EPIDEMIOLOGIC AND MUTATIONAL CHARACTERISTICS OF VARIABLE-NUMBER TANDEM REPEATS IN *ESCHERICHIA COLI* O157:H7

by

Anna Christine Noller

BS, Biology, Virginia Polytechnic Institute & State University, 1998

BS, Forestry & Wildlife, Virginia Polytechnic Institute & State University, 1998

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This dissertation was presented

by

Anna C. Noller

It was defended on

October 28, 2004

and approved by

Lee H. Harrison, M.D.
Dissertation Advisor, Professor
Departments of Epidemiology and Medicine
Graduate School of Public Health and School of Medicine
University of Pittsburgh

M. Catherine McEllistrem, M.D., M.S.
Committee Member, Assistant Professor
Departments of Epidemiology and Medicine
Graduate School of Public Health and School of Medicine
University of Pittsburgh

Phalguni Gupta, Ph.D.
Committee Member, Professor
Department of Infectious Diseases and Microbiology
Graduate School of Public Health
University of Pittsburgh

Timothy A. Mietzner, Ph.D.
Committee Member, Associate Professor
Department of Molecular Genetics and Biochemistry
School of Medicine
University of Pittsburgh

Jeffrey G. Lawrence, Ph.D.
Committee Member, Associate Professor
Department of Biological Sciences
Faculty of Arts and Sciences
University of Pittsburgh

Lee H. Harrison, M.D.

EPIDEMIOLOGIC AND MUTATIONAL CHARACTERISTICS OF VARIABLE-NUMBER TANDEM REPEATS IN *ESCHERICHIA COLI* O157:H7

Anna C. Noller, Ph.D.

University of Pittsburgh, 2004

Escherichia coli O157:H7 is an important food-borne pathogen and public health risk that infects thousands of people a year in the United States alone. While many infections may remain undetected, some develop into hemorrhagic colitis and/or hemolytic uremic syndrome especially in young children and the elderly. The development of the molecular subtyping technique pulsed-field gel electrophoresis (PFGE) greatly enhanced the detection of outbreaks caused by this organism, but its technical limitations had researchers searching for alternative techniques. The use of variable-number tandem repeats (VNTRs) for human forensics and subtyping of extremely clonal bacterial species such as *Bacillus anthracis* provided a potential new technique for examining *E. coli* O157:H7. This new technique, multi-locus VNTR analysis (MLVA), examines multiple VNTR loci, which are some of the most rapidly evolving genetic elements in the genome. We demonstrated the utility and superiority of MLVA over PFGE as a molecular subtyping technique for *E. coli* O157:H7. With the establishment of the MLVA protocol, the need arose to understand how often the MLVA loci mutate to help characterize which isolates are highly related. Using an experimental protocol of 10 serial subcultures, one of the 7 MLVA loci was found to be hypervariable with a tendency of single, addition TR mutations. Two other loci were found to be slightly variable while the remaining 4 loci had no mutation events during the experiments. The establishment of a protocol based on VNTRs and the initial understanding of mutational dynamics only touched upon the genotypic roles of VNTRs but not the functional

roles. A preliminary examination of the functional roles of a few selected VNTRs was undertaken by performing a variety of tests. A detailed description of all this project's results is presented in the following work.

ACKNOWLEDGMENTS

I would like to thank my graduate advisor Dr. Lee Harrison for his guidance and support on my project and my development as an independent researcher, as well as my other committee members Drs. Catherine McEllistrem, Phalguni Gupta, Jeffery Lawrence, and Timothy Mieztner. Special thanks go to Dr. Cathy McEllistrem who has served as my unofficial second advisor and has helped me every step of the way. I would like to thank my departmental faculty members and students who all have provided an enormous source of advice and knowledge. I would like to extend a special thank you to Mimi Ghosh, Urvi Parikh, and Betsy Schauer: you all have provided a wonderful support system both in the academic and personal world that graduate school has brought us. Finally, an enormous thanks to my family and friends who through their love and encouragement have helped me to complete this journey and prepare me for my next adventure.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Overview of *Escherichia coli* O157:H7 biology and epidemiology

Escherichia coli O157:H7 is a major cause of foodborne and waterborne illness in the United States and the world. *Escherichia coli* O157:H7 was first recognized in 1982 in association with a food-borne outbreak and is now recognized as causing an estimated 74,000 infections a year (10, 16). Most *E. coli* O157:H7 infections are caused by exposure to bovine fecal contaminated food or water.

E. coli O157:H7 infection occurs when the bacteria enter the intestine and adheres to the epithelium of Peyer's patches (14). This allows for the translocation of shiga toxins and binding to the lining of blood vessels. The intestinal epithelium cells begin to die due to the effect of the toxin and the cells slough off. While many of these cases remain undetectable or mild, a proportion of people develop the characteristic bloody diarrhea and/or hemolytic uremic syndrome (HUS), which may require hospitalization, and in some cases long recuperation periods and permanent disability. Many of those most severely afflicted with this infection are the elderly and the very young. Additionally, *E. coli* O157:H7 is the primary cause of acute renal failure in children (2). In recent years, there have been numerous large outbreaks of *E. coli* O157:H7-related bloody diarrhea and HUS (3, 4, 6).

The disease impact of *E. coli* O157:H7 has created a need for increased preventative food handling techniques and surveillance for outbreaks. In addition to traditional epidemiological

investigations, the major molecular method for surveillance has been pulsed-field gel electrophoresis (PFGE), which is the preferred method of the Centers for Disease Control and Prevention (CDC). The CDC's PulseNet program has created a protocol and centralized database for state public health laboratories to compare their *E. coli* O157:H7 isolates to isolates across the country (18). The PulseNet system and PFGE in general allow epidemiologists to discriminate between isolates and identify those that are potentially involved in an outbreak (1). While there has been great success with PFGE, several factors have researchers searching for an alternative method. This method, while simple, takes several days to complete, and the results can be difficult to compare between labs due to different protocols, reagents, and other parameters. Additionally, the resulting banding patterns of PFGE can be difficult to determine; for example, it can be problematic to tell if a thick band is due to a large amount of DNA at that molecular weight or due to unresolved multiple bands. Finally, we cannot say definitively that 2 bands that migrate to the same molecular weight from 2 different isolates are the same segments of the bacterial chromosome. However, sequence-based methods offer several important potential advantages over PFGE; including shorter assay times and fully comparable and transferable data (9).

1.2. New Sequenced-Based Methods for Outbreak Detection

The advent of faster and more reliable techniques using sequence-based technologies offers a plethora of advantages over pulsed-field gel electrophoresis. With this knowledge, we first examined multi-locus sequence typing (MLST) as a possible subtyping technique for *E. coli* O157:H7. MLST examines the alleles of selected housekeeping genes by nucleotide sequencing a 500 to 600 base pair segment of the gene (9). The sequence data are then analyzed to determine

the genetic relatedness of the bacterial isolates. MLST has been successful for a variety of bacterial species including, *Neisseria meningitidis* and *Streptococcus pneumoniae* (5, 11).

Multi-locus variable-number tandem repeat (VNTR) analysis (MLVA) examines specific tandem repeats (TR) found at a single genetic locus within the genome and then the copy number of each specific tandem repeat from one isolate can be compared to another isolate. VNTRs are among the most rapidly evolving elements in the genome possibly allowing an array of different alleles in different isolates. This variation may be useful for outbreak detection as long as the targeted loci are not under selective pressure, in which the variation would not be neutral and potentially confounding for molecular typing (20). MLVA already has been successful in differentiating *Bacillus anthracis* isolates, one of the most clonal bacterial species (7).

1.3. Mutation rates of tandem repeats

The basis of molecular typing using VNTRs is that these elements can mutate, creating different length alleles at the same VNTR locus. The mechanism by which tandem repeats mutate has been suggested to be the result of slippage and mispairing during DNA replication (17, 20, 21). Multiple factors influence how tandem repeats change; such as which TRs will mutate more quickly than others and if an addition or a loss of a repeat is more likely to occur. As the length of the VNTR increases, the mutation rate increases greatly; resulting in long VNTRs tending to be unstable ensuing in the loss of tandem repeats more frequently than the addition of a TR. This phenomenon results in the rarity of long tandem repeats (8). Additionally, the nucleotide length and composition of the TR can affect the rate of mutation: the shorter the TR unit length is, the higher the mutation rate (19).

The rate of change of these tandem repeats becomes important when trying to analyze a group of isolates to determine their genetic relatedness. The *E. coli* O157:H7 MLVA protocol demonstrated that sporadic isolates could be separated by their very different MLVA types, while the outbreak isolates had identical MLVA types or were single locus variants (12, 13). We need to understand how the 7 MLVA loci change in order to accurately describe the genetic relationship between *E. coli* O157:H7 cases. *E. coli* O157:H7 isolates that are highly related, i.e. an outbreak, will typically have a single MLVA type. However, during DNA replication slippage of the DNA polymerase may result in the loss or addition of a tandem repeat(s), just as we see PFGE patterns change during an outbreak. Certain loci are more likely to change and therefore calling of related isolates must consider the propensity of these loci to change.

1.4. Biological functions of tandem repeats

Some VNTRs may have functional roles in *E. coli* O157:H7. However, a majority of the research on TR functionality has focused on humans. Tandem repeats have been associated with several genetic disorders in humans either by influencing transcription through the promoter region or altering the actual coding region. A 12 nucleotide repeat found within the promoter region for the *EPM1* locus had been linked to epilepsy as the repeat expands beyond the normal allele (15). TRs also have been found within the coding regions that affect normal expression. For example, a 24-bp repeat in the gene for the prion protein has been associated with Creutzfeldt-Jackob disease (15). The normal gene has 5 repeats, while increases of 6-14 repeats are associated with the disease. While less VNTR research has been conducted in bacteria, TRs

are known to influence gene expression at both the transcriptional and translation levels. *Haemophilus influenzae* contains a tetranucleotide unit in the gene *lic*1 resulting in different patterns of LPS expression (20). In *Mycoplasma hyorhinis* two different repeats are involved in antigen variation. One repeat within the coding sequence results in a size variation in membrane surface lipoproteins while the other can turn on or off gene expression of these same genes by variation in the promoter region (20). The limited research for VNTRs in bacteria has been partly due to the lack of fully sequenced bacterial genomes, which now is greatly changing. We have found many tandem repeats by using the 2 fully sequenced *E. coli* O157:H7 genomes. The goal is to focus on a few of these TRs and determine if they vary between isolates and then if the variability has an effect on function.

1.5. Literature Cited

1. **Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm.** 1997. Surveillance for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.

2. **Besser, R. E., P. M. Griffin, and L. Slutsker.** 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease. Annu. Rev. Med. **50:**355–367.

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

4. **Centers for Disease Control and Prevention.** 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers—Western United States, 1992–1993. Morb. Mortal. Wkly. Rep. **42:**258–263.

5. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics **154:**1439–1450.

6. **Izumiya, H., J. Terajima, A. Wada, Y. Inagaki, K.-I. Itoh, K. Tamura, and H. Watanabe.** 1997. Molecular typing of enterohemorrhagic *Escherichia coli* O157:H7 isolates in Japan by using pulsed-field gel electrophoresis. J. Clin. Microbiol. **35:**1675–1680.

7. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.* J. Bacteriol. **182:**2928–2936.

8. **Lai Y. & F. Sun.** 2003. The Relationship Between Microsatellite Slippage Mutation Rate and the Number of Repeat Units. Mol. Biol. Evol. **20:** 2123-2131.

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

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

11. **Nicolas, P., G. Raphenon, M. Guibourdenche, L. Decousset, R. Stor, and A. B. Gaye.** 2000. The 1998 Senegal epidemic of meningitis was due to the clonal expansion of A:4:P1.9, clone III-1, sequence type 5 *Neisseria menin-gitidis* strains. J. Clin. Microbiol. **38:**198–200.

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

13. **Noller, A.C., M.C. McEllistrem, and L.H. Harrison.** 2004. Genotyping Primers for the Fully Automated Multi-Locus Variable-Number Tandem Repeat Analysis of *Escherichia coli* O157:H7. J. Clin. Micr. **42:**3908.

14. **Phillips, A D, Navabpour, S, Hicks, S, Dougan, G, Wallis, T, Frankel, G.** 2000. Enterohaemorrhagic Escherichia coli O157:H7 target Peyer's patches in humans and cause attaching/effacing lesions in both human and bovine intestine. Gut. **47**: 377-381.

15. **Richards, R.I. and G.R. Sutherland.** 1997. Dynamic Mutation: Possible Mechanisms and Significance in Human Disease. TIBS. **22:**432-436.

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

17. **Sharma, R., S. Bhatti, M. Gomez, R.M. Clark, C. Murray, T. Ashizawa, and S.I. Bidichandani.** 2002. The GAA Triplet-Repeat Sequence in Friedreich Ataxia Shows a High Level of Somatic Instability *In Vivo*, with a Significant Predilection for Large Contractions. Hum. Mol. Gen. **11:**2175-2187.

18. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.

19. **Symonds V.V. & A.M. Lloyd.** 2003. An Analysis of Microsatellite Loci in *Arabidopsis thaliana*: Mutational Dynamics and Application. Genetics. **165:** 1475-1488.

20. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequences DNA repeats in Prokaryotic Genomes. Micro. Mol. Biol. Rev. **62:** 275-293.

21. **Viguera, E., D. Canceill, and S.D. Ehrlich.** 2001. Replication Slippage Involves DNA Polymerase Pausing and Dissociation. EMBO. **20:**2587-2595.

2. SPECIFIC AIMS

The overall goal of this project is to identify genetic elements to molecularly subtype *Escherichia coli* O157:H7 and determine how these elements influence this organism's biology. The specific aims of the project are:

1) To develop and validate a PCR-based molecular subtyping technique. *A PCR-based subtyping technique may be superior to the current technique used, pulsed-field gel electrophoresis, to detect outbreak and sporadic cases.* Pulsed-field gel electrophoresis (PFGE) is the current gold standard for E. coli O157:H7 outbreak detection and will be the comparison method for our trial techniques. We studied both multi-locus sequencing typing (MLST), which utilizes sequence variation in housekeeping genes and multi-locus variable-number tandem repeat analysis (MLVA), which uses the variability in tandem repeats as potential assays to replace pulsed-field gel electrophoresis.

2) Determine the rate of change of variable-number tandem repeats. *The hypervariability of certain MLVA loci, especially TR2, will result in highly related strains possibly becoming single or double locus variants.* The establishment of MLVA as a fast, reproducible, sensitive, and automated method for outbreak demands an understanding of the mutational dynamics of each TR locus. A single colony was grown for multiple generations to analyze the number of mutation events that occur at the seven MLVA VNTR loci to determine mutational dynamics.

3) To determine possible functions of several variable-number tandem repeats. *Increasing the number of tandem repeats in the α-helical region of the gene* tol*A will*

decrease the cell's sensitivity to indirect attacks, such as detergent, but will enhance a cell's susceptibility to filamentous phages. The variability of tandem repeats provokes the question of whether these changes are neutral or whether this variability results in changes in protein expression or behavior. First, a locus that contains a VNTR and has a known function needed to be made. Several potential loci were found and a variety of techniques including sequencing, bacterial kill assays, and phage infection assays, were performed to make inferences about the potential roles of selected VNTRs in the *E. coli* O157:H7 genome.

3. RESULTS

3.1. Chapter 1

Published in:

***Journal of Clinical Microbiology* 2003; 41:675-679**

MULTILOCUS SEQUENCE TYPING REVEALS A LACK OF DIVERSITY AMONG *ESCHERICHIA COLI* O157:H7 ISOLATES THAT ARE DISTINCT BY PULSED-FIELD GEL ELECTROPHORESIS

Anna C. Noller[1,2], M. Catherine McEllistrem[1], O. Colin Stine[3], J. Glenn Morris, Jr.[3], David J. Boxrud[4], Bruce Dixon[5], and Lee H. Harrison[1]

Infectious Diseases Epidemiology Research Unit[1] and Department of Infectious Diseases and Microbiology[2], University of Pittsburgh Graduate School of Public Health and School of Medicine, and Allegheny County Health Department[5], Pittsburgh, Pennsylvania; Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, Maryland[3]; and Microbiology Laboratory, Minnesota Department of Health, Minneapolis, Minnesota[4]

3.1.1. Preface

PFGE had been a reliable subtyping technique for identifying *Escherichia coli* O157:H7, but newer sequenced-based techniques provide several advantages over PFGE including unambiguous and faster results. We decided to examine MLST's utility for subtyping O157 and published our results in a peer-reviewed journal.

3.1.2. Abstract

Escherichia coli O157:H7 is a major cause of foodborne illness in the United States. Pulsed-field gel electrophoresis (PFGE) is the molecular epidemiologic method mostly commonly used to identify foodborne outbreaks. Although PFGE is a powerful epidemiologic tool, it has disadvantages that make a DNA sequence-based approach potentially attractive. Multi-locus sequence typing (MLST) analyzes the internal fragments of housekeeping genes to establish genetic relatedness between isolates. We sequenced selected portions of 7 housekeeping genes and 2 membrane protein genes (*omp*A and *esp*A) of 77 isolates that were diverse by PFGE to determine whether there was sufficient sequence variation to be useful as an epidemiologic tool. There was no DNA sequence diversity in the sequenced portions of the 7 housekeeping genes and *esp*A. For *omp*A, all but 5 isolates had the identical sequence as the reference strains. *E. coli* O157:H7 has a striking lack of genetic diversity in the genes we explored, even among isolates that are clearly distinct by PFGE. Other approaches to identify improved molecular subtyping methods for *E. coli* 0157:H7 are needed.

3.1.3. Introduction

Escherichia coli O157:H7 was first recognized in 1982 in association with a food-borne outbreak and is now recognized as an important cause of foodborne illness in the United States, causing an estimated 74,000 infections a year (26, 20). This pathogen causes both bloody diarrhea and hemolytic uremic syndrome (HUS), a severe illness characterized by hemolytic anemia and acute renal failure (2). In recent years, there have been numerous large outbreaks of *E. coli* O157:H7-related bloody diarrhea and HUS (3, 4, 11). Most *E. coli* O157:H7 infections are caused by exposure to food or water that has bovine fecal contamination.

Pulsed-field gel electrophoresis (PFGE) is currently the most widely utilized molecular subtyping method for detecting outbreaks of *E. coli* O157:H7. In fact, PFGE has been found to identify outbreaks of *E. coli* 0157:H7 that were not detected by traditional epidemiologic methods (1). However, DNA sequence-based methods offer several important potential advantages over PFGE; including shorter assay times and fully comparable and transferable data between laboratories (5, 16, 19, 21). Additionally, while PFGE is relatively simple and inexpensive, it is labor intensive, the interpretation of banding patterns can be subjective, and it does not easily handle large sample sets (13).

Multi-locus sequence typing (MLST) is a DNA sequence-based molecular subtyping method that has been used successfully for other bacteria, such as *Neisseria meningitidis, Streptococcus pneumoniae*, and *Salmonella* for both evolutionary and epidemiologic studies (8, 16, 22, 29). Briefly, MLST examines the alleles of selected housekeeping genes by nucleotide sequencing a 500 to 600 base pair segment of the gene. The sequence data are then analyzed to determine the

genetic relatedness of the bacterial isolates. In this study, we performed MLST on a set of *E. coli* O157:H7 isolates that had been characterized epidemiologically and by PFGE to determine the potential utility of MLST for the molecular subtyping of this organism.

3.1.4. Materials and Methods

E. coli O157:H7 Isolates.

E. coli O157:H7 isolates were obtained from several sources for this study. Isolates were selected to include groups of strains from known outbreaks that were indistinguishable by PFGE, groups that were indistinguishable by PFGE but not associated with a known outbreak, and strains with a unique PFGE pattern that were not known to be associated with an outbreak. The Public Health Infectious Disease Laboratory (PHIDL) obtained all *E. coli* O157:H7 strains isolated by the Allegheny County Health Department (ACHD) from 1999 to 2001 (n=59). These strains were not associated with known outbreaks, with the exception of seven isolates from a single restaurant-associated outbreak in August and September 2001. A sample of isolates from the Minnesota Department of Health (MDH) was also included; these were outbreak-associated isolates (n=14) and sporadic isolates (n=4) from 1996 and 1997. ATCC strain EDL933 and the Sakai, Japan, strain RIMD 0509952 were used as reference strains (10, 23).

PFGE.

PFGE analysis was performed according to the Centers for Disease Control and Prevention PulseNet protocol with minor variations (25, 28). Briefly, pure isolates were grown overnight on blood agar. Equal amounts of bacterial suspension, represented by an optical density at 610 nm of 1.3 in 1X TE buffer (10 mM Tris, 1 mM EDTA [pH 8.0]; Sigma, St. Louis, Mo.), 1% SeaKem

Gold agarose (BioWhittaker, Rockland, Maine), and 1% sodium dodecyl sulfate (Sigma) were added to 0.5 mg of proteinase K per ml (Sigma) and mixed to form plugs. The bacteria were lysed within the plugs with a cell lysis buffer (50 mM Tris, 50 mM EDTA [pH 8], 1% Sarcosine, 0.1 mg of proteinase K per ml [Sigma]) and incubated overnight at 37°C. The plugs were then washed four times with 1X TE buffer. Two-millimeter slices of plugs were incubated overnight with either *Xba*I or *Spe*I (New England Biolabs, Beverly, Mass.) at 37°C. The plugs were then loaded onto a 1% SeaKem Gold agarose gel. PFGE was performed with the CHEF III system (Bio-Rad, Hercules, Calif.) with the following run parameters: *Xba*I with a switch time of 3 to 40 s and a run time of 21 h and *Spe*I with switch time of 3 to 20 s and a run time of 21 h. All gels were run with the Centers for Disease Control and Prevention reference strain, G5244, of *E. coli* O157:H7. After the gel had been stained with ethidium bromide, the gel was captured with the Gel Doc 2000 and Multi-Analyst program (Bio-Rad). Dendrograms were created with Molecular Analyst (Bio-Rad) by using the Dice coefficient, unweighted pair group method with arithmetic means (UPGMA), and a position tolerance of 1.3%. Isolates were considered highly related with 0 or 1 band difference with both *Xba*I and *Spe*I.

MLST.

