 1908-11.

160. Uhlich, G.A., et al., *Characterization of Shiga toxin-producing Escherichia coli isolates associated with two multistate food-borne outbreaks that occurred in 2006.* Appl Environ Microbiol, 2008. **74**(4): p. 1268-72.