Genomic DNA was isolated with Prepman Ultra according to the manufacturer's instructions (Applied Biosystems, Foster City, Calif.). Seven housekeeping genes were amplified from the genomic DNA by using recombinant *Taq* DNA polymerase (Gibco-Invitrogen, Gaithersburg, Md.); reaction parameters varied depending on the primer set (Table 1). The following genes (coding for the proteins in parentheses) were included: *arcA* (aerobic respiratory control protein) and *aroE* (shikimate dehydrogenase) with primers as described by Reid et al. (24), *dnaE* (DNA

Table 1. The forward and reverse sequences of the primers; based upon sequences found in Genbank or [un]published primers.

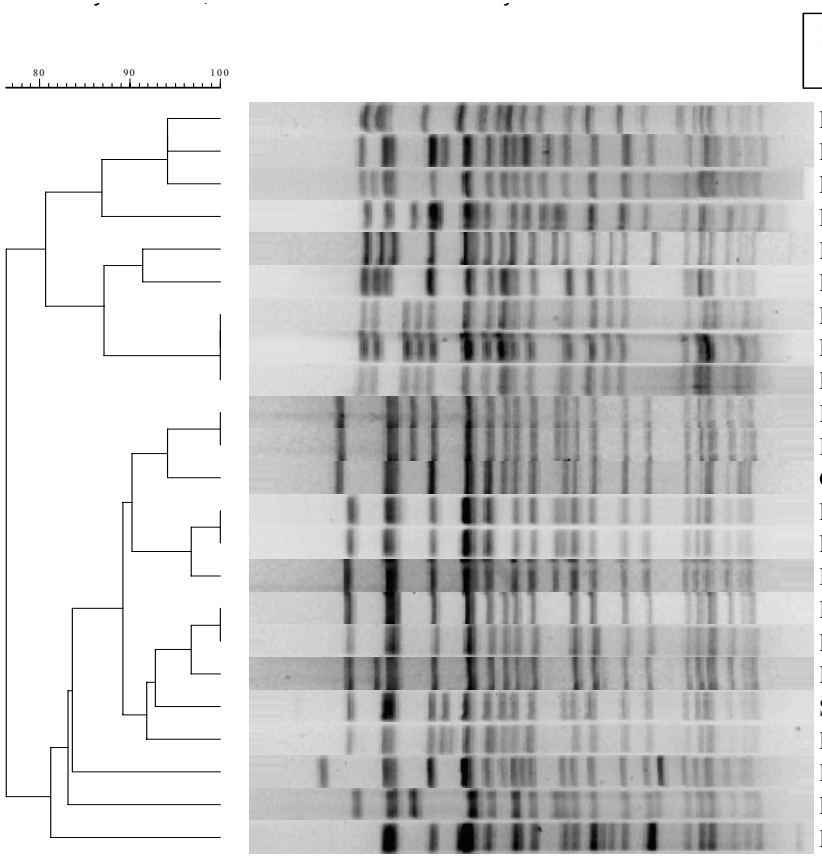
Gene	Primer Sequence	Reaction Parameters[1]	Amplicon Size[2]	Accession # or Reference
*arc*A	F: 5'-GAAGACGAGTTGGTAACACG-3' R: 5'-CTTCCAGATCACCGCAGAAGC-3	95°C (1 m) 55°C (2 m) 72°C (3 m) 30X	680bp	Reid, *et al.* (17)
*aro*E	F: 5'-AAGGTGCGAATGTGACGGTG-3' R: 5'-AACTGGTTCTACGTCAGGCA-3'	95°C (1 m) 57°C (2 m) 72°C (3 m) 28X	620bp	Reid, *et al.* (17)
*dna*E	F: 5'-GA G/T ATGTGTGAGCTGTTTGC-3' R: 5'-CG A/G AT A/C ACCGCTTTCGCCG-3'	94°C (45 s) 45°C (45 s) 72°C (1 m) 30X	550bp	Pallavi Garg, Personal Communication
mdh	F: 5'-CAACTGCCTTCAGGTTCAGAA-3' R: 5'-GCGTTCTGGATGCGTTTGGT-3'	94°C (45 s) 50°C (45 s) 72°C (1 m) 30X	580bp	AE005551 AP002564
gnd	F: 5'-GGCTTTAACTTCATCGGTAC--3' R: 5'-TCGCCGTAGTTCAGATCCCA-3'	94°C (45 s) 50°C (45 s) 72°C (1 m 10 s) 30X	590bp	AE005428 AP002559
*gap*A	F: 5'-GATTACATGGCATACATGCTG-3' R: 5'-CAGACGAACGGTCAGGTCAAC-3'	94°C (45 s) 50°C (45 s) 72°C (1 m 10 s) 30X	535bp	AE005401 AP002558
pgm	F: 5'-CC G/T TC G/C CA G/T AACCCGCC-3' R: 5'-TC A/G AC A/G AACCATTTGAA A/G/T CC-3'	94°C (45 s) 50°C (45 s) 72°C (1 m) 35X	600bp	Kotetishvili, *et al.* (13)
*esp*A	F: 5'-ATGGATACATCAA A/C TG G/C A/C AC-3' R: 5'-TTATTTACCAAGGGATATT-3'	94°C (45 s) 50°C (45 s) 72°C (1 m) 35X	579bp	AE005594 AP002566
*omp*A	F: 5'-AGACAGCTATCGCGATTGC-3' R: 5'-GCTTTGTTGAAGTTGAACAC-3'	94°C (45 s) 50°C (45 s) 72°C (1 m) 30X	691bp	AE005286 AP002554

[1] All reactions had initial denaturation at 94°C (4 m) & final extension at 72°C (4 m)
[2] The sequenced DNA strands were typically shorter than the original PCR amplicon.

polymerase III, α subunit), *mdh* (malate dehydrogenase), *gnd* (6-phosphogluconate dehydrogenase), *gapA* (glyceraldehydes-3-phophate dehydrogenase), and *pgm* (phosphoglucomutase). Also sequenced were the membrane protein coding genes *espA* (*E. coli* secreting protein A) and *ompA* (outer membrane protein A). The oligonucleotide primers were designed based on the published sequences of the genes found in GenBank (Table 1). PCR products were purified with Multiscreen PCR plates (Millipore, Bedford, Mass.). PCR products were sequenced with the Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). Initial denaturation was for 4 min at 94°C, followed by 25 cycles of denaturation at 96°C (30 s), annealing at 50°C (5 s), and extension at 60°C (4 min), with a final extension at 72°C (1 min). The sequencing products were run on an Applied Biosystems 3700 DNA sequencer. Both the forward and reverse strands were sequenced with the PCR primer set (Table 1). Raw sequences were interpreted with Phred (a base-caller program) and Phrap (an assembly program) and verified with Consed (a Unix-based graphical editor) (6, 7, 9). All sequences were aligned and compared by using ClustalX, a graphical multiple alignment program (12). Sequence results were compared to the reference strains from the National Center for Biotechnology Information (NCBI) EDL933 (AE005174) and Sakai RIMD 0509952 (BA000007) by using ClustalX.

3.1.5. Results

A total of 77 *E. coli* O157:H7 isolates were studied: 59 from Pennsylvania and 18 from Minnesota. The PFGE patterns of a selected group of these isolates, chosen to demonstrate the range of diversity of isolates that were sequenced, are shown in Fig. 1. The genetic relatedness of



Isolate #	Isolation Date	County, State
PHIDL 26	5/23/00	Allegheny
PHIDL 34	7/19/00	Allegheny
PHIDL 19	1/24/00	Allegheny
PHIDL 61	9/18/01	Allegheny
E97001249	10/29/96	Stearns, MN
PHIDL 58	8/24/01	Allegheny
PHIDL 35	7/21/00	Allegheny
PHIDL 62	9/13/01	Erie, PA
PHIDL 12	8/18/99	Allegheny
I97000024	6/30/96	Anoka, MN
I97000025	6/29/96	Anoka, MN
G5244	Reference	CDC
PHIDL 57	8/27/01	Allegheny
PHIDL 59	9/4/01	Allegheny
I97001176	9/29/96	Scott, MN
E97001162	10/14/96	Anoka, MN
PHIDL 42	9/17/00	Allegheny
E96001177	10/29/95	Winona, MN
Sakai	1996	Japan
PHIDL 36	8/3/00	Allegheny
EDL 933	1982	Michigan
PHIDL 46	2/1/01	Allegheny
PHIDL 33	7/13/00	Allegheny

Figure 1. **PFGE analysis of selected strains from the Allegheny County Department of Health and Minnesota Department of Health restricted with *Xba*I.** Reference strains include those from; the CDC, G5244; Japan, Sakai RIMD 0509952; and American Type Culture Collection, EDL 933. Isolates are designated with county and state, except for Allegheny County isolates, which are all from Pennsylvania.

these isolates ranged from around 75 to 100% with *Xba*I and around 80 to 100% with *Spe*I (data not shown).

Initially, housekeeping genes were targeted because they have successfully been used for other organisms (8, 22, 29). The seven selected housekeeping genes were chosen for their potential sequence diversity. Three of the genes, *aroE*, *arcA*, and *mdh*, have been used to determine the evolution of pathogenic *E. coli* (24). Two genes, *dnaE* and *pgm*, were chosen because they were found to be informative for *Salmonella* and *Vibrio cholerae* (16; Pallavi Garg, personal communication). The final two housekeeping genes, *gapA* and *gnd*, were chosen because they were transferred into the O157 genome at different evolutionary times. We hoped to find diversity as these genes reached G-C equilibrium with the new host (18). Finally, the two membrane proteins were chosen as being potential targets of the immune system and under possible pressure to mutate.

MLST analysis of the seven housekeeping genes demonstrated that the PHIDL and Minnesota strains had identical sequences at all seven loci. Similar to the housekeeping genes, there were no nucleotide differences in *esp*A. All of the housekeeping genes and *esp*A loci had identical sequences compared to the reference NCBI sequences EDL933 and Sakai RIMD 0509952. For *omp*A, all of the isolates had the same allele as the reference sequences in NCBI, except for five isolates that demonstrated two minor alleles. There was a single nucleotide polymorphism (SNP) (cytosine to thymidine) that was present at base 301 in 4 PHIDL isolates (PHIDL no. 19, 26, 34, and 61). These isolates were clustered together by PFGE, but were not indistinguishable (Fig. 1). Additionally, PHIDL 61 did not remain in the cluster when digested with *Spe*I (data not shown).

PHIDL 62 had an SNP further downstream in *ompA* at nucleotide 560 (guanine to adenine); but the other isolates in its *Xba*I PFGE cluster, PHIDL 12 and 35, did not have this nucleotide polymorphism and instead had the most common allele.

3.1.6. Discussion

In this study, we found a striking lack of DNA sequence diversity for all seven housekeeping genes, with not a single difference in the approximately 311,000 nucleotides (over 4,000 nucleotides per isolate) that were sequenced. Additionally, two other genes which might have been expected to have a higher degree of diversity, *ompA* and *espA*, exhibited either minimal or no diversity, respectively. We had included these genes because we hypothesized that, with products exposed on the surface of the cell, they could be under immune pressure and therefore might exhibit a higher degree of genetic diversity than the housekeeping genes, as was seen in *Neisseria menigitidis* subgroup III (30).

In studies conducted with other bacterial species, strain-to-strain variations in nucleotide sequence are commonly seen, even among strains within a single serotype (8, 19, 27, 29) and PFGE type (16) and/or associated with a common source. The observed sequence conservation could be interpreted as indicative of strong selection, as has been suggested for conserved genotypes of strains of *N. menigitidis* (19). An alternative interpretation—that the strains are clonal due to the organism's recent evolutionary appearance as a recently emerged human pathogen—is consistent with our sequence data. The contrast between the sequence conservation

and presence of diversity as measured by PFGE that we observed could be explained if the PFGE pattern changes resulted from insertions and deletions of DNA that included a restriction site.

Three lines of evidence suggest that an important source of genetic diversity of *E. coli* O157:H7 is based on insertions and deletions of DNA sequences. First, octamer-based genome scanning has revealed distinct lineages of *E. coli* O157 strains. The polymorphic markers that distinguish the lines of descent have been shown to be the result of insertion and deletion of phages and prophages (14, 15). Second, the different banding patterns by PFGE have been shown to result from insertions and deletions containing the *Xba*I restriction sites, not SNPs (17). These deletion/insertion sites all were localized within O157-specific regions (O-islands) of the genome compared to the restriction sites *E. coli* O157:H7 has in common with *E. coli* K-12. Third, an analysis of the differences between the two published *E. coli* O157:H7 genomes indicated substantial differences attributable to insertions and deletions, because the total number of potential protein-encoding genes differs between the genomes by several dozen (10, 23). Additionally, Sakai RIMD 095520 has 1,632 O-island genes that are not found in *E. coli* K-12, while EDL933 has only 1,387 of these genes.

Sequence analysis has multiple advantages over fingerprinting-based methods, including shorter assay time, less subjectivity in interpretation of results, fully transferable data that are comparable among laboratories, and greater ease of automated computer analysis. Our study indicates that the genes we selected for analysis did not have sufficient variation to be useful as an epidemiological tool in *E. coli* O157:H7. Clearly, other approaches to identify informative

regions of the genome will be required to develop improved methods for molecular subtyping of this important pathogen.

3.1.7. Literature Cited

1. **Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm.** 1997. Surveillance for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.

2. **Besser, R. E., P. M. Griffin, and L. Slutsker.** 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease. Annu. Rev. Med. **50:**355–367.

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

4. **Centers for Disease Control and Prevention.** 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers—Western United States, 1992–1993. Morb. Mortal. Wkly. Rep. **42:**258–263.

5. **Enright, M. C., N. P. J. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt.** 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus.* J. Clin. Microbiol. **38:**1008–1015.

6. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.

7. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.

8. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics **154:**1439–1450.

9. **Gordon, D., C. Abajian, and P. Green.** 1998. Consed: a graphical tool for sequence finishing. Genome Res. **8:**195–202.

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

11. **Izumiya, H., J. Terajima, A. Wada, Y. Inagaki, K.-I. Itoh, K. Tamura, and H. Watanabe.** 1997. Molecular typing of enterohemorrhagic *Escherichia coli* O157:H7 isolates in Japan by using pulsed-field gel electrophoresis. J. Clin. Microbiol. **35:**1675–1680.

12. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson.** 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:**403–405.

13. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182:**2928–2936.

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

15. **Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier, and A. K. Benson.** 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, β-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. J. Bacteriol. **183:**6885–6897.

16. **Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulakvelidze.** 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J. Clin. Microbiol. **40:**1626–1635.

17. **Kudva, I. T., P. S. Evans, N. T. Perna, T. J. Barrett, F. M. Ausubel, F. R. Blattner, and S. B. Calderwood.** 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. J. Bacteriol. **184:**1873–1879.

18. **Lawrence, J. G., and H. Ochman.** 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95:**9413–9417.

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

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

21. **Nallapareddy, S. R., R. W. Duh, K. V. Singh, and B. E. Murray.** 2002. Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. J. Clin. Microbiol. **40:**868–876.

22. **Nicolas, P., G. Raphenon, M. Guibourdenche, L. Decousset, R. Stor, andA. B. Gaye.** 2000. The 1998 Senegal epidemic of meningitis was due to the clonal expansion of A:4:P1.9, clone III-1, sequence type 5 *Neisseria menin-gitidis* strains. J. Clin. Microbiol. **38:**198–200.

23. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Ananthara-man, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.

24. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64–67.

25. **Ribot, E. M., C. Fitzgerald, K. Kubota, B. Swaminathan, and T. J. Barrett.** 2001. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. J. Clin. Microbiol. **39:**1889–1894.

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

27. **Stine, O. C., S. Sozhamannan, Q. Gou, S. Zheng, J. G. Morris, and J. A. Johnson.** 2000. Phylogeny of *Vibrio cholerae* based on *recA* sequence. Infect. Immun. **69:**7180–7185.

28. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.

29. **Zhou, J., M. C. Enright, and B. G. Spratt.** 2000. Identification of the major Spanish clones of penicillin-resistant pneumococci via the internet using multilocus sequence typing. J. Clin. Microbiol. **38:**977–986.

30. **Zhu, P., A. van der Ende, D. Falush, N. Brieske, G. Morelli, B. Linz, T. Popovic, I. G. Schuurman, R. A. Adegbola, K. Zurth, S. Gagneux, A. E. Platonov, J. Y. Riou, D. A. Caugant, P. Nicholas, and M. Achtman.** 2001. Fit genotypes and escape variant of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. Proc. Natl. Acad. Sci. USA **98:**5234–5239.

3.2. Chapter 2

Published in

***Journal of Clinical Microbiology* 2003; 41:5389-5397**

MULTI-LOCUS VARIABLE-NUMBER TANDEM REPEAT ANALYSIS DISTINGUISHES OUTBREAK & SPORADIC *ESCHERICHIA COLI* O157:H7 ISOLATES

Anna C. Noller[1,2], M. Catherine McEllistrem, MD, MS[1], Antonio G.F. Pacheco MD, MSc[3], David J. Boxrud, MS[4], and Lee H. Harrison, MD[1]

Infectious Diseases Epidemiology Research Unit, University of Pittsburgh Graduate School of Public Health and School of Medicine[1], and Department of Infectious Diseases and Microbiology[2] and Department of Epidemiology[3], University of Pittsburgh Graduate School of Public Health and School of Medicine, Pittsburgh, Pennsylvania; Departamento de Epidemiologia e Métodos Quantitativos em Saúde, Escola Nacional de Saúde Pública, FIOCRUZ, Rio de Janeiro, Brazil[4]; and Microbiology Laboratory, Minnesota Department of Health, Minneapolis, Minnesota[5]

3.2.1. Preface

MLST showed that *E. coli* O157:H7 was too clonal to target its housekeeping genes, so we needed to concentrate on elements that mutated at a much higher rate. VNTRs were found to differ within the highly clonal species *Bacillus anthracis*; therefore we decided to focus on these elements as a subtyping technique. We published our results in a peer-reviewed journal.

3.2.2. Abstract

Escherichia coli O157:H7 is a major cause of foodborne illness in the United States. Outbreak detection involves traditional epidemiological methods and routine molecular subtyping using pulsed-field gel electrophoresis (PFGE). PFGE is labor intensive, difficult to analyze, and not easily transferable between laboratories. Multi-locus variable-number tandem repeat (VNTR) analysis (MLVA) is a fast, portable method that analyzes multiple VNTR loci, areas of the bacterial genome that evolve quickly. Eighty isolates, including 21 from 5 epidemiologically well-characterized outbreaks from Pennsylvania and Minnesota, were analyzed by PFGE and MLVA. PFGE clusters were defined as strains that differed by ≤1 band using *Xba*I and the confirmatory enzyme, *Spe*I. MLVA was performed by comparing the number of tandem repeats (TRs) at 7 loci. A range of 6-30 alleles was found at the 7 loci resulting in 64 MLVA types among the 80 isolates. MLVA correctly identified all 5 outbreaks if only a single-locus variant was allowed. MLVA differentiated strains with unique PFGE types. Additionally, MLVA discriminated strains within PFGE-defined clusters that were not known to be part of an outbreak. In addition to being a simple and validated method for *E. coli* O157:H7 outbreak detection, MLVA appears to have equal sensitivity and superior specificity compared to PFGE.

3.2.3. Introduction

Escherichia coli O157:H7 has emerged as an important food-borne pathogen infecting thousands of people per year (17). Most *E. coli* O157:H7 infections are caused by exposure to bovine fecal contaminated food or water. The clinical syndromes caused by this organism include bloody diarrhea and hemolytic uremic syndrome (HUS) (4). There have been numerous large foodborne outbreaks of *E. coli* O157:H7-related bloody diarrhea and HUS (1, 5, 6, 21).

The public health impact of *E. coli* O157:H7 has created a need for improved preventative food handling techniques and enhanced surveillance for outbreaks. In addition to traditional epidemiological investigations, pulsed-field gel electrophoresis (PFGE) is used to discriminate between outbreak and sporadic strains (2). Although PFGE has been successful, several factors have led researchers to search for alternative methods. This method, while simple and inexpensive, takes several days to complete, produces results that are suboptimal for interlaboratory comparisons, and can be subjective because it is based on banding patterns (19).

Sequenced-based methods, such as multi-locus sequence typing (MLST), are becoming powerful subtyping tools in molecular epidemiology. These methods have the advantage of being easily standardized and automated. MLST, while successful for other organisms (9, 16, 18, 25) has been unable to discriminate among *E. coli* O157:H7 isolates (19). In one study, no variation was detected in 7 housekeeping genes and little variation was noted in 2 surface protein genes (19).

Given the poor discriminatory power of MLST for *E. coli* O157:H7, we decided to target short tandem repeats (TRs), areas of the bacterial genome that evolve rapidly. Targeting these

elements, which often vary in number among different strains of the same species (the definition of a variable number TR, or VNTR), has successfully been used to discriminate between strains of prokaryotes (24). Multiple-locus VNTR analysis (MLVA) involves determining the number of repeats at multiple loci, thereby providing a powerful tool for assessing the genetic relationships between bacterial strains of the same species. In a study of the highly clonal *Bacillus anthracis*, 426 isolates that were previously homogeneous by other molecular subtyping methods, including PFGE, were separated into 89 distinct genotypes by MLVA (14) (18). MLVA has several advantages over PFGE because, like MLST, the output is highly objective, making the data amenable to automated computer analysis for the rapid detection of outbreaks and easy to compare across laboratories.

The 2 completely sequenced *E. coli* O157:H7 genomes have allowed us to identify many tandem repeats (11, 20). We initially focused on short TRs that varied in the number of times repeated between the 2 reference genomes. We then were able to compare MLVA and PFGE in their ability to detect outbreaks. The highly discriminatory power of PFGE demands that a competing technique be equal, if not superior, in its ability to differentiate between isolates. In this study, we sought to develop a MLVA assay that is useful for detecting outbreaks while being at least as discriminatory and easier to perform compared to PFGE.

3.2.4. Materials and Methods

E. coli O157:H7 Strains.

All *E. coli* O157:H7 strains collected by the Allegheny County Health Department (ACHD) from 1999-2001 were provided to the Public Health Infectious Disease Laboratory (PHIDL) at the

University of Pittsburgh (n=58) (Table 2). These strains were not associated with known outbreaks, with the exception of 7 isolates from a single restaurant-associated outbreak in August and September 2001. Two strains collected from ACHD were Shiga toxin-positive *E. coli* O157:NM (PHIDL isolate numbers 27 and 28). A sample of isolates from the Minnesota Department of Health (MDH) was also included; these were isolates from 4 outbreaks (n=14) and sporadic isolates (n=4) from 1995 and 1996 (Table 2). ATCC strain EDL933 and the Sakai, Japan strain RIMD 0509952 were used as reference strains for MLVA (11, 20), while G5244 from the CDC was used as the reference strain for PFGE. [All of these strains were analyzed previously by our MLST protocol (19)].

Each isolate was classified into 1 of 3 groups. Group 1 isolates were from known outbreaks and were associated with a specific PFGE cluster. Strains with ≥2 band difference by *Xba*I and *Spe*I that were not known to be associated with an outbreak were classified as Group 2. Finally, strains that were ≤1 band different by PFGE but not associated with a known outbreak were classified as Group 3.

PFGE.

PFGE analysis was performed according to the Centers for Disease Control and Prevention PulseNet protocol with minor variations as described previously (19). The bacterial DNA was restricted with *Xba*I or the confirmatory enzyme, *Spe*I (New England Biolabs, Beverly, MA). The switch times for *Xba*I and *Spe*I were 3-40 sec and 3-20 sec, respective, and both ran for 21 hours. Dendrograms were created with Molecular Analyst (BioRad, Hercules, CA) using the

Table 2. Isolate information. Isolates (n= 80) included in this study, including year, state of isolation, and outbreak number.

Group[1]	Year	Location	Outbreak # [2]	Strain I.D.	
1	1995	MN (Daycare)	4	E96001161	E96001162
				E96001177	I96001815
	1996	MN (Daycare)	1	I96003168	I97000025
				I97000024	I97000027
	1996	MN (Daycare)	2	I97001003	I97001180
				I97001017	I97001040
	1996	MN (Daycare)	3	I97000770	
				I97001176	
	2001	PA (Restaurant)	5	PHIDL 51	PHIDL 57
				PHIDL 52	PHIDL 59
				PHIDL 53	PHIDL 60
				PHIDL 54	
2	1995	MN		E96000049	
	1996	MN		E97001162	E97001249
	1999	PA		PHIDL 5	PHIDL 14
				PHIDL 7	PHIDL 15
				PHIDL 11	PHIDL 18
	2000	PA		PHIDL 19	PHIDL 37
				PHIDL 21	PHIDL 38
				PHIDL 25	PHIDL 41
				PHIDL 26	PHIDL 42
				PHIDL 29	PHIDL 43
				PHIDL 33	PHIDL 44
				PHIDL 34	PHIDL 45
				PHIDL 36	
	2001	PA		PHIDL 46	PHIDL 56
				PHIDL 47	PHIDL 58
				PHIDL 48	PHIDL 61
				PHIDL 49	PHIDL 62
				PHIDL 55	
3	1996	MN		E97001568	
	1999	PA		PHIDL 2	PHIDL 10
				PHIDL 3	PHIDL 12
				PHIDL 4	PHIDL 13
				PHIDL 8	PHIDL 16
				PHIDL 9	PHIDL 17
	2000	PA		PHIDL 20	PHIDL 30
				PHIDL 22	PHIDL 31
				PHIDL 23	PHIDL 35
				PHIDL 24	PHIDL 39
				PHIDL 27	PHIDL 40
				PHIDL 28	
	2001	PA		PHIDL 50	PHIDL 63
REFERENCE	1982	MI		EDL933	
	1993	CDC		G5244	
	1996	Japan		Sakai RIMD 0509952	

[1] See text for definition of groups 1-3
[2] Information only for outbreak isolates

Dice coefficient, and a position tolerance of 1.3%. Isolates were classified as belonging to the same PFGE cluster if they had ≤1 band difference with both *Xba*I and *Spe*I.

Potential Variable Number Tandem Repeats.

Over one hundred potential TRs were found in the two fully sequenced *E. coli* O157:H7 genomes, EDL933 (AE005174) and Sakai (BA000007) using the Tandem Repeats Finder software (3). After identifying all TRs that were common to both strains, we chose 6 TRs that were different in number between the two strains. Among TRs that were not variable between the 2 reference genomes, some were found to be variable among the study isolates. For example, because of success with several 6 base pair TRs that were variable between the reference genomes, we tested some that were not variable and found them to be variable among the study isolates, such as TR5.

DNA Isolation & PCR Amplification & Sequencing.

DNA was isolated using the Prepman Ultra Protocol (Applied Biosystems, Foster City, CA). All Allegheny County and Minnesota isolates were analyzed at 7 loci. Primers were based on the sequences from Sakai and EDL933 genomes (11, 19) and were designed using the Primer Finder website (http://eatworms.swmed.edu/~tim/primerfinder/).

Primers were designed for the amplification and sequencing of the targeted repeat region (Table 3) (IDT Inc., Coralville, IA) to verify that the differences seen were due to the variability in the TR region rather than another genetic event (proof of concept primers). Each 30 µL PCR reaction contained 3 µL of 10X PCR buffer, 1.5 mM $MgCl_2$, 0.33 µM of each primer, 25 µM of

Table 3. VNTR loci primers & characteristics. Primers used for the initial amplification and sequencing of the selected tandem repeats for all isolates and characteristics of each tandem repeat locus.

TR Name	Forward Primer Sequence	Reverse Primer Sequence	T_a(ºC)	Tandem Repeat Sequence	Number of Repeats Mimimum	Number of Repeats Maximum	# of Alleles	Diversity[1]	Found in *E. coli* K12	Inside ORF[2]
TR1	ACTGCATGATAAGCCTCAGG	CACTGAAGCCTGTTCCGTTC	57	AAATAG	4	20	12	0.88	No	No
TR2	CGCAGTTGATACCTACGG	GGAAGGAAGCTGATAGGT	53	TGGCTC	7	58	30	0.96	No	Yes
TR3	TCTTGTCAATATAGATTGG	TGATTAAGCGTGTACTGA	50	TATCTT	3	10	8	0.71	No	Yes
TR4	GGTGATGGCTTGATATTGA	GCCACACTGCGAGTATAGAG	53	TGCAAA	2	9	7	0.57	Yes	No
TR5	GTTGATTATCATGGTATGTC	GGACAACTTGTAGTACAAG	51	AAGGTG	6	26	15	0.86	No	Yes
TR6	GATGGTTCGACTAACCGTTAT	TAGCAGATGTTCGTTCCT	53	TTAAATAATCTACAGAAG	7	12	6	0.69	Yes	Yes
TR7	CGCAGTGATCATTATTAGC	TGCTGAAACTGACGACCAGT	50	GACCAC	4	9	6	0.67	Yes	Yes

[1]Diversity based on Nei's marker diversity: 1-Σ(allele frequency)^2, and based on 63 or 64 unique genotypes.
[2]Most of the open reading frames were hypothetical based on either Sakai or EDL933 in NCBI (11,19).

each deoxyribonucleotide, 1.5U of the recombinant Taq DNA polymerase (Invitrogen, Carlsbad, CA), and 1 μL of DNA template. The PCR thermocycling program was identical for the 7reactions, except for the annealing temperatures (Table 3). The samples were placed on a GeneAmp PCR System 9700 (Applied Biosystems) and raised to 94ºC for 4 min, followed by 35 cycles of 94ºC for 45 sec, 50ºC-57ºC for 45 sec, 72ºC for 1 min. The final hold was for 5 min at 72ºC. PCR products were purified using Exo-Sap It (USB Corporation, Cleveland, OH).

Forward and reverse strands of the PCR products were sequenced with an ABI PRISM® 3700 Genetic Analyzer using the Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) following the previously described protocol (19). Contigs were created using Phred and Phrap (7,8). Once the sequences were aligned, the number of repeats was counted using ClustalX (13) or Chromas (Technelysium Pty Ltd.).

Data Analysis.

The unweighted pair group method with arithmetic mean (UPGMA) was used to generate the PFGE and MLVA dendrograms.

The sensitivity and specificity of MLVA for detecting outbreaks were calculated using the pair-wise distances between isolates after being analyzed by UPGMA (also known as cophenetic distances) to determine which cut-point would yield the highest values for both of them. Sensitivity, a measure of the ability to detect outbreaks, was defined as the ability of the MLVA-derived dendrogram to classify a pair of Group 1 isolates as belonging to an outbreak. Specificity, a measure of the discriminatory power for unrelated isolates, was defined as the

ability of a MLVA-derived dendrogram to classify a pair of Group 2 isolates as not belonging to an outbreak.

We observed that the single locus variants that occurred during outbreaks differed by only a single TR. To test the hypothesis that only a single TR difference would likely occur during an outbreak, we determined the likelihood of such a difference to occur between isolates that belong to Group 2. This was achieved by constructing an empirical distribution of the distances in that group after performing logarithmic transformation to account for normality and allow negative values. Using another approach, we compared the mean distance among Group 1 and Group 2 pairs, employing a Student's t-test. All analyses were done with the statistical package R (12).

3.2.5. Results

PCR Amplification & Sequence Analysis of Potential VNTRs.

Initially, a subset of 16 PHIDL isolates was sequenced at 11 loci to determine if the TR locus had sufficient variability (data not shown). If variation existed in this small subset at a particular locus, the remaining isolates were amplified. We found that 7 loci had multiple alleles with substantial variability (Table 3 & Figure 2). The seven primer sets amplified all isolates at all loci with 2 exceptions: isolate E96001161 with the TR2 primers and E97001249 with the TR5 primers. These data were counted as missing for the MLVA analysis. We sequenced the 7 loci of all of our isolates to confirm that the size variations seen in the PCR products were due to the number of TRs. In all cases, the size variation we observed was due to the number of TRs. Rarely, there was sequence variation within the repeat.

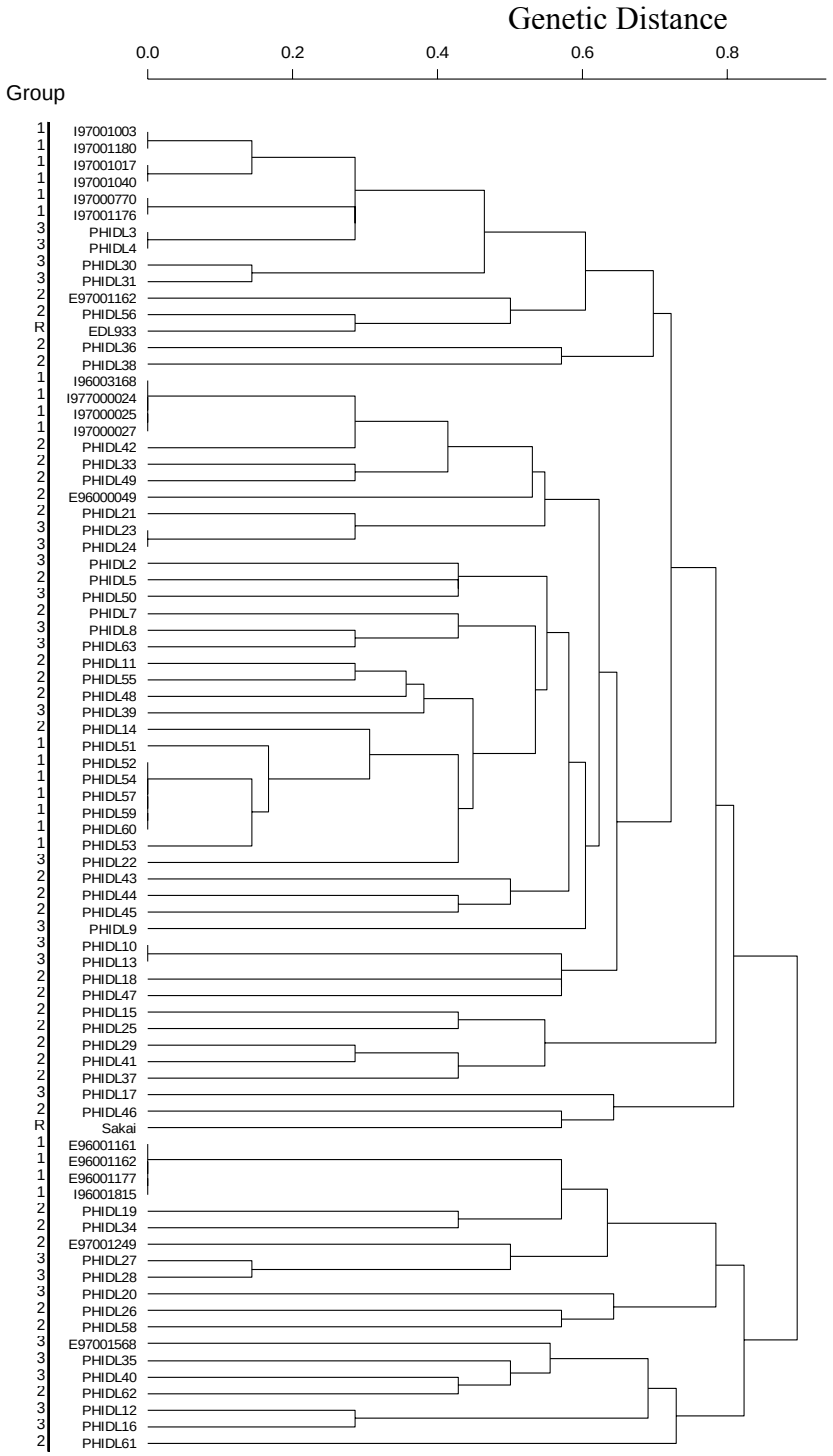


Figure 2. **MVLA dendrogram of PHIDL and MN isolates.** Dendrogram based on the allelic profile of the 80 *E. coli* O157:H7 isolates. See Table 2 for isolate details.

Locus Characteristics.

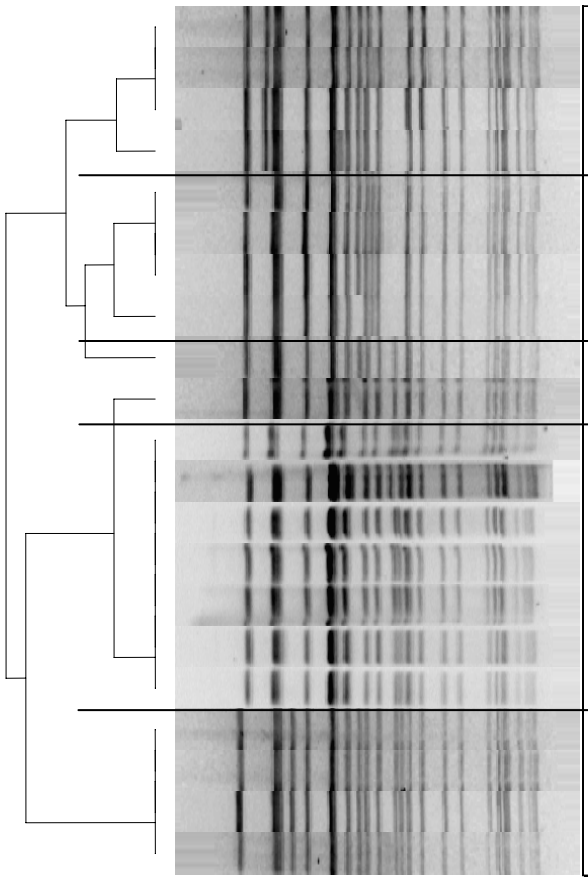
A range of 6 to 30 alleles was found for the seven loci, with VNTRs repeating as few as 2 times at one locus to 58 times at another (Table 3). The diversity for each locus was calculated based on either 63 or 64 unique genotypes; the former was used for TR2 and TR5 because of unsuccessful PCR amplification.

MLVA for outbreak detection.

Group 1 included organisms from 5 separate outbreaks, each associated with a specific PFGE cluster (Figure 3). For outbreak numbers 1, 3, and 4, all isolates had an identical MLVA type. For the remaining 2 outbreaks, there were single locus variants (SLV) that were a result of single TR differences in all instances. In outbreak 2, 2 isolates had 30, rather than 31 repeats at locus TR2. For outbreak number 5, one isolate had 16, rather than 15 repeats at locus TR4 and another isolate had 9, rather than 10 repeats at locus TR6. Allowing SLVs to be considered part of the same outbreak cluster, the sensitivity of MLVA for identifying outbreak strains as such using all 7 loci was 100% (21/21).

The isolates from outbreak 2 differed at 2 loci from the outbreak 3 isolates. The outbreaks involved person-to-person transmission, were separated in time by 2 weeks in September 1996, and occurred in cities that are about approximately 75 miles apart. There was no known epidemiologic connection between the 2 outbreaks.

The probability of a pair of isolates not belonging to an outbreak having at most a 1 TR difference was estimated to be 1.02×10^{-5}, when all 7 loci were taken into account. The difference



Isolate	Outbreak Number	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7
E96001161	4	9	40	7	4	7	11	7
E96001177	4	9	40	7	4	7	11	7
E96001162	4	9	40	7	4	7	11	7
I96001815	4	9	40	7	4	7	11	7
I97001003	2	11	30	7	5	9	11	7
I97001180	2	11	30	7	5	9	11	7
I97001040	2	11	31	7	5	9	11	7
I97001017	2	11	31	7	5	9	11	7
I97000770	3	10	24	7	5	9	11	7
I97001176	3	10	24	7	5	9	11	7
PHIDL60	5	15	25	7	5	11	10	6
PHIDL51	5	15	25	7	5	11	9	6
PHIDL52	5	15	25	7	5	11	10	6
PHIDL53	5	16	25	7	5	11	10	6
PHIDL54	5	15	25	7	5	11	10	6
PHIDL57	5	15	25	7	5	11	10	6
PHIDL59	5	15	25	7	5	11	10	6
I97000024	1	13	19	7	5	11	10	5
I97000025	1	13	19	7	5	11	10	5
I96003168	1	13	19	7	5	11	10	5
I97000027	1	13	19	7	5	11	10	5

Figure 3. **Pulsed-field gel electrophoresis using *Xba*I of all Group 1 isolates, representing 5 outbreaks and corresponding MLVA types.** The numbers under each tandem repeat locus reflect the number of times the TR is found in that isolate. The horizontal lines through the dendrogram and chart are used to visually demarcate the outbreak isolates.

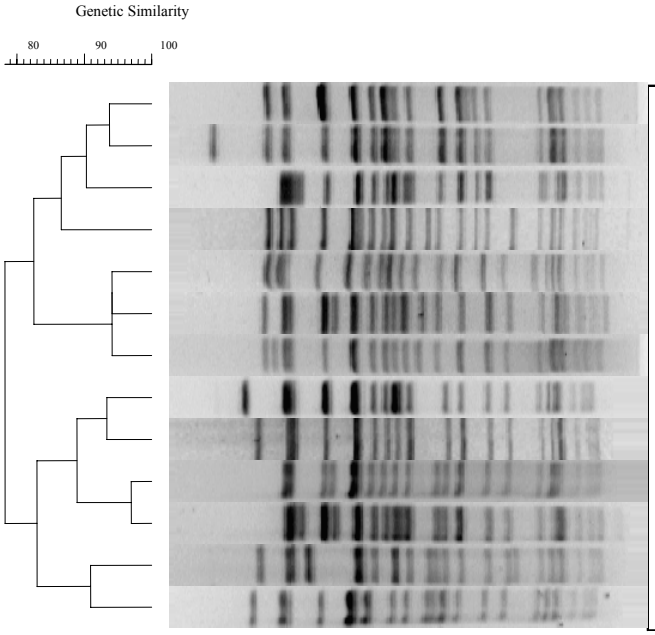
of the average distances in Groups 1 (0.4) and 2 (14.4) was also highly significant ($p < 0.0001$) These data suggest that intra-locus differences that occur during outbreaks occur 1 TR at a time, whereas unrelated isolates are much more likely to differ by more than 1 TR.

MLVA for discriminating isolates from sporadic cases.

Each Group 2 isolate had a unique MLVA type (Figure 4). Additionally, these isolates differed by at least 2 VNTR loci when compared to all other isolates included in this study, for a specificity of 100% (35/35). Discriminatory power was less with all possible combinations of 6 loci. For example, when TR1or TR2 was excluded, PHIDL #14, a Group 2 isolate, was included in outbreak 5 if a single locus difference was allowed. In addition, PHIDL #3 and PHIDL #4, both Group 3 isolates, differed from outbreaks 2 and 3 by 1 locus. When TR7 was excluded, PHIDL #30 and PHIDL #31, Group 3 isolates, differed from outbreak 3 isolates by only a single locus. Similar results were encountered with the exclusion of each of the remaining loci.

MLVA for discriminating strains related by PFGE.

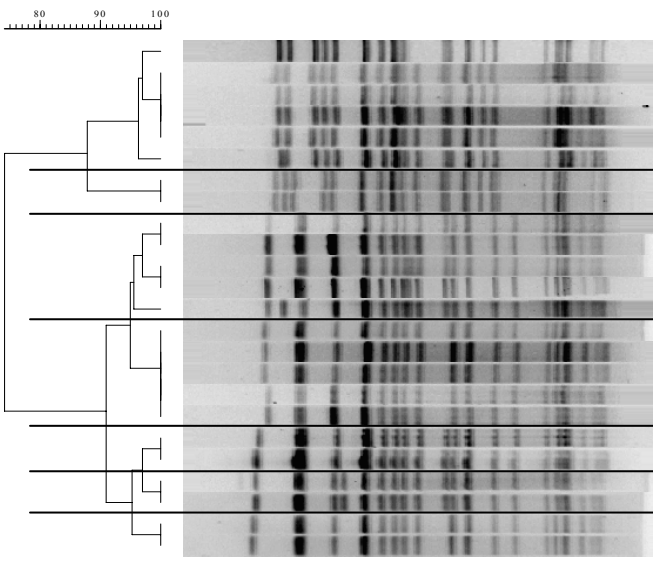
After restriction with *Xba*I, the 24 Group 3 isolates were found to group together in 7 PFGE-based clusters, despite not being part of any identified outbreaks. After restriction with *Spe*I, according to the PulseNet Protocol (22), some of the isolates were further subgrouped (data not shown). Since these strains had not been identified as part of an outbreak, they could not be included in the calculation of sensitivity and specificity. However, the PFGE and MLVA results were compared to provide insights about the relative discriminatory power of these 2 methods using the limited epidemiologic information that was available for these isolates (Figure 5).



Isolate	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7
PHIDL15	17	39	4	5	15	8	7
PHIDL41	15	15	4	4	13	8	6
PHIDL37	12	58	4	5	9	8	6
E97001249	8	20	3	4	No[1]	8	7
PHIDL26	11	29	5	4	12	9	6
PHIDL34	11	15	5	4	7	10	7
PHIDL19	12	18	5	4	7	9	7
PHIDL55	13	28	7	5	10	10	6
E96000049	16	17	7	5	10	10	7
PHIDL43	15	28	9	5	8	11	6
PHIDL47	20	35	7	9	14	10	6
PHIDL46	14	26	8	4	10	10	5
PHIDL60	15	25	7	5	11	10	6

[1]No product for this PCR reaction.

Figure 4. **Pulsed-field gel electrophoresis using *Xba*I of a sample of Group 2 isolates and corresponding MLVA types.** The numbers under each tandem repeat locus reflect the number of times the TR is found in that isolate.



Isolate	Onset Date	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7
E97001568	12/4/1996[1]	10	No[2]	5	3	26	7	8
PHIDL12	8/11/1999	11	14	7	3	21	8	7
PHIDL35	7/21/2000	10	14	5	4	14	8	9
PHIDL16	8/31/1999	11	14	6	3	9	8	7
PHIDL20	3/6/2000[1]	8	23	5	4	18	8	8
PHIDL40	8/7/2000	10	9	5	3	9	8	9
PHIDL27	5/21/2000	9	33	6	4	14	8	7
PHIDL28	5/27/2000	9	35	6	4	14	8	7
PHIDL8	3/13/1999	15	16	8	5	10	10	6
PHIDL23	3/21/2000	17	22	7	5	9	10	5
PHIDL24	3/25/2000	17	22	7	5	9	10	5
PHIDL63	10/5/2001	14	19	8	5	10	10	6
PHIDL17	10/18/1999	15	31	8	5	6	10	4
PHIDL22	3/1/2000	11	11	7	5	11	11	6
PHIDL30	6/27/2000	10	38	7	5	9	10	6
PHIDL31	6/26/2000	10	38	7	5	9	11	6
PHIDL9	5/5/1999	13	18	7	5	8	9	6
PHIDL2	6/21/1999	14	16	7	4	13	10	6
PHIDL39	8/8/2000	15	24	7	7	10	10	6
PHIDL50	6/6/2001	12	29	7	4	10	10	6
PHIDL10	8/6/1999	16	34	7	8	12	10	6
PHIDL13	8/24/1999	16	34	7	8	12	10	6
PHIDL3	7/18/1999	13	10	7	5	9	11	7
PHIDL4	7/17/1999	13	10	7	5	9	11	7

[1]This date reflects the culture date and not the onset of clinical symptoms.
[2]No product for this PCR reaction.

Figure 5. **Pulsed-field gel electrophoresis using *Xba*I of a sample Group 3 isolates and corresponding MLVA types.** The numbers under each tandem repeat locus reflect the number of times the TR is found in that isolate. The horizontal lines through the dendrogram are used to visually demarcate the PFGE-grouped isolates.

The *Xba*I-based cluster containing PHIDL #2, #9, #22, #30, and #31 was subdivided by *Spe*I into 2 clusters, with 1 cluster consisting of the 2 1999 isolates and the second cluster consisting of the 3 2000 isolates, PHIDL #22, #30, and #31. The MLVA provided further discrimination among some of the isolates from 2000. PHIDL #30 and PHIDL #31 had identical MLVA types and these 2 organisms were only isolated one day apart. In contrast, PHIDL #22 differs at 3 loci from #30 and #31 and is separated in time by 3 months from the other 2 isolates.

In addition to PHIDL #30 and #31, other isolates that were clustered by PFGE were also highly related by MLVA. For example, PHIDL # 3 and #4 were identical by MLVA and were isolated in Allegheny County 1 day apart from each other. Taken together with the analysis of the Group 1 isolates, the data suggest that these isolates were part of an unrecognized outbreak.

On the other hand, MLVA also differentiated some Group 3 strains. For example, PHIDL #39 and #50 were different at 3 MLVA loci and were isolated 11 months apart. PHIDL #8 and #23 were also clustered by PFGE (indistinguishable by *Xba*I and 1 band difference by *Spe*I), were detected over a year apart, and differed at 5 MLVA loci. These data suggest that MLVA is able to distinguish among unrelated strains that may be falsely clustered together by PFGE.

The preliminary data suggest that even with a second enzyme, PFGE is unable to differentiate strains as well as MLVA. These data suggest that Group 3 isolates consist of both previously unrecognized *E. coli* 0157:H7 outbreaks and unrelated isolates that PFGE erroneously clustered together. If confirmed in future studies, these data indicate that MLVA is more specific than PFGE for detecting outbreaks caused by this organism.

3.2.6. Discussion

The MLVA assay we developed was highly sensitive in identifying *E. coli* O157:H7 outbreaks while at the same time able to accurately discriminate among sporadic isolates. Using cutoffs of a difference of ≤1 locus with 2 TR differences allowed us to correctly classify all Group 1 and Group 2 isolates, respectively. The data from our Group 1 isolates, consisting of well-characterized outbreaks, suggest that isolates that differ at no more than a single locus are highly related and should be considered to be part of the same outbreak. It was striking that the SLVs we identified among Group 1 isolates all differed by a single repeat, suggesting that SLVs that occur during outbreaks are likely to differ by a small number of repeats (C. Keys, Z. Jay, A. Fleishman, J. Fox, G. Evans, and P. Keim, poster, 103rd General Meeting American Society for Microbiology, Washington, D.C., 2003). Whether all intra-outbreak SLVs will differ by a single repeat remains to be seen. However, the data from our Group 3 isolates suggest that the difference may not always be a single TR because PHIDL numbers 27 and 28, which were likely from a point source, differed at locus TR2 by 2 repeats

Importantly, MLVA was able to distinguish among some Group 3 isolates that appeared to be highly related by PFGE. Based on the comparison of the results for these two assays and the available epidemiologic information, it appears that this group included both sporadic and outbreak-related strains. Thus, MLVA was more discriminatory than PFGE with the group of isolates we studied.

The major implication of this finding is that, if used as part of routine public health surveillance, MLVA may result in fewer false positive signals suggestive of an outbreak. This finding, in

addition to the fact that MLVA has many other advantages over PFGE, suggests that MLVA is superior to PFGE. We are currently automating this process by analyzing fluorescently tagged PCR amplicons of the 7 TR loci on a 3700 DNA analyzer as described by Keim (14). This will eliminate the sequencing step that was described in this experiment and further reduce user intervention, thereby increasing the efficiency of this protocol.

VNTRs are rapidly evolving genomic elements that have been used successfully for the molecular typing of other pathogens such as *Bacillus anthracis, Yersinia pestis,* and *Mycobacterium tuberculosis* (10, 14, 15). One potential concern is that VNTRs evolve so rapidly that multiple MLVA types would emerge during an outbreak initially caused by a single clone. In fact, we observed SLVs in 2 of the 5 outbreaks we studied. This is similar to PFGE, where differences of up to several bands can be observed by PFGE during outbreaks (23). Whether MLVA frequently exhibits a degree of diversity that diminishes its utility for outbreak detection will need to be studied with additional isolates.

We primarily chose relatively short TRs for two reasons. First, shorter repeats may be associated with an increased potential of DNA polymerase slippage resulting in either a loss or gain of a TR (24). Second, shorter repeat sizes may facilitate automation by reducing the potential overlap of different loci during the run on the DNA sequencer. Of the 7 VNTR loci we analyzed, there was a minimum of 6 alleles found at one locus and a maximum of 30, which gives MLVA tremendous discriminatory abilities that are superior to PFGE based on the results for our isolates.

3.2.7. Literature Cited

1. **Bell, B. P., M. Goldoft, P. M. Griffin, M. A. Davis, D. C. Gordon, P. I. Tarr, C. A. Bartleson, J. H. Lewis, T. J. Barrett, J. G. Wells, et al.** 1994. A multistate outbreak of *Escherichia col*i O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington Experience. JAMA **272**:1349–1353.

2. **Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm.** 1997. Surveillance by molecular subtype for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.

3. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.

4. **Besser, R. E., P. M. Griffin, and L. Slutsker.** 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: an emerging infectious disease. Annu. Rev. Med. **50:**355–367.

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

6. **Centers for Disease Control and Prevention.** 2002. Multistate outbreak of *Escherichia coli* O157:H7 infections associated with eating ground beef—United States, June-July 2002. Morb. Mortal. Wkly. Rep. **51:**637–639.

7. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res. **8:**186–194.

8. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res. **8:**175–185.

9. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics **154:**1439–1450.

10. **Frothingham, R., and W. A. Meeker-O'Connell.** 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144**(Pt 5)**:**1189–1196.

11. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori,**

and H. Shinagawa. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8:**11–22.

12. **Ihaka, R., and R. Gentleman.** 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. **5:**299–314.

13. **Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson.** 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:**403–405.

14. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.* J. Bacteriol. **182:**2928–2936.

15. **Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim.** 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pesti*s genome. J. Clin. Microbiol. **39:**3179–3185.

16. **Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulakvelidze.** 2002. Multilocus sequence typing for characterization of clinical and environmental S*almonella* strains. J. Clin. Microbiol. **40**:1626–1635.

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

18. **Nicolas, P., G. Raphenon, M. Guibourdenche, L. Decousset, R. Stor, and A. B. Gaye.** 2000. 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. **38:**198–200.

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

20. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Ananthara-man, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner.** 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature. **409:**529–533.

21. **Samadpour, M., J. Stewart, K. Steingart, C. Addy, J. Louderback, M. McGinn, J. Ellington, and T. Newman.** 2002. Laboratory investigation of an *E. coli* O157:H7

outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington. J. Environ. Health. **64:**16–20, 25, 26.

22. **Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe.** 2001. PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.

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

24. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62:**275–293.

25. **Zhou, J., M. C. Enright, and B. G. Spratt.** 2000. Identification of the major Spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing. J. Clin. Microbiol. **38**:977–986.

3.3. Chapter 3

Published in:

***Journal of Clinical Microbiology* 2004; 42:3908.**

GENOTYPING PRIMERS FOR THE FULLY AUTOMATED MULTI-LOCUS VARIABLE-NUMBER TANDEM REPEAT ANALYSIS OF *ESCHERICHIA COLI* O157:H7

Anna C. Noller[1,2], M. Catherine McEllistrem, MD, MS[1], and Lee H. Harrison, MD[1]

Infectious Diseases Epidemiology Research Unit, University of Pittsburgh Graduate School of Public Health and School of Medicine[1], and Department of Infectious Diseases and Microbiology[2]

3.3.1. Preface

We wrote a data letter to demonstrate the automation of our MLVA protocol and published it in the *Journal of Clinical Microbiology*. We also have additional data that were not published, but enhances the published data and provides additional validation of our MLVA protocol.

3.3.2. Brief Report

We recently developed a multi-locus variable-number tandem repeat (TR) analysis (MLVA) assay for *Escherichia coli* O157:H7 (2). In this assay, we identified 7 loci that, when used in combination, were able to identify *E. coli* O157:H7 outbreaks, discriminate among genetically diverse isolates, and discriminate among isolates that were found to be highly related by pulsed-field gel electrophoresis but not known to be associated with an outbreak.

In our paper, we supplied primers for the 7 loci (2): these primers successfully amplified the appropriate tandem repeat loci, and by sequencing each locus and counting the number of repeats we were able to assign a MLVA type for each isolate. We now have automated the process with new primers, using the approach described for *Bacillus anthracis* (1). Our fluorescently-labeled primers have been designed to give each locus a discrete range on the sequencer, when possible, and color (Table 4).

The PCR reactions were multiplexed to reduce the overall number of reactions needed: TR1, TR5, and TR6 in the first reaction; TR3, TR4, and TR7 in a second reaction; and TR2 alone in a third reaction. All 3 PCR reactions were based on a 30µL reaction volume with 3µL of 10X PCR buffer, 1.5mM MgCl2, 25µM of each deoxyribonucleotide, 1.5U of Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, Calif.), and 1.0 µL of DNA. After primers were added to the reactions, water was added to a final volume of 30µL. The first multiplex reaction contained the following primer concentrations: 0.27µM each of TR1 and TR6, and 0.13µM of TR5. The second multiplex reaction contained 0.17µM each of TR3 and TR4, and 0.2µM of TR7. The single PCR reaction contained 0.33µM of TR2. The samples were placed on a GeneAmp PCR

Table 4. Genotyping primers for automation of MLVA for *E. coli* O157:H7. Table includes fluorescent tags, annealing temperatures, and product ranges in basepairs.

Genotyping Primers (5'-3')				
Locus	Forward Primer Sequence	Reverse Primer Sequence	Annealing Temp (^{o}C)	Genotyping Product Range (bp)
TR1	6Fam-CTCAGGGAAAAGGGAAGACAC	TTTCCTCTGCTGTAAATGTTCG	58	208-304
TR2	Ned-AAGTGATTATCTTTTTCAGCCTCC	GAACAACCCATTTCATTATCTGAT	53	205-515
TR3	Vic-CAGTTGCTCGGTTTTAACATTG	CGACAAGATGATAATGAAAGCG	56	85-128
TR4	6Fam-AGGAGGGTGATGAGCGGTTA	CATTATTCCCATTTCTGCCTG	56	147-183
TR5	Vic-ACTTTGGAGTGAGGGCTGCTA	ATAACCGATACTGAGCTGTCGC	58	318-405
TR6	6Fam-GACTCGTCAAGGACATACAGCC	CAGATGTTCGTTCCTTATTGCTAA	58	395-483
TR7	Vic-ACGCAACTGGCTGGAGAATA	TGACCTCTCTACCATCAAAGCG	56	186-216

System 9700 (Applied Biosystems) and the temperature was raised to 94°C for 4 min, followed by 32 cycles of 94°C for 45 s, 53 to 58°C for 45 s, and 72°C for 1 min. The final hold was for 5 min at 72°C. The 3 PCR reaction products were pooled so that each isolate had its 7 loci analyzed on one lane on the ABI PRISM 3700 Genetic Analyzer (Applied Biosystems).

Our genotyping results correlated with our previous results with one exception (2). PHIDL 18 (1/80 isolates, 1.25 %) could not be genotyped after repeated attempts, even though the PCR was successful. Moreover, sequencing of 8 isolates demonstrated that the genotyping primers were amplifying the tandem repeats of interest.

Due to migration differences of the ladder and products, the genotyping size differed from the expected size based on the length of the flanking region and TR (Appendix A). For all isolates, the observed size for each locus was reproducible from at least four independent runs. With the verification of these new primers, our MLVA protocol provides an automated, reproducible, highly discriminatory method for detected outbreaks caused by *E. coli* O157:H7.

3.3.3. Additional Introduction

The creation of our MLVA protocol and subsequent automation of this procedure offers a rapid, reproducible, and almost completely objective analysis of *E. coli* O157:H7 isolates and their potential to be part of outbreaks. The previous investigation into *E. coli* O157:H7 MLVA used 80 clinical isolates from restricted geographical regions: Allegheny County, Pennsylvania and Minnesota. The following, additional results demonstrate the ability of our MLVA protocol to

classify 100 isolates form across the United States including both known outbreak isolates and known sporadics.

3.3.4. Additional Methods & Materials

E. coli O157:H7 Strains.

One hundred isolates from around the United States were provided by the Centers for Disease Control & Prevention. A combination of sporadic and outbreak isolates were sent without epidemiological data. Once PFGE and MLVA had been completed, the epidemiological information was obtained. Twelve outbreaks with a total of 22 isolates were included, while the remaining isolates were either sporadic isolates or were isolates that were genetically and temporally related but were unconfirmed outbreaks.

Pulsed-Field Gel Electrophoresis.

Isolates were examined using PFGE as previously described (2). PFGE and MLVA genotyping were performed by different individuals to not influence the analysis of the results from both techniques. Briefly, pure isolates were grown overnight on blood agar. Equal amounts of bacterial suspension were added to 1X TE buffer (Sigma), 1% SeaKem Gold Agarose (BioWhittaker), 1% sodium dodecyl sulfate (Sigma), and proteinase K (Sigma). Suspensed cells were lysed overnight with cell lysis buffer and washed four times with 1X TE buffer. Two-millimeter slices of plugs were incubated overnight with either *Xba*I or *Spe*I (New England Biolabs). PFGE was performed with the CHEF III system (Bio-Rad). After the gel had been stained with ethidium bromide, the gel was captured with the Gel Doc 2000 and the Quantity One program (Bio-Rad). Dendrograms were created with the program, Bionumerics (Bio-Rad)

by using the Dice coefficient, unweighted pair group method with arithmetic means (UPGMA). Isolates were considered highly related with 0 or 1 band difference with both *Xba*I and *Spe*I.

DNA Isolation.

DNA from the *E. coli* O157:H7 isolates was isolated using the Prepman Ultra technique (Applied Biosystems) as described in Chapter 1.

MLVA Genotyping.

All isolates were examined at the 7 MLVA loci using the primer sets and PCR reactions as described in the Brief Report above. More specifically, 5uL of each of the 3 PCR reactions were run on a 3% high-resolution gel (Sigma) to examine the intensity of the bands. From those results, the amount of each sample to be added to the pooled sample was determined. The pooled samples from all 100 isolates were run on a 3700 DNA Analyzer using the Rox® Ladder (Applied Biosystems) as a size standard. The software package Genescan® (Applied Biosystems) algorithm automatically identifies and sizes each peak to the Rox® Ladder (Applied Biosystems), and provides peak area and peak height information. Genotyper® software (Applied Biosystems) was used to call and catalog alleles for all 7 loci and build tables automatically. Allele sizes for each of the 7 loci were converted to number of tandem repeats by using an equation specific for each MLVA locus:

$$\text{\# of Tandem Repeats} = \frac{\text{(Genotyper Peak Size [bp] – Constant Size of Region Flanking TR[bp]*)}}{\text{\# of Nucleotides in Repeat (bp)}}$$

*Constant size is different for each of the 7 loci.

Each isolate had a 14 digit MLVA type which could then be compared to one other to determine relatedness.

3.3.5. Additional Results

PFGE Results.

Of the 100 CDC isolates, 96 of the isolates were successfully digested with both *Xba*I and *Spe*I while the remaining 4 isolates had degraded DNA despite several attempts. During the primary analysis of the PFGE data, we were blinded regarding the epidemiologic information for each isolate. Therefore, we classified our isolates into 2 groups: highly related isolates, defined as $\leq$1 band different with both restriction enzymes and sporadic isolates, defined as $\geq$2 bands different to their closest neighbor (Figures 6 & 7). Fifty-eight of the 100 isolates were clustered in the highly related group into 19 groups and 2 subgroups (labeled PFGE-based groups A-S).

MLVA Results.

Of the 100 CDC isolates, 98 of the isolates had all 7 alleles and the other 2 isolates had results for 6 of the 7 loci (Figures 6 & 7). Isolates were suspected to be highly related with identical MLVA types or single locus variants (or double locus variants in particular situations). Forty-one of the isolates were characterized as highly related by MLVA alone.

Comparison of PFGE and MLVA with the Inclusion of Epidemiological Data.

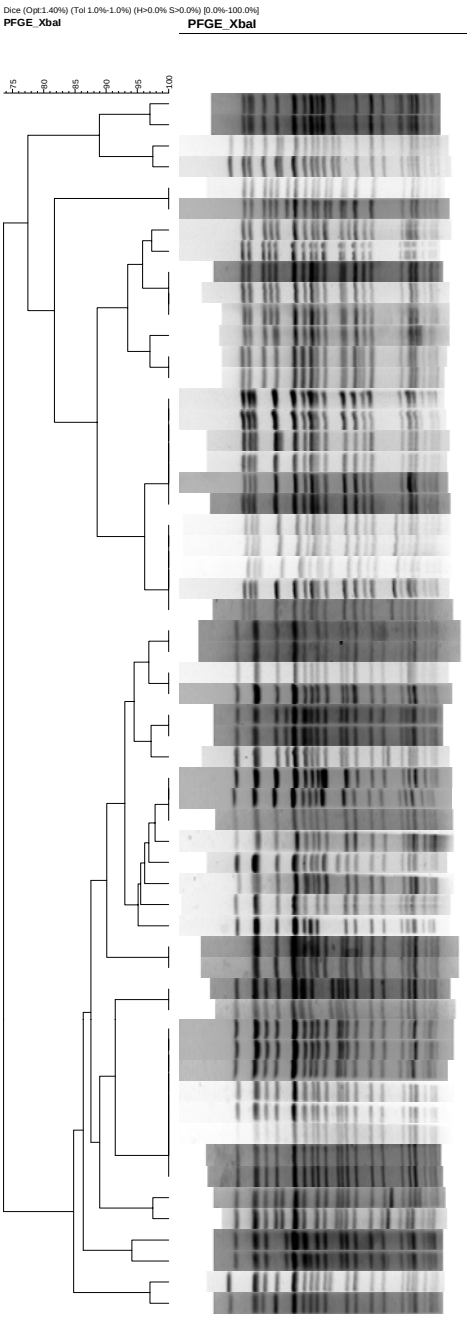
Fifteen of the 19 groups that were deemed highly related by individual methods were shown to remain so with the combination of PFGE, MLVA, and epidemiological information. There were multiple examples of isolates that were grouped together due to their PFGE patterns, but were separated by the MLVA and epidemiological data. For example, PFGE-based group H1 had 6 isolates in which the PFGE patterns were identical with both enzymes. The MLVA data

demonstrated that only 2 isolates (9082 & 9220) had identical MLVA types, while the remaining 4 isolates had 3 to 4 different loci (Figure 6).

In the group defined as sporadic by our PFGE definition, 29 out of 42 isolates were confirmed as sporadic individually and together with the PFGE, MLVA, and epidemiological information. The majority of the remaining 13 isolates had the potential to be highly related and were deemed sporadic due to our very strict definition of highly related versus sporadic isolates. For example, isolates 8945 and 8948 were categorized as sporadic because while identical under *Xba*I (Figure 7), the isolates had a 2-band difference under *Spe*I (Data not shown). These 2 isolates had an identical MLVA type and epidemiologically appeared to be highly related because they were both from Oregon and were isolated 2 days from each other. A third isolate, 8947, also had an identical MLVA type and epidemiological data to suggest being part of the same group, but the PFGE data placed this isolate even further outside the definition of highly related.

Four of the sporadic isolates were untypeable by PFGE, so we automatically assigned them to the sporadic group because we initially had no additional information. The MLVA data suggested that 2 of the isolates were indeed sporadic, but the other 2 isolates (G5289 & G5290) had the identical MLVA type and were confirmed as part of a known outbreak by the epidemiological information.

Mixed between the highly related and sporadic groups were 10 outbreak groups and 2 individual outbreak isolates. Eight of the outbreaks were correctly assigned to the highly related group by the PFGE and MLVA data. The remaining 4 isolates were assigned to the sporadic group. Two of these isolates were described previously, as they were untypeable by PFGE. The other 2

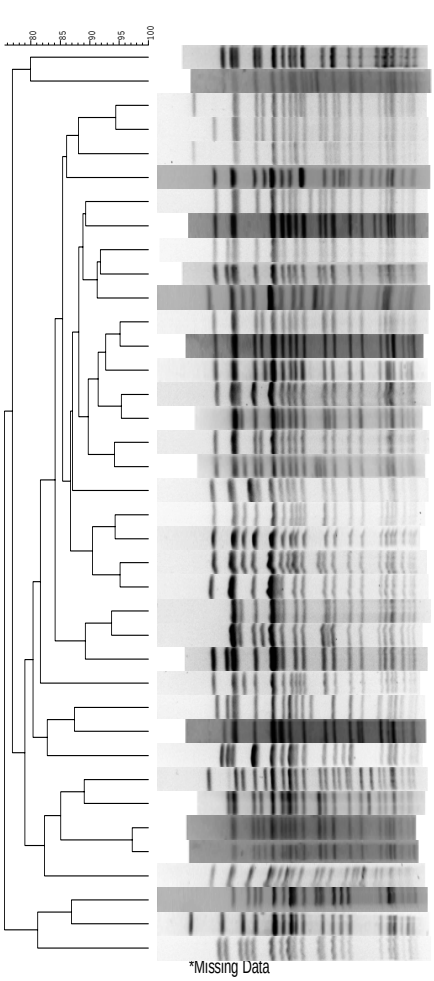


	PFGE-Based Groups	MLVA Results TR1	TR2	TR3	TR4	TR5	TR6	TR7	Date	State**	County
CDC G5295	A	11	20	5	3	9	9	6	1988	MN	
CDC G5296	A	11	20	5	3	9	9	6	1988	MN	
CDC 6905	B	10	35	5	4	10	9	7	1/7/00	CT	NEW HAVEN
CDC 7447	B	10	36	5	4	10	9	7	1/19/00	OR	CLATSOP
CDC 9051	C	9	12	3	4	7	9	5	9/14/00	MD	ANNE ARUNDEL
CDC 9090	C	13	13	3	4	7	9	5	11/20/00	CT	NEW HAVEN
CDC 7478	D	9	29	6	4	M*	8	8	3/21/00	OR	UMATILLA
CDC 8949	D	9	28	6	4	22	8	8	8/31/00	OR	CROOK
CDC G5307	E	10	9	5	3	19	8	9	1992	ME	
CDC 8993	E	10	13	5	3	18	8	10	10/7/00	NX	SUFFOLK
CDC G5284	F	7	27	5	3	24	8	9	1986	NC	
CDC G5283	F	7	27	5	3	24	8	9	1986	NC	
CDC G5324	G	9	34	5	4	8	8	6	1984	NC	
CDC G5325	G	9	34	5	4	8	8	6	1984	NC	
CDC 8960	H1	16	36	4	4	14	8	7	10/11/00	KS	STAFFORD
CDC 8967	H1	11	27	4	4	14	8	8	10/12/00	MN	DOUGLAS
CDC 9038	H1	15	18	4	5	13	8	7	4/17/00	MD	BALTIMORE
CDC 9040	H1	14	34	4	5	13	8	7	6/29/00	MD	BALTIMORE
CDC 9082	H1	15	14	4	4	13	8	6	10/13/00	GA	WALKER
CDC 9220	H1	15	14	4	4	13	8	6	11/5/00	NX	SUFFOLK
CDC 7441	H2	15	23	4	5	14	8	7	1/9/00	NX	WESTCHESTER
CDC 7442	H2	15	23	4	5	14	8	7	1/11/00	NX	WESTCHESTER
CDC 7443	H2	15	23	4	5	14	8	7	1/14/00	NX	WESTCHESTER
CDC 7445	H2	15	23	4	5	14	8	7	1/22/00	NX	WESTCHESTER
CDC 9228	H2	15	18	4	4	17	8	8	12/1/00	CO	DENVER
CDC G5287	I	13	11	7	4	13	10	6	1986	WA	
CDC G5288	I	13	10	7	4	13	10	6	1986	WA	
CDC 6906	J	11	6	8	5	11	10	6	1/17/00	CT	LITCHFIELD
CDC 9107	J	4	20	8	7	8	12	6	9/21/00	WA	CLARK
CDC G5301	K	12	21	8	3	11	10	6	1991	MT	
CDC G5302	K	12	21	8	3	11	10	6	1991	MT	
CDC 8994	K	9	33	6	3	14	9	7	10/9/00	NX	SARATOGA
CDC 9102	L	10	36	7	5	3	11	6	8/25/00	WA	KING
CDC 9104	L	10	36	7	5	3	11	6	9/12/00	WA	SNOHOMISH
CDC 9239	L	13	23	6	5	9	11	8	11/28/00	NX	ONTARIO
CDC 9166	M	13	39	9	6	8	11	8	11/6/00	NY	NEW YORK
CDC 9039	L	7	20	7	3	9	10	4	6/16/00	MD	BALTIMORE
CDC 9180	L	12	13	7	5	10	10	4	11/13/00	MN	HENNEPIN
CDC 7416	M	13	16	7	5	9	11	8	2/21/00	TN	DAVIDSON
CDC 9125	M	9	20	8	5	10	11	4	9/28/00	CA	ALAMEDA
CDC G5291	N	16	23	4	6	12	10	8	1987	UT	
CDC G5292	N	16	23	4	6	12	10	8	1987	UT	
CDC 9221	O	10	14	6	5	8	10	6	11/21/00	NX	ONONDAGA
CDC 9240	O	10	14	6	5	8	10	6	12/8/00	NX	MONROE
CDC 9105	P1	14	41	6	5	14	10	6	9/17/00	WA	KING
CDC 9106	P1	14	41	6	5	14	10	6	9/15/00	WA	KING
CDC 9109	P1	14	41	6	5	14	10	6	10/15/00	WA	KING
CDC 9181	P2	12	14	7	5	10	10	4	11/16/00	MN	ANOKA
CDC 9182	P2	12	13	7	5	10	10	4	11/20/00	MN	HENNEPIN
CDC 9183	P2	12	13	7	5	10	10	4	11/23/00	MN	DAKOTA
CDC 9184	P2	12	13	7	5	10	10	4	11/26/00	MN	RAMSEY
CDC 9185	P2	12	13	7	5	10	10	4	11/29/00	MN	RAMSEY
CDC G5313	Q	13	17	7	5	11	10	4	1993	OR	
CDC G5314	Q	14	18	7	5	11	10	4	1993	OR	
CDC G5299	R	4	42	6	5	11	10	6	1990	ID	
CDC G5300	R	4	42	6	5	11	10	6	1990	ID	
CDC 9050	S	12	40	6	5	8	11	8	8/15/00	MD	HOWARD
CDC G5309	S	12	18	7	5	10	10	6	1993	WA	

*Missing Data

**NX = New York State, NY = New York City

Figure 6. PFGE, MLVA, and epidemiological data of the highly related isolates. The dendrogram is based on both the *Xba*I and *Spe*I data, but only the *Xba*I patterns are seen. The corresponding MLVA types and any known epidemiological information are also presented.



	MLVA Results										
	TR1	TR2	TR3	TR4	TR5	TR6	TR7	Date	State**	County	Known Outbreak
CDC 9002	4	20	7	4	6	6	7	10/22/00	FL	ORANGE	
CDC 9238	18	34	4	3	6	8	6	11/26/00	NX	ONEIDA	
CDC 7327	16	29	7	6	9	11	6	1/1/00	MN	SCOTT	
CDC 7329	16	29	7	6	9	11	6	1/20/00	MN	DAKOTA	
CDC 7328	12	24	6	5	10	9	6	1/13/00	MN	BLUE EARTH	
CDC 9103	16	11	8	4	11	10	4	8/31/00	WA	KING	
CDC 7523	11	16	6	5	8	10	6	3/3/00	MA	WORCESTER	
CDC 9219	16	20	9	5	10	11	6	10/29/00	NX	SARATOGA	
CDC 7440	14	23	6	6	11	10	6	1/1/00	NX	HERKIMER	
CDC 9003	5	29	7	6	13	10	6	10/22/00	FL	LEE	
CDC 9108	12	21	7	5	11	10	5	10/8/00	WA	SPOKANE	
CDC 7346	10	34	8	7	8	11	7	2/20/00	MN	CROW WING	
CDC 9197	10	27	8	5	11	10	4	11/3/00	CA	ALAMEDA	
CDC 9134	9	20	10	5	11	10	4	11/22/00	CO	EL PASO	
CDC 7476	12	M*	9	5	11	10	6	1/26/00	OR	WASHINGTON	
CDC 9248	10	20	7	6	8	11	7	10/1/00	WV	MERCER	
CDC 7522	20	33	7	10	14	10	6	1/10/00	MA	MIDDLESEX	
CDC 9249	13	22	8	5	18	10	6	10/19/00	WV	OUT OF STATE	
CDC 8991	22	30	8	8	14	10	6	9/27/00	NX	NIAGRA	
CDC 8945	10	9	8	5	9	12	6	8/20/00	OR	MARION	
CDC 8948	10	9	8	5	9	12	6	8/22/00	OR	LINCOLN	
CDC 8968	13	8	8	4	12	10	6	10/24/00	MN	ANOKA	
CDC 8990	13	30	7	5	8	12	8	9/23/00	NX	ONONDAGA	
CDC 7477	10	30	9	5	9	10	6	2/18/00	OR	CLATSOP	
CDC 8950	11	33	9	5	9	10	6	9/9/00	OR	UMATILLA	
CDC 9037	13	15	7	5	8	11	7	4/13/00	MD	HARFORD	
CDC 8947	10	9	8	5	9	12	6	8/24/00	OR	MARION	
CDC 8944	9	33	6	3	14	9	7	8/15/00	OR	COOS	
CDC 9218	17	21	6	5	10	10	6	10/25/00	NX	PUTNAM	
CDC 8946	14	35	4	5	13	8	7	8/19/00	OR	DOUGLAS	
CDC 7446	16	11	8	4	11	10	4	1/21/00	NX		
CDC 9250	9	22	9	5	7	11	5	11/7/00	WV	KANAWHA	
CDC G5293	10	14	8	5	8	11	7	1988	WI		college outbreak
CDC G5294	10	14	8	5	8	11	7	1988	WI		college outbreak
CDC 7444	13	44	12	5	9	12	4	1/18/00	NX	ALBANY	
CDC 9089	11	40	6	3	9	7	6	10/19/00	CT	FAIRFIELD	
CDC 9144	10	13	5	3	18	8	10	9/20/00	NY	KINGS	
CDC 8992	10	14	5	3	18	8	10	10/5/00	NX	SCHENECTADY	
CDC 9004	5	29	6	4	9	10	8	1/9/00	MD	BALTIMORE	
CDC 9167	14	8	9	3	15	10	8	11/13/00	NY	KINGS	
CDC G5289	11	29	6	3	12	10	8	1986	WA		Walla Walla outbreak
CDC G5290	11	29	6	3	12	10	8	1986	WA		Walla Walla outbreak

*Missing Data

**NX = New York State, NY = New York City

Figure 7. PFGE, MLVA, and epidemiological data on the sporadic isolates. The dendrogram is based on both the *Xba*I and *Spe*I data, but only the *Xba*I patterns are seen. The corresponding MLVA types and any known epidemiological information is also presented. The highlighted colors represent identical or SLV MLVA types.

isolates were part of a known college outbreak and had identical MLVA types (Figure 7). Again, they were classified as sporadic because of our strict PFGE criteria.

3.3.6. Discussion

The goal of evaluating the 100 CDC isolates was to further validate our MLVA genotyping protocol as explained in the Brief Report. These isolates represented a diverse collection of sporadic and outbreak cases from across the country collected over 16 years. These isolates would truly test if our MLVA protocol was sensitive and specific enough to replace the current molecular subtyping technique, PFGE.

Overall, MLVA and PFGE results correlated with most of the isolates determined to be highly related by PFGE also being highly related by MLVA. Similar concurrent PFGE and MLVA results are seen for the sporadic isolates. Where the 2 techniques differ is where MLVA is able to more precisely separate isolates that cluster by PFGE but that are not related. This can be seen in examples in which PFGE has identical patterns (ex. 9050 & G5309), but the MLVA types differ at multiple loci, in this example 5 loci, and the epidemiological information confirms that the isolates are not related. More important are those isolates that are not clustered epidemiologically but have identical PFGE patterns. This could set off a potentially expensive investigation. Some of these proposed outbreaks are in fact not related events, but instead are due to PFGE's inability to discriminate. MLVA is able to identify that these groups are in fact not part of an outbreak. For example, PFGE-based group H2 had 5 isolates in it (Figure 6), with 4 isolates from the New York area and 1 from Denver, Colorado. The identical PFGE patterns

could trigger suspicions of a nationwide outbreak, but the MLVA data clearly show that the Colorado isolate is not part of the outbreak as it differs at 4 MLVA loci.

In the last few chapters, we have shown the superiority of MLVA as a subtyping technique. Not only does it discriminate between outbreak and sporadic isolates as well as PFGE, but it is able to discriminate isolates that PFGE is unable to differentiate. Our MLVA genotyping technique also surpasses PFGE in its high reproducibility, faster protocol times, objective data output, and automated analysis. MLVA may represent the next generation molecular subtyping technique for *E. coli* O157:H7.

3.3.7. Literature Cited

1. **Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis.* J. Bacteriol. **182:**2928–2936.

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

3.4. Chapter 4

Evaluation of Multi-Locus Variable Number Tandem Repeat Analysis for non-O157 *Escherichia coli*

3.4.1. Preface

With the establishment of the MLVA protocol for *Escherichia coli* O157:H7,we wanted to determine whether our assay could be used to detect outbreaks of non-O157 enterohemorrhagic *E. coli*s (EHECs). Non-O157s EHECs represent an underreported portion of gastrointestinal disease in the United States. The following displays the results we obtained looking at a wide array of EHECs from Baltimore, Maryland and São Paulo, Brazil. These results have not been published but instead are the framework for the continuing investigation into MLVA of non-O157 EHECs.

3.4.2. Introduction

Escherichia coli O157:H7 is responsible for over 74,000 illnesses a year in the United States alone, but other serotypes do occur in the U.S. (5). Worldwide, non-O157 strains continue to be an increasingly important cause of hemorrhagic colitis. Unfortunately, most screening has been for *E. coli* O157:H7 resulting in an underreporting of these non-O157 strains. The enterohemorrhagic *E. coli* (EHEC) serogroup O111 is becoming an increasingly important cause of human illness, as it is second only to O157 (1). In 1999, 55 individuals were infected with *E. coli* O111:H8 while attending a high school cheering camp in Texas; a 16-year and 19-year old developed hemolytic-uremic syndrome (2). Examples of other outbreaks caused by non-O157 strains include an outbreak in 1994 where 11 confirmed cases and 7 suspected case-patients were identified in an outbreak of *E. coli* O104:H21 in association with the consumption of milk in Montana (3). Additionally, other important serotypes include O111:H2 in Germany, O103:H2 in France, and O145:H5 in Japan; but multiple serotypes have global distribution including O5:H-, O26:H11, O91:H-, O113:H21, O116:H-, O123:H-, and O128:H2 (1).

Currently pulsed-field gel elecrophoresis (PFGE) is a highly used method in which *E. coli* O157:H7 and other EHECs are molecular subtyped using restriction enzymes. We have established a protocol that is superior to PFGE for molecular typing of *E. coli* O157:H7 with the completion of the multi-locus variable-number tandem repeat (VNTR) analysis (MLVA) protocol (6,7). MLVA discriminates among target isolates allowing for rapid identification of potential outbreak and/or sporadic strains. The advantages of MLVA for *E. coli* O157:H7 are its rapid procedure, more objective data, and high discriminatory powers. With other EHEC strains playing an important role worldwide, a subtyping technique that can be used to type any strain

using the same parameters would be very beneficial. While laboratories are not always capable of determining which EHEC serotype they have, besides *E. coli* O157, this should not inhibit the ability of officials to examine the potential of an outbreak situation. If our MLVA protocol works for non-O157 isolates, another powerful tool will be added to the arsenal of EHEC surveillance.

3.4.3. Materials and Methods

Isolates.

Forty strains of various EHECs were obtained from the Instituto Adolfo Lutz, a large public research and reference laboratory in São Paulo, Brazil; strains were from human and bovine sources. A human strain of *E. coli* O157:H7 was used as a positive control for PCR amplification (SPB24). The remaining 8 EHEC strains were from the University of Maryland, Baltimore, Maryland.

PCR Amplification.

DNA was isolated using the Prepman Ultra protocol as previously described (6) All isolates were PCR amplified using the protocol as previously described (6), but using the genotyping primers (7).

Sequencing.

Amplified products from the Baltimore, Maryland isolates were sequenced as previously described (6). Sequenced products from Brazil were purified using the protocol from Instituto Adolfo Lutz. Briefly, 80 uL of 75% isopropanol were added to each sample in a 96- well plate

Table 5. Isolates used in this experiment. Information included: serotype and source; those strains identified by "PHIDL" are from Baltimore, Maryland, and those marked "SPB" are from São, Paulo, Brazil.

Organism	Isolate #	Source
E. coli OX3:H21	PHIDL1	UNKNOWN
E. coli O185:H25	PHIDL2	UNKNOWN
E. coli O45:H2	PHIDL3	UNKNOWN
E. coli O128:H27	PHIDL4	UNKNOWN
E. coli O121:H19	PHIDL5	UNKNOWN
E. coli O91:H21	PHIDL6	UNKNOWN
E. coli O146:H21	PHIDL7	UNKNOWN
E. coli O137:H41	PHIDL8	UNKNOWN
E. coli O22:H8	SPB1	BOVINE
E. coli O22:H16	SPB2	BOVINE
E. coli O26:H11	SPB3	HUMAN
E. coli O44:H25	SPB4	BOVINE
E. coli O55:H19	SPB5	HUMAN
E. coli O77:H18	SPB6	BOVINE
E. coli O79:H14	SPB7	BOVINE
E. coli O87:H16	SPB8	BOVINE
E. coli O88:H25	SPB9	BOVINE
E. coli O91:H21	SPB10	HUMAN
E. coli O93:H19	SPB11	HUMAN
E. coli O98:H4	SPB12	BOVINE
E. coli O98:H17	SPB13	BOVINE
E. coli O103:H2	SPB14	HUMAN
E. coli O105:H18	SPB15	BOVINE
E. coli O111:HNM	SPB16	BOVINE
E. coli O111:H8	SPB17	HUMAN
E. coli O111:H11	SPB18	HUMAN
E. coli O111:HNM	SPB19	HUMAN
E. coli O112:H21	SPB20	BOVINE
E. coli O113:H21	SPB21	BOVINE
E. coli O118:H16	SPB22	HUMAN
E. coli O136:HNT	SPB23	BOVINE
E. coli O157:H7	SPB24	HUMAN
E. coli O157:H7	SPB25	BOVINE
E. coli O178:H18	SPB26	BOVINE
E. coli O178:H19	SPB27	BOVINE
E. coli O181:H4	SPB28	BOVINE
E. coli O181:HNT	SPB29	BOVINE
E. coli ONT:H2	SPB30	HUMAN
E. coli ONT:H8	SPB31	BOVINE
E. coli ONT:H16	SPB32	BOVINE
E. coli ONT:H19	SPB33	BOVINE
E. coli ONT:H21	SPB34	BOVINE
E. coli ONT:H28	SPB35	BOVINE
E. coli O178:H19	SPB36	BOVINE
E. coli ONT:H46	SPB37	BOVINE
E. coli ONT:H49	SPB38	BOVINE
E. coli ONT:HNM	SPB39	BOVINE
E. coli ONT:HNT	SPB40	BOVINE
E. coli OR:H46	SPB41	BOVINE

and allowed to sit at room temperature for 30 minutes. The plate was centrifuged for 45 minutes at 4,000 rpm at room temperature; then the supernatant was discarded. One hundred uL of 70% ethanol were added to each sample and then centrifuged for 15 minutes at 4,000 rpm at RT. The supernatant was discarded and then the ethanol step was repeated. Finally, the inverted plate was centrifuged at 200 rpm for a short pulse. The plate was dried by heating it on a thermocycler for 2 minutes at 94^{o}C. The dried products were stored at –20^{o}C until ready for the sequencing run. Immediately prior to sequencing, 10uL of Hi-Di formamide were added to each well to resuspend the pellet. The plates then were placed on a 3100 DNA Analyzer (Applied Biosystems, Foster City, Calif). The 8 remaining isolates from Baltimore, MD were sequenced as previously described (6).

Analysis of Sequenced Products.

Sequencing results were analyzed using Phred, Phrap, and Chromas as previously described (6).

3.4.4. Results

PCR Results.

All 49 isolates were analyzed using the 7 sets of genotyping primers from the *E. coli* O157:H7 MLVA protocol (Table 6). The TR3 primers did not amplify any of the samples, except for the human O157:H7 isolate. TR1 and TR5 had very limited success with a PCR product only from *E. coli* O121:H19. TR2 had limited success with 7/49 isolates amplifying. TR6 had partial success with 35/49 isolates amplifying. TR4 had almost complete amplification success (48/49). TR7 had complete success with all isolates resulting in a PCR product (49/49).

Table 6. **PCR and sequencing results of non-O157 EHECs using the 7 MLVA loci.** The sequence shown below for each locus is the known *E. coli* O157:H7 sequence. All isolates that successfully amplified using the MLVA loci primers were then sequenced to examine the similarities or differences to the known sequence.

			Sequence Compared to O157:H7 TR	
TR Locus	O157:H7 TR Sequence	PCR[a]	Identical	Different
TR1	AAATAG	1/48 (2%)	1/1 (100%)	0/1 (0%)
TR2	TGGCTC	6/48 (12.5%)	1/6 (16.7%)	5/6[b] (83.3%)
TR3	TATCTT	0/48 (0%)	--	--
TR4	TGCAAA	47/48 (98%)	2/47 (4.2%)	45/47[c] (95.8%)
TR5	AAGGTG	1/48 (2%)	1/1 (100%)	0/1 (0%)
TR6	TTAAATAATCTACAGAAG	34/48 (70.8%)	3/34 (8.8%)	31/34[d] (91.2%)
TR7	GACCAC	48/48 (100%)	2/48[e] (4.2%)	46/48[f] (95.8%)

[a] All isolates that amplified, also sequenced successfully.

[b] Single copy of "TGGCT**G**"

[c] Single copy of "TGCAAA"

[d] TR begins identically to O157 TR than degenerates

[e] Bovine O157:H7 had single copy

[f] Similar sequence of "**C**ACCACGACCA**T**"

Sequencing Results.

TR1 and TR 5 had successful sequencing products for *E. coli* O121:H19; TR1 ("AAATAG") repeated 14 times and TR5 ("AAGGTG"), 8 times. The sequences of the flanking regions and the TRs themselves were identical to that seen in *E. coli* O157:H7 (6) (Table 6).

The TR2 locus was amplified for 7 isolates, in which only *E. coli* O121:H19 had the repeat seen in O157:H7; "TGGTCT" was repeated 34 times in this isolate. The remaining isolates had flanking regions with low homology, with around 30% in the 50 bp before and after the repeat, to O157:H7 and did not contain the known repeat. Instead a single copy of "TGGCTG" was found within the range of the O157:H7 repeat (Figure 8). Those non-O157 isolates without the TR sequence had highly homologous sequences (>95%) (Table 6).

The TR3 locus in *E. coli* O157:H7 was a 6-basepair repeat containing the sequence "TATCTT", which was found to repeat from 3 to 10 times in the previous investigation (6). The Brazilian human O157:H7 isolate (SPB24) contained the same sequence repeated 8 times and the flanking regions were homologous (Table 6).

TR4 in *E. coli* O157:H7 was a 6-basepair repeat containing the repeat "TGCAAA." In our previous study, we found that TR4 was minimally repeated 2 to 9 times (6). In this study, the human O157:H7 was repeated 4 times. The remaining isolates, including the bovine O157:H7 (SBP 25), had a single copy of "TGCAAA" excluding 2 isolates: *E. coli* O87:H16 (SPB 8) had 2 copies and *E. coli* O121:H19 (PHIDL 5) had 4 copies (Table 6). The flanking regions surrounding the target sequence were highly similar to the O157 (around 96% homology) for the single copy isolates and identical for *E. coli* O87:H16 and *E. coli* O121:H19 (Figure 9).

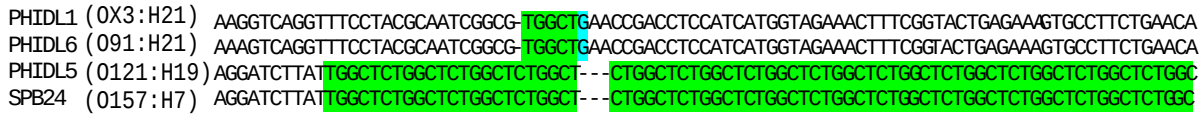


Figure 8. A portion of the TR2 locus. PHIDL 5 (*E.*coli O121:H19) shows the known TR2 repeat as compared to SPB 24 (human *E.* coli O157:H7). PHIDL 1 & 6 show highly homolgous sequences to the O157 TR2 sequence, but they differ at a single nucleotide and are present only 1 time. These sequences were typical of the other nonO157s that were sequenced at the TR2 locus (except for PHIDL 5). The green highlighted region show the repeat "TGGCTC," while the blue highlighted nucleotides show the difference in the non-O157 sequence.

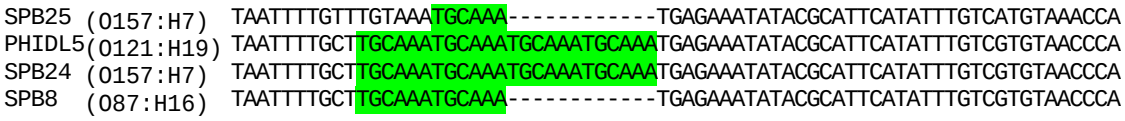


Figure 9. The TR4 locus containing the flanking regions and the repeat itself. The flanking regions of this VNTR were highly homologous even in those isolates that only had a single copy of the repeat. The green highlights show the known repeat "TGCAAA."

TR6 in *E. coli* O157:H7 was an 18-basepair repeat containing the sequence "TTAAATAATCTACAGAAG," but this repeat did tend to have some slight variation in its sequence. We found a range of 7-12 repeats in our collection of *E. coli* O157:H7 (6). The human isolate of *E. coli* O157:H7 in this study repeated 9 times. Three non-O157s did contain the repeats seen in O157:H7: O121:H19, O137:H41, and OX3:H21 which repeated 10, 14, and 16 times, respectively. In the remaining non-O157s analyzed, several copies of a repeat were found that began identically to the O157:H7 isolate but then varied dramatically. The 5' end-flanking region had 93 nucleotides that were identical between O157 and the non-O157s, minus 2 nucleotides leading into the first repeat, in which all isolates had the beginning sequence of "TTAAATA." The O157 repeat continued as previously described (6), but the majority of the non-O157 had a variant sequence (Table 6). The second repeat began similarly between the O157 and most of the non-O157 isolates, but again, the non-O157s quickly diverged from the O157 sequence and from the first non-O157 repeat (Figure 10). There was a possibility of more repeats in the non-O157 isolates, but there was substantial sequence diversity making it difficult to determine where the repeat terminated.

TR7 in *E. coli* O157:H7 was another 6-basepair repeat containing the sequence "GACCAC." Previously, we found a range of 4-9 tandem repeats in our collection of *E. coli* O157:H7 (6). In this study, the human O157:H7 was repeated 8 times while the bovine isolate had a single copy. The remaining 47 isolates did not contain the sequence "GACCAC", except for *E. coli* O121:H19 in which "GACCAC" repeated 6 times (Figure 11, Table 6). The flanking regions were highly similar; although minimal sequence was obtained on the 5' end, the 3' end was

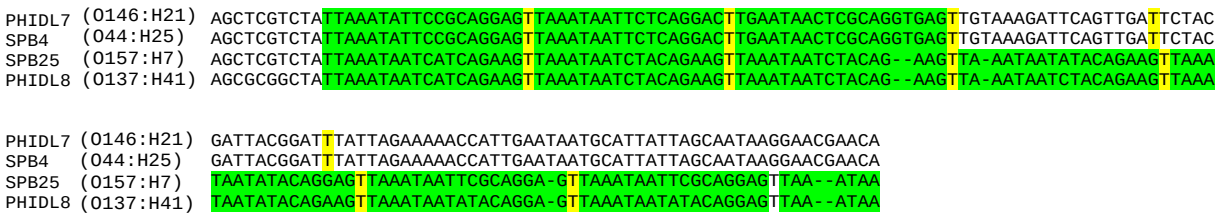


Figure 10. **The 5' end of the TR6 locus.** The first 2 repeats began the same regardless if the isolate was an O157:H7, a non-O157 with the O157 repeat, or a divergent non-O157. The divergent non-O157s all had sequences that were identical to each other. Additionally, the O157:H7 and the non-O157s with the typical repeat also were identical to each other. But these 2 groups differed when compared to each other. The repeat is highlighted in green while the beginning of each repeat is highlighted in yellow.

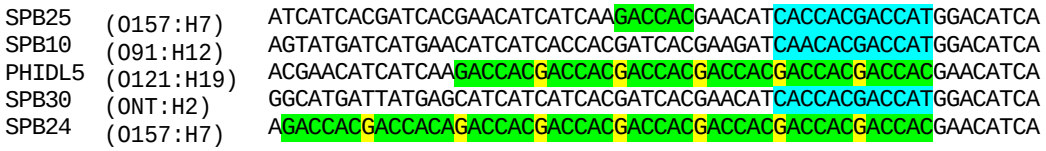


Figure 11. **Sequences of several non-O157 & O157 isolates for TR7.** The isolates that contain the known sequence of "GACCAC" were highly homologous to each other and differed by the number of times the repeat repeated. The other isolates, while highly homologous to the O157 sequence, were even more similar to each other. The typical repeat is highlighted in green with the beginning of each typical repeat highlighted in yellow. The blue highlighted region represents the similar sequence found in some of the non-O157s.

identical except for a 9-nucleotide insert in the human O157:H7 isolates. The area in which the tandem repeat was located in O157:H7, the non-O157s contained a similar sequence of "CACCACGACCAT."

3.4.5. Discussion

Several important and interesting results were gained in this pilot study on the use of our MLVA assay for molecular subtyping of non-O157 EHEC. Initially, the results were surprising in that 4 of the 7 loci did not amplify the majority of strains, and the remaining 3 loci had different sequences at the targeted loci. But after further analysis of the results, including the fact that a majority of these strains are bovine isolates, the results may not be so surprising. Additionally, EHECs are characterized as such because of the presence of shiga toxins and a few other virulence genes. This allows for the potential for great variation among the remaining genome from one serotype to the next, as they have most likely evolved from the same lineage that have been diverging from each other for a long time (8).

There is a possibility that several of these loci may be true VNTRs within specific serotypes. This study examined only 1 isolate from 47 different serotypes, many of them bovine isolates. One thought is that looking at 10-15 isolates from a single serotype may result in the discovery of multiple alleles at a specific locus. For example, the serotype *E. coli* O87:H16 at TR4 may have multiple alleles as this one isolate did have 2 copies of the tandem repeat. Additionally, *E. coli* O121:H19 had 6 of the 7 MLVA loci showing a "typical" repeat behavior. But more likely, the majority of sequenced loci are not VNTRs in these 47 serotypes. The reasoning behind this is that the vast majority either only had a single copy (hence not a tandem repeat by definition) or

the tandem repeat was absent. The chance that all the loci examined for these isolates had these results but are still VNTRs is highly unlikely.

Interestingly, the majority of the non-O157 loci that were amplified had a high sequence homology to O157:H7 and an even higher, if not identical, sequence to each other even though there were 47 serotypes. Strangely, the O157:H7 isolates appear to have developed VNTRs at these loci while the non-O157s have not (6). Whether these VNTRs confer advantages or disadvantages to *E. coli* O157:H7 is not known.

These findings all indicate that further examination of specific serotypes is needed. Ideally a panel of VNTR loci can be identified that can be used for all of the major EHEC serotypes. Future plans should include obtaining 10 or so isolates from each serotype of interest, representing the other major serotypes that cause disease both in the United States and worldwide (eg. serotypes O111 and O104:H21). Both bovine and human isolates should be examined to determine if there is a propensity for VNTRs to be present in human versus bovine isolates. Perhaps portions of the genomes will have been sequenced and can be searched for VNTRs; otherwise AFLP may be used to detect potential VNTRs (4). Additionally, there may be some O157:H7 VNTRs that were not used for the original MLVA protocol that would be useful with some of the non-O157s. Consideration should also be given to designing a less expensive MLVA protocol that may be applicable for labs that do not have the resources for the standard protocol. This includes choosing VNTRs that are large (>50bp) allowing the elucidation of alleles using regular gel agarose to examine differences. With the ever-increasing awareness of disease caused by non-O157 EHECs and the globalization of food distribution, the

need for a subtyping technique that can quickly identify potential EHEC outbreaks has become a pressing issue.

3.4.6. Literature Cited

1. **Bettelheim, K.A.** 2003. Non-O157 verotoxin-producing *Escherichia coli*: A problem, paradox, and paradigm. Exp. Biol. Med. **228**:333-344.

2. **Brooks, J.T., D. Bergmire-Sweat, M. Kennedy, K. Hendricks, M. Garcia et al.** 2004. Outbreak of shiga toxin-producing *Escherichia coli* O111:H8 infections among attendees of a high school cheerleading camp. Clin. Inf. Dis. **38**:190-198.

3. **Centers for Disease Control and Prevention.** 1995. Outbreak of acute gastroenteritis attributable to *Escherichia coli* serotype O104:H21 – Helena, Montana, 1994. Morbid. Mortal. Wkly. Rep. **44**:501-503.

4. **Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P.Jackson.** 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymophism markers. J. Bacteriol. **179:**818-824.

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

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

7. **Noller, A.C., M.C. McEllistrem, and L.H. Harrison.** 2004. Genotyping primers for the fully automated multi-locus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Micr. **42:**3908.

8. **Reid, S.D., C.J. Herbelin, A.C. Bumbaugh, R.K. Selander, and T.S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature. **406:**64-67.

3.5. Chapter 5

Mutational Events of the Seven Loci of the Multi-Locus Variable-Number Tandem Repeat Analysis Assay for *Escherichia coli* O157:H7

Manuscript in Preparation:

3.5.1. Preface

This work will be submitted to a peer-reviewed journal, and is representative of the work done to answer the questions posed in Specific Aim 2. Additional work will be needed to completely characterize the manner in which VNTRs mutate and how this affects in turn the interpretation of MLVA data.

3.5.2. Abstract

Multi-locus variable-number tandem repeat (VNTR) analysis (MLVA) for molecular subtyping of *Escherichia coli* O157:H7 is a fast, reproducible, and sensitive method for determining genetic relatedness and detecting outbreaks. However, the high mutability of VNTR loci creates the need to understand the dynamics of how these loci change. Using a representative *E. coli* O157:H7 outbreak isolate, two types of mutations experiments were performed. An analysis of 384 isolates derived from a single culture revealed no changes in the seven MLVA loci (TR1-TR7) that were studied. In contrast, serial subcultures collected twice daily for five days revealed a total of 41 mutation events. TR2 had 35 of the 41 total events with an average mutation rate of 3.5×10^{-3}. Additionally, 27/35 mutation events in TR2 were single additions. TR1 and TR5 also had events, but at a much lower rate of 2 and 4 events, respectively and an average mutation rate of 1.9×10^{-4} for TR1, and 4.0×10^{-4} for TR5. The remaining four loci had no slippage events in either of the experiments. These data indicate that growth conditions influence mutational dynamics of TR loci and that the tendency to change varies by locus. These locus-specific differences must be taken into account when interpreting MLVA data from epidemiologic investigations.

3.5.3. Introduction

Prokaryotic genomes contain a wide array of repetitive DNA elements ranging from single nucleotide repeats to large, complicated repeats of dozens of nucleotides. Variable-number tandem repeats (VNTRs) are repeats that are found in tandem and demonstrate inter-strain variability. Multi-locus VNTR analysis (MLVA) has become a reliable way to establish genetic relatedness for epidemiological surveillance and molecular subtyping of organisms such as *Escherichia coli* O157:H7 (17, 18), *Salmonella typhimurium* (14), *Francisella tularensis* (5), and *Bacillus anthracis* (9). The basis of molecular typing using VNTRs is that these elements can mutate creating different alleles at the same VNTR locus.

Tandem repeat (TR) loci are among the most variables regions of bacterial genomes. The mechanism by which tandem repeats mutate has been suggested to be the result of slippage and mispairing during DNA replication because of pausing and dissociation by DNA polymerase while within the tandem repeat (19, 21, 22). Repeats can be inserted or deleted depending on the strand orientation (15). If the tertiary structure occurs in the template strand during replication, this results in a loss of at least one tandem repeat in the new DNA strand, while an event in the nascent strand results in the addition of one or more tandem repeat. Additionally, several models have been proposed to explain the mutation process seen in VNTRs. The stepwise mutation model proposes that VNTR alleles evolve through a gain or loss of a single repeat (2, 11). The infinite allele model assumes that the size of a new allele is independent of the ancestral allele (1, 2). Finally, the two-phase mutation model combines the 2 previous ideas and proposes that most changes lead to a change of one repeat change and that only a small portion of mutations involves large changes (2, 3). Estimations on the rate of change have been performed to a limited

degree with human VNTRs due to their involvement with heritable diseases, but there is a scarcity of literature on the rate of mutational changes of VNTRs in bacteria.

Multiple factors influence the frequency and type of TR mutation, such as the number of TRs and the unit size of the TR. As the number of TRs increases, the slippage mutation rate dramatically increases due to instability, which accounts for the fact that long tandem repeats are relatively uncommon (12). The mutation rate also increases with perfectly homologous repeats (15). In contrast, a disruption of the repeat, caused by indels or point mutations, decreases the mutation rate by decreasing the length of the homologous repeat. In addition to the number of TRs, the nucleotide length and composition of each TR can affect the rate of slippage mutation: the shorter the TR unit length, the higher the mutation rate (20). Poly (G-T) tracts and polypyrimidine tracts also have been shown to be associated with high mutation rates (13). Multiple studies have shown that certain repeats, such as repeating purine-pyrimidine sequences, result in a bias towards expansion if the sequence is on the leading strand (3, 7, 8).

Knowledge of the rate of TR change is important when using MLVA for epidemiologic purposes, such as outbreak detection. In a previous study, we demonstrated that isolates from the same outbreak either had an identical MLVA type or were single locus variants (SLVs), suggesting that mutations occur during the course of an outbreak (17, 18). Intra-outbreak events have been observed using other molecular subtyping methods, such as pulsed-field gel electrophoresis (PFGE) (10, 17). All SLVs that were observed during outbreaks differed from the predominant MLVA type by a single repeat. In addition, analysis of our MLVA data demonstrates that the TR1, TR2, and TR5 loci had a greater number of alleles than the other 4

loci (17). These observations underscore the need to understand the dynamics of tandem repeat mutation events at each locus for the optimal interpretation of MLVA results.

Using a known outbreak strain of *E. coli* O157:H7 that was an SLV of the predominant MLVA type, we performed both dependent and an independent mutation experiments to examine the MLVA VNTR mutation rates. The issue of independence arises when determining if a detected mutation is an original event or is just a detection of the offspring of the original mutation event. The purpose of this study was to examine tandem repeat mutational events to improve our understanding of MLVA as an outbreak detection tool.

3.5.4. Materials and Methods

Isolate.

The *Escherichia coli* O157:H7 isolate (PHIDL #53) chosen was an outbreak isolate received from the Allegheny County Health Department. This isolate was selected for two reasons. The outbreak strain was a SLV of the predominant MLVA outbreak type, with 16 repeats rather than 15 repeats at the TR1 locus. Moreover, the strain's MLVA alleles were close to the median number of repeats of the 80 isolates in our previous study (Table 7) (17).

Parallel (Independent) Mutation Experiment.

To alleviate the potential issue of non-independent events, 10 parallel cultures were cultured simultaneously. From a frozen culture, the study isolate was streaked for isolation on 5% sheep's blood agar and incubated for 24 hours at 37^{o}C. A single colony was picked and inoculated into 10mL of Luria broth (LB); a small portion of the colony was used to determine

Table 7. **Range of alleles for the 7 MLVA loci**. The 7 loci of PHIDL #53, the isolate chosen for the present study, were close to the median range for the alleles. The numbers represent the number of times the TR repeated at the particular locus. The minimum and maximum TR sizes represent the ranges seen in our previous study (17).

	TR1	TR2	TR3	TR4	TR5	TR6	TR7
Min TR	4	7	3	2	6	7	4
Median	13	24	7	5	10	10	6
PHIDL #53	16	25	7	5	10	10	6
Max TR	20	58	10	9	12	12	9

the starting number of tandem repeats at the 7 MLVA loci by sequencing as described previously (17). The suspended colony was divided into 10 flasks each with 9.9mL of LB in a 25 mL Erlenmeyer flask. The cultures were incubated for 12 hours at 37^{o}C in a rotating water bath. One hundred uL of a 10^{-4} dilution from each flask was inoculated onto blood agar plates. After a 24-hour incubation at 37^{o}C, the number of colony forming units (CFUs) was counted. Eight of ten plates had at least 48 colonies/plate and were analyzed via MLVA as previously described (17, 18)

Serial (Dependent) Mutation Experiment.

The initial incubation of the study isolate from a frozen culture to a single colony in 10mL of LB was identical to the parallel study. Again, a small portion of the colony was used to verify the starting number of tandem repeats at the 7 MLVA loci as described above. The LB culture was incubated at 37^{o}C in a rotating water bath for 12 hours. Dilutions of 1:100 were made using LB. One hundred uL of a 10^{-4} dilution was added to 9.9mL of fresh broth and placed into the rotating water bath for 12 hours. Dilutions from the original flask were made: 10^{-4}, 10^{-6}, and 10^{-8}; 100uL of each of these dilutions were plated onto blood agar plates and incubated at 37^{o}C for 24 hours. This serial subculturing and plating was performed 9 more times for a total of 10 12-hour serial subcultures. The number of generations per time point was calculated using the following calculation:

$$\frac{\text{Log}_{10}(\#\text{CFU}_{\text{End of 12-hr}}) - \text{Log}_{10}(\#\text{CFU}_{\text{Beginning of 12-hr}})}{\text{Log}_{10}2} \quad \text{(Eq. 1)}$$

After 24 hours of incubation on blood plates, the total number of CFU was counted from the 10^{-6} plates and 48 isolated colonies were picked and each added to 50mL of sterile water. The suspended colonies were then boiled for 2 minutes. DNA was stored at -20^{o}C until evaluation by

MLVA. The MLVA protocol was identical to that previously described except 2.5mL of DNA was used in the 30uL reaction (17, 18).

Data Analysis.

The mutation rate was based on the mutant fraction and number of generations. The mutant fraction was the number of mutation events divided by the number of colonies screened at that time (16). This value was multiplied by 2/number of generations to compute the mutation rate. At each time point over the five day period, the mutation rate for TR2 was calculated. These nine rates were averaged to determine an overall rate for each experiment.

3.5.5. Results

Parallel (Independent) Mutation Experiment.

Due to insufficient numbers of colonies, two of the 10 subcultured blood plates were therefore excluded from the analysis. An analysis of 48 colonies per 8 plates after the 12 hour incubation revealed no mutation events (0/384). In other words, all colonies had the identical MLVA type compared to the original colony.

Serial (Dependent) Mutation Experiment.

These experiments were performed in duplicate, resulting in similar numbers of generations and a similar pattern of mutation events (Table 8). The largest number of mutation events was seen in TR2 for both experiments: 11/13 (84.6%) in experiment 1 and 24/28 (85.7%) in the second experiment. TR2 mutation events were seen throughout the time course of the experiments; the majority of events were single TR additions, but both deletion and addition events of single and

Table 8. **Number of observed mutation events for the 2 serial mutation experiments by tandem repeat (TR) locus.**

	Ave Gen/ Time Pt	Total # Gen	Observed Mutation Events							Total Events
			TR1	TR2	TR3	TR4	TR5	TR6	TR7	
Experiment 1	21.8	218	1	11	0	0	1	0	0	13
Experiment 2	21.6	216	1	24	0	0	3	0	0	28
Total Events			2	35	0	0	4	0	0	

multiple TR units were observed (Figure 12). The average mutation rate for TR2 in experiment 1 was $2.1x10^{-3}$ while in experiment 2 the average rate was $4.8x10^{-3}$. TR1 and TR5 had single addition events but resulted in a lower mutation rate than TR2. TR1 was associated with one event in both experiments and TR5 was associated with one event in the first experiment and three in the second experiment. For TR1 the average mutation rate for experiment 1 was $1.7x10^{-4}$ and for experiment 2 it was $2.1x10^{-4}$. For TR5 experiment 1 demonstrated an average mutation rate of $1.7x10^{-4}$ and experiment 2 an average rate of $6.2x10^{-4}$. The remaining four loci had no slippage events in either of the experiments, including the locus associated with the SLV *in vivo*.

3.5.6. Discussion

In this study, we detected no mutation events in the parallel mutation study, but in contrast observed 41 mutation events seen in the serial mutation study. Thirty-five of those 42 events were found at the TR2 locus, with only 2 and 4 events occurring at the TR1 and TR5 loci, respectively. During both serial mutation experiments, there was an overwhelming tendency towards the addition of tandem repeats, and most involved a single tandem repeat change at all 4 loci that mutated. This is consistent with our previous study in which we observed 3 intra-outbreak SLVs, all differing by a single tandem repeat (17).

VNTRs have been shown to be a powerful tool for the molecular subtyping of a wide range of organisms, including humans, plants, and bacteria (2-5, 9, 14, 17). However, interpretation of assays that exploit these genetic elements for epidemiologic purposes requires an understanding of the dynamics of TR mutations. Many studies on VNTR mutations have been focused on those VNTRs associated with human disease (3, 8, 19). While general comparisons about VNTR

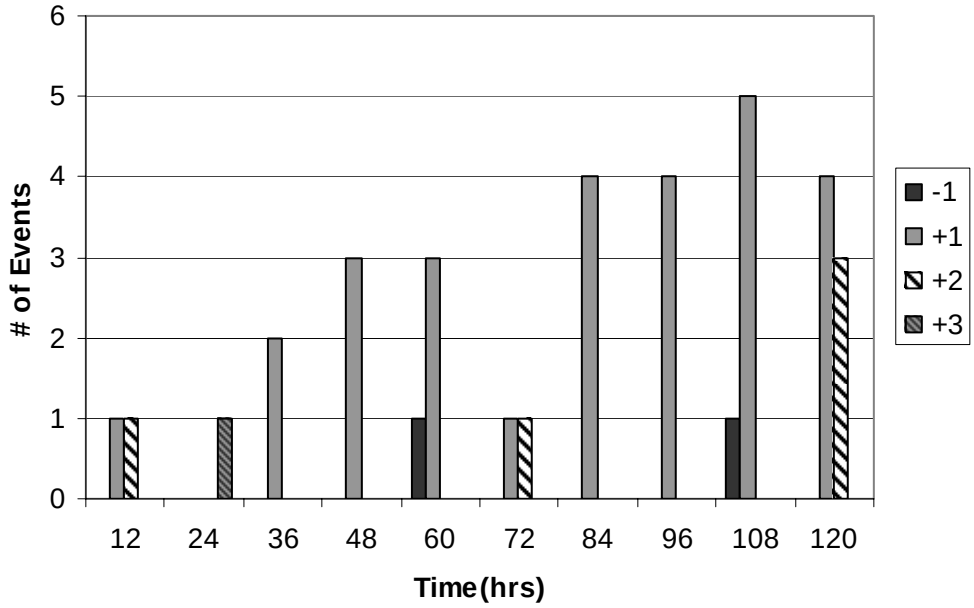


Figure 12. **Mutation events in TR2 for both serial mutation experiments combined.** The majority of observed events were single additions (+1).

mutations can be made between bacteria and eukaryotes due to their highly conserved genes involved in mutation repair, a detailed MLVA mutation analysis needs to be made for each locus in each bacterial species (6).

Our serial mutation experimental design allowed for a completely random sampling of the original bacterial culture with subsequent analysis at nine more time points. Due the long interval between time points the bacteria reached stationary phase prior to sampling. We hypothesize that the multiple cycles through the stationary phase may have been important for these mutations to occur. This is supported by lack of observed mutation events during the parallel study, even though roughly the same number of isolates was analyzed. The major difference between these two experiments was that the parallel cultures entered stationary phase only once. We believe that the serial mutation experiment more closely reflects what occurs *in vivo* than the parallel experiment. During the course of an outbreak, a strain likely goes through serial exponential and stationary phases due to passage from food to humans and subsequent human to human transmission.

The most mutable locus in our MLVA assay was TR2. In an unpublished study of *E. coli* O157:H7, a single colony seeded 96 parallel cultures that subsequently was subcultured 40 times (C. Keys, Z. Jay, A. Fleishman, J. Fox, G. Evans, and P. Keim, poster, 103rd General Meeting American Society for Microbiology, Washington, D.C., 2003, 9). The study had the advantage of removing the issue of independence, but had some similar observations to this study. For example, their study also demonstrated the hypermutability of TR2. The current literature suggests that tandem repeats can have a tendency towards expansion depending on the

nucleotide sequence composition. Most research has been done with mono-, di-, tri-, and tetranucleotide sequences showing that certain sequences can form tertiary sequences increasing the rate of mutation (4, 7, 8). To our knowledge, there is no literature described hexanucleotide tandem repeats and their mutation propensities. Perhaps the repeat sequence "TGGCTC" in the leading strand may form a tertiary structure that promotes additional events.

The other 2 loci in which events were observed had lower rates of change than TR2. TR1 had a single addition event in both 5-day experiments, while TR5 had several single addition events in the 5-day experiments. Since only 6 events were observed, no firm conclusions can be made about the propensity towards additions versus deletions and single versus multiple TR changes.

Although no mutations were observed in TR3, TR4, TR6 and TR7, our previous study indicates that these loci do mutate in natural populations (16). Further experiments will be required to better understand the mutational dynamics at these loci. In addition, to fully characterize how the MLVA loci mutate, a more detailed investigation would need to be undertaken. The allelic extremes of each locus would need to be examined to truly determine how the length of the TR influences the mutational dynamics. These questions must be answered to completely understand how TRs change for the optimal interpretation of MLVA data.

In light of the results from these experiments and our previous study, several generalizations can be made about the interpretation of MLVA data during outbreak investigations. Patients with isolates with an identical MLVA type likely acquired their infection from the same source. In addition, SLVs that differ by either a single or double TR then should also be considered to be

suspicious for having come from the same source and should be investigated accordingly. If an isolate differs from other suspected outbreak isolates at two different loci, then the specific loci involved and the difference in the number of TR would become important to assess. A double locus variant that includes an increase in TR size by two within the TR2 locus might part of the outbreak based on these data. In general, precise cut-offs will be difficult to develop because some variables, such as the duration of the outbreak, will likely increase the likelihood that additional intra-outbreak variability will be observed. Therefore, as with all molecular subtyping methods, the interpretation of MLVA results must be done in conjunction with the epidemiologic data.

3.5.7. Literature Cited

1. **Ballous, F. and N. Lugon-Moulin.** 2002. The estimation of population differentiation with microsatellite markers. Mol. Evol. **11:**155-165.

2. **Cozzolino, S., D. Cafasso, G. Pellegrino, A. Musacchio, and A. Widmer.** 2003. Molecular evolution of a plastid tandem repeat locus in an orchid lineage. J. Mol. Evol. **57:**S41-S49.

3. **Di Rienzo, A., A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, and N.B. Freimer.** 1994. Mutational processes of simple-sequence repeat loci in human populations. PNAS. **91:**3166-3170.

4. **Eckert, K.A. and G. Yan.** 2000. Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: Influence of sequence on expansion mutagenesis. Nuc. Acids Res. **28:**2831-2838.

5. **Farlow, J., K.L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim.** 2001. *Francisella tularensis* strain typing using multiple-locus variable-number tandem repeat analysis. J.Clin. Micr. **39:**3186-3192.

6. **Field, D. and C. Wills.** 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellite in eight prokaryotes and *S. cerevisiae* result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad.Sci. USA. **95:**1647-1652.

7. **Freudenreich, C.H., J.B. Stavenhagen, and V.A. Zakian.** 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol. Cell. Bio. **17:**2090-2098.

8. **Kang, S., A. Jaworski, K. Ohshima, and R.D. Wells.** 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. Nature Gen. **10:**213-218.

9. **Keim, P., L.B. Price, A.M. Klevytska, K.L. Smith, J.M. Schupp, R. Okinaka, P.J. Jackson, and M.E. Hugh-Jones.** 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bact. **182:**2928-2936.

10. **Kimura, A.C., K. Johnson, M.S. Palumbo, J. Hopkins, J.C. Boase, R. Reporter, M. Goldoft, K.R. Stefonek, J.A. Farrar, T.J. Van Giler, & D.J. Vugia.** 2004. Multistate Shigellosis outbreak and commercially prepared food, United States. Emerg. Inf. Dis. **10:**1147-1149.

11. **Kimura, M. and T. Ohtia.** 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. Proc. Natl. Acad. Sci. USA. **75:**2868-2872.

12. **Lai Y. & F. Sun.** 2003. The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol. Evol. **20:** 2123-2131.

13. **Levinson G. & G.A. Gutman.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:** 203-221.

14. **Lindstedt, B., E. Heir, E. Gjernes, and G. Kapperud.** 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat loci. J. Clin. Micr. **41:**1469-1479.

15. **Lovett, S.T. & V.V. Feschenko.** 1996. Stabilization of diverged tandem repeats by mismatch repair: Evidence for deletion formation via a misaligned replication intermediate. Proc. Natl. Acad. Sci. USA. **93:** 7120-7124.

16. **Nadas, A, E.I. Goncharova and T.G. Rossman.** 1996. Mutations and infinity: improved statistical methods for estimating spontaneous rates. Environ. Mol. Mutagen. **28**:90-99.

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

18. **Noller, A.C., M.C. McEllistrem, and L.H. Harrison.** 2004. Genotyping primers for the fully automated multi-locus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Micr. **42:**3908.

19. **Sharma, R., S. Bhatti, M. Gomez, R.M. Clark, C. Murray, T. Ashizawa, and S.I. Bidichandani.** 2002. The GAA triplet-repeat sequence in friedreich ataxia shows a high level of somatic instability *in vivo*, with a significant predilection for large contractions. Hum. Mol. Gen. **11:**2175-2187.

20. **Symonds V.V. & A.M. Lloyd.** 2003. An analysis of microsatellite loci in *Arabidopsis thaliana*: Mutational dynamics and application. Genetics. **165:** 1475-1488.

21. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequences DNA repeats in prokaryotic genomes. Micro & Mol Biol Rev. **62:** 275-293.

22. **Viguera, E., D. Canceill, and S.D. Ehrlich.** 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO. **20:**2587-2595.

3.6. Chapter 6

The Functional Roles of Variable Number Tandem Repeats in *Escherichia coli* O157:H7

3.6.1. Preface

The work in the following chapter reflects the preliminary steps towards examining the potential, functional roles of variable-number tandem repeats in *Escherichia coli* O157:H7. This work will not be published but can be used as a building block for future experiments.

3.6.2. Introduction

Eukaryotic and prokaryotic genomes contain tandemly-repeated elements that have been utilized to differentiate strains. In humans, many of these repeated nucleotide tracts have been linked to heritable diseases, such as Fragile X syndrome and Huntington Disease (12). Many bacterial TRs also have been shown to have a functional role. For example, *Haemophilus influenzae* contains a tetranucleotide repeat "CAAT" that is located in a gene resulting in different patterns of lipopolysaccharide expression (14). Not only are TRs found directly within affected genes, but can also effect gene transcription by regulating promoters. In *Neisseria menigitidis*, *por*A encodes an outer membrane protein that is a vaccine target. A polyguanine tract found between the –35 and –10 promoters modulates the level of expression (14). This modulation of expression could in turn affect the efficacy of a vaccine that utilized the porA outer membrane protein as a key antigen.

Escherichia coli O157:H7 contains a large number of tandem repeats ranging from simple mono-nucleotides to complicated tracts of nucleotides over 150 bases long; some are found within genes while others are extragenic. No literature to date has examined any of the TRs in *E. coli* O157:H7 to classify their functional roles. We had hoped to study the 7 variable-number tandem repeats (VNTRs) in our multi-locus VNTR analysis (MLVA) protocol (9). Unfortunately, the 7 VNTRs are located within or between hypothetical open reading frames. The identification of these genes was beyond the scope of our study, so we focused on TRs with known or highly homologous gene products. We chose 6 genes or extragenic regions to determine if 1) the TRs were VNTRs within our sample isolates and 2) if a function of the VNTR could be determined.

One of the 6 chosen TRs with the most interesting gene was the TR located within *tol*A. TolA is a periplasmic protein with its N-terminus anchored in the inner membrane, an α-helical central domain, and a C-terminus that interacts with the inner portion of the outer membrane (8). This protein is thought to have 3 roles in the *E. coli* genome: 1) structural stability of the cell, 2) group A colicin uptake and 3) filamentous phage import (2, 6, 7). The C-terminal region appears to be the active portion of the protein that interacts with invading phage and colicin molecules. The N-terminus region interacts with other tol proteins as part of an energy-dependent process involved in the outer membrane stability (4). Separating the 2 ends of *tol*A is a 750-nucleotide region flanked on both sides with a polyglycine sequence; these regions may encode a flexible hinge. This central region is mainly an uninterrupted α-helix containing the basic amino acid sequence KA_3D/E; this is the 15-basepair tandem repeat of interest (8). The addition or loss of a TR may lengthen or shorten, respectively, the α-helix altering its proximity to the outer membrane resulting in a functional change in either membrane stability or infection by colicin and/or filamentous phage.

3.6.3. Materials & Methods

Bacterial Isolates.

E. coli O157:H7 isolates were obtained from several sources for this study. The Public Health Infectious Disease Laboratory (PHIDL) obtained all *E. coli* O157:H7 strains isolated by the Allegheny County Health Department (ACHD) from 1999 to 2003 (n=70). A sample of 18 isolates from the Minnesota Department of Health (MDH) was also included from 1996 and 1997. ATCC strain EDL933 and the Sakai, Japan, strain RIMD 0509952 were used as reference strains (5, 11). The *tol*A assays used 3 isolates representing the 3 alleles detected in the

sequencing reactions: PHIDL #60, #61, and #62. A *Salmonella typhimurium* (F'400,his+,zee-1::Tn10 [tetracycline resistance]) strain was used for the conjugation assay. JM109 *E. coli* cells (F' traD36 proA+ proB+ lacIq delta[lacZ]M15 delta[lac-proAB] supE44 hsdR17 recA1 gyrA96 thi-1 endA1 relA1 e14- lambda-) were used as a positive control for the phage assay as these cells were known M13 phage hosts.

Potential Variable Number Tandem Repeats.

During our previous investigation into VNTRs, we found over one hundred TRs in the two fully sequenced *E. coli* O157:H7 genomes, EDL933 (AE005174) and Sakai (BA000007) using the Tandem Repeats Finder software (1, 9). We chose 6 new TRs for this investigation in which 5 of the TRs had the same number of repeats in both of the reference strains. The final TR was located on the O157 plasmid in both reference strains and was variable between the 2 strains.

DNA Isolation, PCR Amplification, and Sequencing.

DNA was isolated using the Prepman Ultra Protocol as previously described (9, 10). All Allegheny County and Minnesota isolates were analyzed at the 6 loci, plus 3 isolates were analyzed at all the genes in the *tol*QRAB operon and related genes (Table 8). Primers were based on the sequences from Sakai and EDL933 genomes (5, 11). PCR amplifications were performed as previously described, but with primers and annealing temperatures specific for each of the primer pairs (9, 10) (Table 9). Sequencing of the targeted genes was done as previously described (9, 10).

Table 9. **Primer sequences and annealing temperatures of potential VNTRs and accessory proteins.**

Locus	Sequence (5'-3') Forward	Reverse	Annealing Temp (^{o}C)
hem X[1]	TCATCGGGCCAACCACGATC	TGCAAAGCCAGGCGATGCTGG	54
IS3[1]	TGAGGTACTTTGCCTCAGG	TTGCCACAGATGCAATATC	54
mop A[1]	CGAATGCATGGTTACCGACCT	CTCGCGTCGTCCGTGTCTGA	54
tir [1]	TATCAGACCTGTAGCAACAGTC	CATGCTATGGTCACCGTT	54
tol A[1]	ATGACGCAGCCTATCTAAACATCTG	ACTACCAGAACCCCGTGGCAA	54
yci [1]	GACTCGGTACGACTGGATC	ATCAGAGCGATTGTTCAGAG	54
ygb C[2]	CGCGTATAGTAGCAGCGTTT	CAGAAGGCTAGCCTTCAGG	54
tol Q[2]	TGCTGAATGAAGCAGAGGTTC	CACCAGCAGTACGTCCA	52
tol R[2]	TTTACCGCGATTCTGCACC	CTGTTCGCCTGTTACCCG	52
tol B[2]	CACGGGGTTCTGGTAGTTT	ATCATCAGCCCTTTCAGCAC	52
pal [2]	TGGACAGGTCAAATTCCCTG	CCAACCAGTAACGACAGACA	52
ygb F[2]	ACCTGCAGTACTGGGTCAT	ATGATTCGCACGACACGAC	52
btu B[2]	TCTGGTTCTCATCATCGCG	CGGATCTCGTCATAGACCG	52
omp F[2]	TGAGATTGCTCTGGAAGGC	GGAAAGATGCCTGCAGACA	54
omp A[2,3]	AGACAGCTATCGCGATTGC	GCTTTGTTGAAGTTGAACAC	50

[1]Loci associated with VNTR determination

[2]Loci associated with *tol* A studies

[3]Primers previously described (10)

Deoxycholate Assay.

Three isolates of *E. coli* O157:H7 containing the 3 different alleles of *tol*A were incubated overnight at 37°C in Luria broth (LB). The isolate concentrations were equalized to each other by measuring the OD_{650nm}. Susceptibility to deoxycholate (DOC) was measured two ways: in solid or liquid media. For the liquid experiment, 50 μL of each equalized culture was added to 6mL of LB containing 0-1% DOC (wt/vol). Cultures were incubated for 3 hours and then visual observations and the OD_{650nm} were made. One hundred μL aliquots of the bacteria-DOC mixture were subcultured onto 5% sheep's blood agar. Plates were incubated overnight (O/N) at 37°C and the colony forming units (CFU) were counted the next day. This experiment was repeated 3 times. The solid experiment consisted of making LB agar plates with increasing concentrations of DOC from 0-1%. One hundred μL of a 10^{-6} dilution of the equalized cultures were added to the plates. Plates were incubated overnight at 37°C and the CFUs were counted the next day. This experiment also was repeated 3 times. Negative controls for both assays consisted of 0% and 1% DOC media/plates being incubated with no bacteria.

Bacterial Killing Assay.

The 3 *tol*A allele strains were subcultured overnight on 5% sheep's blood plates at 37°C. Isolated colonies were incubated in 5 mL of LB O/N at 37°C with shaking. Five hundred μL of the overnight cultures were added to 20-24 mL of fresh LB and incubated with vigorous shaking for another 3 hours. Meanwhile, 200μL of the 40μM WLBU2 peptide were added to the undiluted peptide-undiluted bacteria well (3). Two-fold dilutions of the peptide were made by adding 100μL of the starting 40μM peptide into 80μL of 1X PBS buffer and continued until a final dilution of 0.075μM was obtained. One hundred eighty μL of 1X PBS buffer were added

to the remaining wells. The bacterial suspension was prepared by centrifuging 10mL of the 3-hour culture at 2800 rpms for 10 minutes. The supernatant was decanted and the pellet resuspended in 7-10mL of 1X PBS. The OD_{600nm} of the resuspended bacteria was measured and the dilution factor obtained by dividing the OD_{600nm} by 0.009. The dilution factor was divided into 3mL to determine the amount of resuspended bacteria to be added to 3mL of 1X PBS. This equation resulted in obtaining an estimated 1 x 10^5 cells for the start of the experiment. Twenty μL of bacterial suspension were added to each of the peptide dilutions and incubated for 30 minutes at 37°C. Dilutions of the bacteria were then made by taking 20μL of the bacteria-specific peptide concentration mixture and diluting 10-fold to a final bacterial concentration of 10^{-3}. Positive controls contained bacteria but no peptide. Finally, 100μL of each bacterial-peptide suspension were added to Luria plates and spread using glass beads. Plates were incubated overnight at 37°C and then the colonies counted.

Conjugation of *E. coli* O157:H7.

To make the 3 *E. coli* O157:H7 isolates (recipient[R]) F' plasmid positive cells, a donor (D) male *Salmonella typhimurium* was used as the F' plasmid source for the conjugation. This plasmid supplied both the pilus apparatus needed for phage infection and the gene for tetracycline (TC) resistance. A single colony was used to inoculate LB and was incubated overnight. Single colonies of the 3 R isolates were used to inoculate 20mL LB containing 0.4% glucose and a phosphate-nitrogen-sulfur buffer and then were incubated overnight at 37°C with shaking. The R cultures were spun down and resuspended into a final volume of 1mL. Equal volumes of the D and R overnight cultures were mixed together to allow for conjugation and incubated overnight without agitation. An aliquot was transferred onto solid LB medium and allowed to conjugate

for another 2 days. The bacteria then were printed first onto minimal-glucose plates, in which only the *E. coli* O157:H7 bacterial cells would grow and not the donor source, followed by printing on minimal-glucose-TC plates where only the conjugated *E. coli* O157:H7 F' positive, TC resistant cells would grow. Finally, the cells were printed to LB-TC plates in which the *Salmonella* and conjugated *E. coli* O157:H7 would grow. The first and last plates were used as controls to ensure that the bacterial cells were growing successfully and to not lose the D and R cells, which neither should grow on the minimal-glucose-TC plates. All plates were incubated at 37^{o}C for 2 days until conjugants appeared on the minimal-glucose-TC plates. Isolated, conjugated colonies were restreaked for clonal expansion and then frozen in 20% glycerol-nutrient broth.

M13 Phage Infection Assay.

A frozen aliquot of the 3 conjugated isolates and the positive control JM109 each were subcultured into 5mL of LB (5μg/mL TC in the conjugate's LB) and agitated at 37^{o}C until OD_{600nm} ~ 0.5. A 10^{-6} phage dilution was made using LB. The M13 phage stock came from a culture propagated in 1985 using *E. coli* KK2186. One hundred μL of bacteria were gently mixed with 100μL of the phage dilution and incubated at room temperature for 2-3 minutes. Three mL of warm soft agar (5μg/mL TC for conjugants) were added to the combination, then mixed gently and poured onto prewarmed LB plates. After the soft agar hardened, the plates were inverted and incubated overnight at 37^{o}C. The number of plaques was then counted for each phage-bacteria combination. Due to individual restriction modification systems in the 3 *E. coli* isolates, each isolate must be tested with a phage that has been modified by the other isolates (11). Three 3 mL of LB were added to the overnight cultured plates of JM109 & the 3 *E. coli*s.

The plates containing the fresh LB alternated between gentle rocking and resting for 10 minutes. The LB supernatants from each plate were decanted into tubes and centrifuged at 10,000 rpms for 1 minute. The supernatant was saved as each supernatant contained phage that had been epigenetically tagged by the specific isolate. Each new specifically tagged phage was then tested on the other 3 hosts by the method described above.

Statistical Analysis.

ANOVA analysis was performed on the M13 phage assay using the statistical package found at the website: http://faculty.vassar.edu/lowry/ank3.html. The statistical significance was determined for the means of the 3 isolates with their 3 specific phages, also the statistically significance of the total means was examined.

3.6.4. Results

Identification of Potential VNTR Loci.

Of the 6 TR loci, we found 4 to have multiple alleles and the remaining loci, *tir* and *mop*A, had only 1 allele each after examining a sample of isolates (n=16) (Table 10). *Yci* and *hem*X both had 2 alleles each so we decided to not pursue their functional roles. The final 2 loci had much more interesting results: the loci located before IS3 had 9 alleles plus 8 isolates had an insert interrupting the tandem repeat. This 1.2kbp insert was sequenced and found to be a shiga toxin-2 (Stx-2) pseudogene. The final gene, *tol*A, was found to have 3 alleles resulting in 7, 9, and 13.7 repeats.

Table 10. **Characteristics of the 6 targeted tandem repeat loci.** Details included are the known functions of the genes containing or flanking the TR and the range of alleles.

TR Locus	Locus Function	Range of Alleles
hem X	Uroporphyrinogen III methylase	3-4X
Hyp-IS3[1, 2](IS3)	Hypothetical protein & Transposase	4-12X [3]
mop A	Chaperone	4X
tir	Extracellualr functions	3X
tol A	Membrane stability, colicin/phage infection	7,9,13.7X [4]
yci D-*int* O[1](yci)	Putative enzyme & Integration	3-4X

[1]TR locus is inbetween 2 open reading frames

[2]Hyp-Hypothetical open reading frame

[3]Some isolates have a 1.2kbp insert

[4]Discrete alleles, not a range

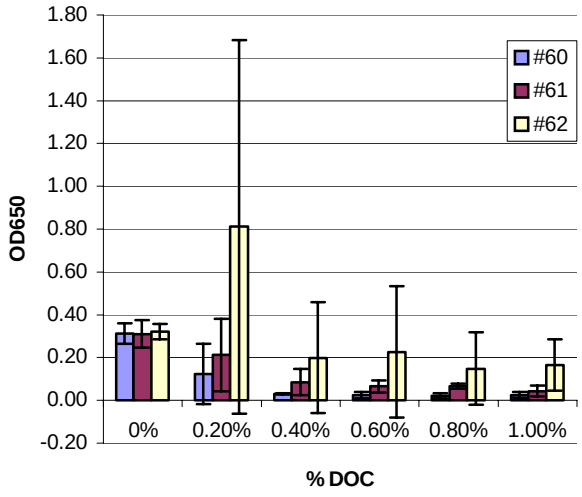
Growth in Liquid Medium in Increasing Amounts of DOC.

Three independent experiments were performed evaluating the growth of our 3 conjugated *tol*A allele isolates, PHIDL #60 with 7 repeats, PHIDL #61 with 9 repeats, and PHIDL #62 with 13.7 repeats, in the presence of increasing amounts of DOC. This assay was an indirect measure of tolA's role on the structural stability of the cell. After 3 hours of coincubation, the OD_{650nm} was measured although there was interference due to the cloudy and proteinacious material in some of the cultures (Figure 13). PHIDL #62 had the highest OD readings, and typically had the largest amount of gelatinous material in the culture suggesting growth and then lysis of the cells. One hundred μL of each bacteria-specific DOC percentage were aliquoted onto blood plates. PHIDL #60 consistently had the least amount of CFUs per plate followed by PHIDL #61 then PHIDL #62. This pattern remained consistent even as the percentage of DOC increased (Figure 14).

Growth on Solid Medium with Increasing Amounts of DOC.

This assay also was an indirect measure of the role of tolA in cell wall stability. The growth patterns on solid DOC medium differed from that of the liquid medium. PHIDL #60 had the highest level of growth against the 0% control compared to the other 2 isolates (Figure 15). PHIDL #62, which had the highest level of growth in the liquid-DOC coincubation, had a growth pattern between the other two isolates.

A)



B)

Figure 13. **Typical results after isolates exposed to increasing concentrations of DOC.** A) Overall, loss of turbidity with increasing DOC. Increased turbidity seen in tubes containing PHIDL #62 compared to the other isolates. B) Sample of growth after 3 hours coincubation of samples with 0.6% DOC; with the order of tubes from left to right being PHIDL #60, PHIDL #61, and then PHIDL #62. Increasing amounts of potentially proteinacious material were seen in PHIDL #62 tubes.

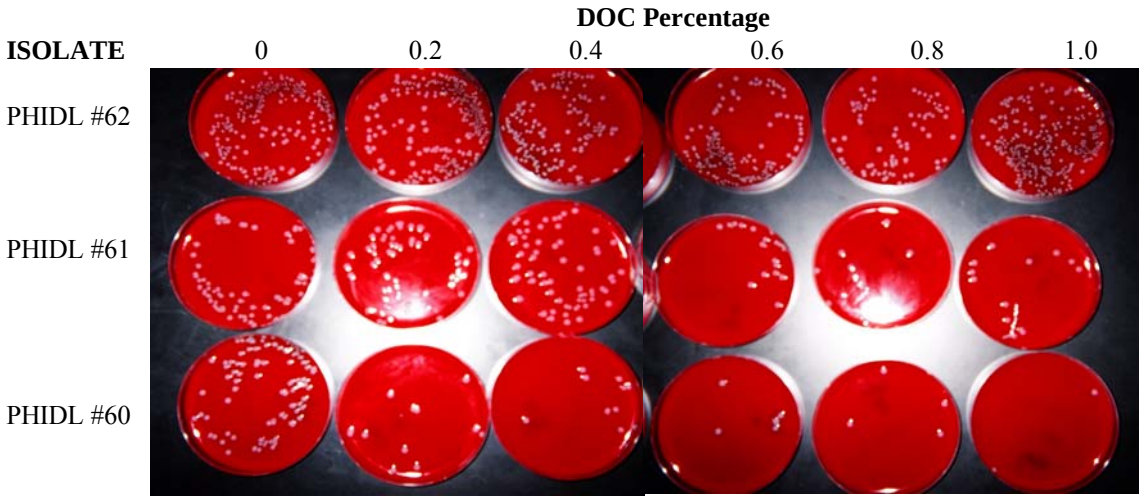


Figure 14. **CFU count of each isolate-DOC combination.** Liquid cultures of isolates with increasing amounts of DOC were incubated together for 3 hours and then plated overnight. Typically, more colonies were seen with lower concentrations of DOC than with the higher percentages. PHIDL #62 had the highest amount of growth than the other isolates in the presence of DOC.

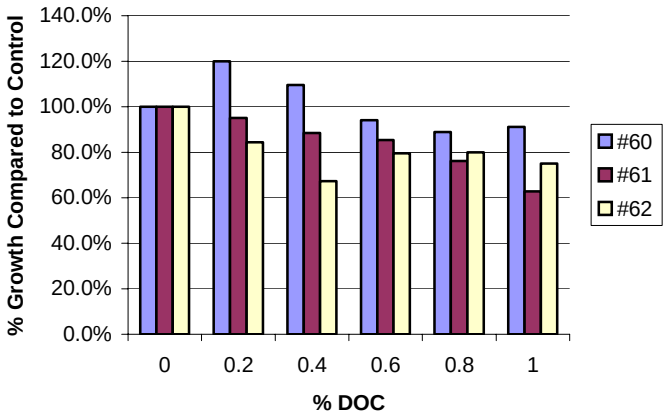


Figure 15. **Percent growth of *E. coli* O157:H7 isolates compared to control on LB agar containing increasing concentrations of DOC.** The average number of CFU/plate for each isolate was compared to growth on the 0% DOC plate.

Bacterial Killing Assay.

The ability of the peptide WLBU2 to kill the 3 *tol*A study isolates and a control isolate, *Pseudomonas aeruginosa* was measured by counting the number of CFUs. The control *Pseudomonas* had a known susceptibility to the WLBU2 peptide at the concentrations tested. We were able to determine if the assay was working correctly by looking at the susceptibility curve of the control bacterium. As the WLBU2 peptide forms pores in cell membranes, this assay was the final indirect measure of tolA's role in maintaining structural stability. Overall, the *E. coli* O157:H7 isolates had at least a 3-fold higher resistance to the peptide compared to the *P. aeruginosa.* While not significant, PHIDL #60 had the highest resistance to the peptide and PHIDL #61 had the least resistance (Figure 16).

Sequence Results of the Genes Associated with M13 Phage Infection.

A direct measure of the role of *tol*A was conducted by looking at the infectivity of M13 phage in *E. coli* O157:H7. The genes associated with M13 infection were sequenced to determine if only *tol*A was different between the 3 isolates. All the genes in the *tol*QRAB operon and the receptors for phage M13 were identical to each other in the 3 isolates except for 2 single nucleotide polymorphisms (SNPs) found in *omp*A. These SNPs were found previously in our MLST study (10). Both PHIDL #61 and PHIDL #62 each contained a SNP but both were synonymous changes, so the amino acid sequence did not change.

Conjugation of *E. coli* O157:H7.

The 3 isolates were successfully conjugated with the F' plasmid, which conferred tetracycline resistance from the *Salmonella* isolate, although at different levels of success. PHIDL #61 had

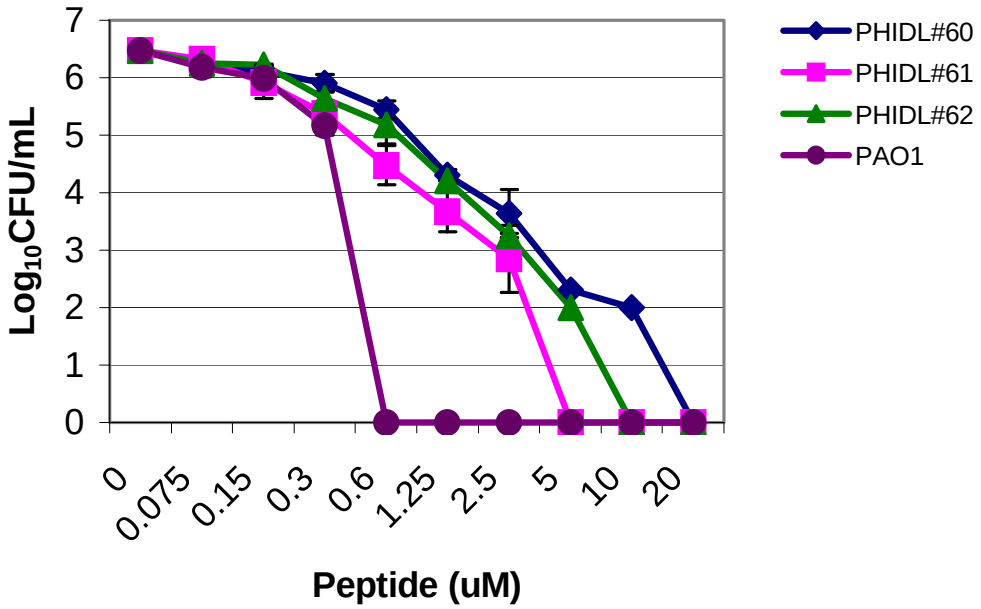


Figure 16. **Bacterial killing after 30 minute exposure to the synthetic peptide WLBU2.** All *E. coli* O157 isolates were more resistant than the control *Pseudomonas* strain (PAO1) to the peptide. PHIDL #60 had a slightly higher resistance to the peptide than the other *E. coli* O157:H7 strains.

the most conjugants followed by PHIDL #60, then PHIDL #62. After conjugation, the *E. coli* O157:H7 isolates had the necessary pilus assembly to be infected by M13. The sex pilus is the coreceptor for M13 attachment to the bacterium.

M13 Phage Assay.

Isolates were infected by each of the 4 specific epigenetically-tagged M13 phage. Plaques were counted for each isolate-phage mix and the counts normalized to the JM109 counts. When looking at specific phages, the isolate that was most susceptible for each phage varied from phage to phage (Table 11). Comparing each isolate to its matched phage showed that isolate #60 had the greatest number of plaques compared to the others, and this was statistically significant ($p=0.004$). When the mean total count of plaques per isolate was calculated, once again PHIDL #60 differed by the others, but a conflicting result arose showing that the total count of PHIDL #60 was not significantly less than the other 2 isolates' counts ($p=0.13$).

3.6.5. Discussion

Variable-number tandem repeats are fascinating genetic elements in both the prokaryotic and eukaryotic genomes. Certain VNTRs in humans cause a wide array of genetic disorders while some VNTRs in bacteria are thought to modulate the immune response (14). *E. coli* O157:H7 also has a wide array of VNTRs, but their roles are undefined. We analyzed several tandem repeats to determine if they were indeed VNTRs and if we could begin to understand their possible functions.

Table 11. **Plaque counts for each isolate-phage combination normalized to the JM109 counts.**

		Phage Grown In:				
Isolate		JM109	60	61	62	Total
60	Mean	190.1	**353**[a]	96.3	132.0	**771.3**[b]
	StDv	33.6	24.0	27.2	28.7	59.1
61	Mean	216.5	463.0	**107.25**[a]	168.8	**955.5**[b]
	StDv	14.9	46.7	35.0	9.0	57.8
62	Mean	247.5	448.0	82.5	**143.4**[a]	**921.4**[b]
	StDv	14.0	62.2	7.8	1.8	82.2

[a]ANOVA analysis (p=0.004)

[b]ANOVA analysis (p=0.13)

We chose to focus on 6 potential VNTRs, as they were located within or between known or highly homologous genes (Table 9). Four of the loci were true VNTRs when we sequenced the loci in at least 16 different isolates. We then chose to focus on 2 loci because they had multiple alleles.

The VNTR loci located before IS3 is a seven-basepair repeat that was found to repeat 4 to 12 times in our collection. Of the 90 isolates that we sequenced, 8 isolates (8.9%) were found to have a 1.2kbp insert that encodes a Stx2 pseudogene. This insertion interrupted the tandem repeat in the middle of the sequence and at the same repeat for all 8 isolates. At the end of the pseudogene, the tandem repeat continued and the sequence was identical to that of the isolates without the insertion. Additionally, 6 of the 8 isolates were related via PFGE, but only 2 of these were found to be related both temporally and by MLVA. This suggests that the pseudogene may have integrated into different genomes at separate, evolutionary events and that this tandem repeat sequence may be ideal sites for integration.

The other locus of interest was that of *tol*A, which contained a 15bp tandem repeat in the central portion of the gene. TolA has 3 roles in the *E. coli* genome that of cell membrane stability, filamentous phage infection, and group A colicin sensitivity (2, 6, 7). Upon sequencing our isolates, we found 3 discrete alleles in which the TR was found to repeat 7, 9, and 13.7 times. We hypothesized that the increasing length of the TR in *tol*A would result in an increased stability of the cell, but an enhanced sensitivity to filamentous phages.

One of the roles of tolA is that of cell membrane stability as tolA is anchored in the inner membrane, spans across the periplasmic space, and interacts with the inner portion of the outer membrane (7, 8). We chose to examine the role of tolA in membrane stability through 2 indirect assays: growth in deoxycholate and exposure to a synthetic antimicrobial peptide. These are indirect assays because these compounds indiscriminately attack the bacteria. The effect of DOC was measured 2 ways: the bacteria were exposed to DOC in a liquid culture for 3 hours or the bacteria were plated directly on hard agar impregnated with DOC. In the liquid media, PHIDL #62 had the highest level of growth in the presence of DOC (Figures 13, 14). PHIDL #62 was the isolate with the longest *tol*A α-helix, while PHIDL #60 with the shortest α-helix had the lowest amount of growth. The assay using DOC in solid agar had discordant results from that of the liquid assay. PHIDL #60 had the best growth in DOC followed by PHIDL #62 and then PHIDL #61 (Figure 15). The difference between the solid and liquid assays may be completely independent of *tol*A. Different genes are turned on or off depending on the medium used, so the different results could be due to this, which leaves the role of *tol*A unanswered.

The other indirect assay used the antimicrobial peptide WLBU2 to again examine the role of *tol*A in membrane stability. This cationic peptide inserts itself into bacterial cell membranes resulting in pore formation and lysis of the bacterium. The results were similar to that of the solid agar-DOC assay. PHIDL #60 had the highest level of resistance to the peptide resulting in the most growth with higher concentrations of the peptide than the other isolates (Figure 16). PHIDL #62 followed with the next highest resistance to the peptide. These results along with the solid agar-DOC assay may suggest that length is not necessarily the main determinant for *tol*A's role in cell membrane stability, but perhaps how the rotation and position of the α-helix against

the outer membrane. Protein chemistry and structure are beyond the scope of this study, so only observations and ideas can be postulated about the results that have been observed.

The M13 phage infectivity assay was the direct assay to determine if the alleles of *tol*A would result in a difference in infectivity between the 3 isolates. This was a direct assay because M13 must use *tol*A and a few accessory proteins to enter into the host, along with the sex pilus encoded for on the F' plasmid. As previously discussed in the results, the 3 isolates had identical accessory protein sequences, besides for *omp*A. The 3 isolates also contained the same F' plasmid, which was supplied during the conjugation assay. Another factor that could influence the infection level is how a phage is modified upon replicating in the host cell and then the this modified phage is introduced into another host *E. coli*. We attempted to control this confounder by having the M13 phage propagate using all 4 hosts and then testing those modified phages on the other 3 isolates and the cognate host. So while the 3 *E. coli* O157:H7 isolates are not identical to each other, we hope that we have controlled for the extraneous variation important to M13 infection. The assay showed us that PHIDL #60 is different from the other two isolates and the *tol*A is most likely the reason for this. Little more can be said about the difference of PHIDL #60's *tol*A because the 3 *E. coli* O157:H7 isolates are not isogenic strains. While as many differences were controlled for as possible, these are wildtype stains and hence other factors may be playing a role.

A comparison of all the results for tolA shows a conflict not only between assays that should have similar results, but also the original hypothesis (Table 12). The DOC experiments have contrasting results, which may be due to the technical issues in the liquid experiment (eg.

Table 12. **Compilation of tolA functionality results.** All assays performed to ascertain the change of function due to the different alleles of tolA. The results are presented as short, medium, and long which refers to the length of the VNTR of the isolate that was most, medium, or least susceptible to the particular assay.

	Susceptibility		
Assay	Most	Medium	Least
Liquid DOC[1]	Short	Medium	Long
Solid DOC	Med/Long	Med/Long	Short
Antimicrobial Peptide	Medium	Long	Short
Restriction-Modified Phage[2]	Short	Med/Long	Med/Long
Total Phage	Med/Long	Med/Long	Short

[1]Technical Issues with Assay

[2]Statistically Significant (p=0.004)

protenacious material). If the liquid DOC experiment is removed, then the results of the 2 indirect assays do correspond, but this result opposed the original hypothesis, which was that the isolate with the shortest tolA allele would be the most susceptible. The direct assay looking at susceptibility of the bacteria to M13 phage also had conflicting results. The only significant finding was that PHIDL #60, the isolate with the shortest allele, was the most susceptible to the phage. Again, this result went against the original hypothesis that the isolate with the longest allele would be the most susceptible to the M13 phage. The tolA protein in PHIDL #60 appears to function differently from the other 2 alleles, but how or what this implies has not been determined during this investigation. As mentioned previously, isogenic strains are the next step to determine if these same differences occur in strains that are identical except for their *tol*A. Additionally protein modeling could be done to determine if there are conformational differences between the 3 proteins that might affect function.

This set of preliminary work suggests that some of *E. coli* O157:H7's tandem repeats may have functional roles. Some TRs may be signals for insertion of foreign or virulence genes as seen with the TR loci before IS3 where a Stx2 pseudogene was inserted. Additionally, an active functional change may be assigned to other TRs as potentially seen with the gene *tol*A with an apparent difference between the tolA of PHIDL #60 with the smallest repeat to the other 2 larger repeats found in the other isolates. Again, the best way to address functional changes would be to have isogenic strains in which the only difference exists in the TR of interest. Care must be given when using VNTRs for epidemiological work versus VNTRs that have functional roles. The VNTRs for epidemiological surveillance should mutate neutrally. This work represents the background work into determining the roles of interesting VNTRs in *E. coli* O157:H7.

3.6.6. Literature Cited

1. **Benson, G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.

2. **Bouveret, E.**, **A. Rigal, C. Lazdunski, and H. Benedetti.** 1998. Distinct regions of the colicin A translocation domain are involved in the interaction with TolA and TolB proteins upon import into *Escherichia coli.* Mol. Microbiol. **27:** 143-157.

3. **Deslouches, B., S.M. Phadke, M. Cascio, K. Islam, R. C. Montelaro, T.A. Mietzner.** 2004. *De novo* generation of cationic antimicrobial peptides: Influence of length and tryptophan substitution on antimicrobial activity. Antimicrobial Agents and Chemotherapy. In press.

4. **Germon, P., M.C. Ray, A. Vianney, J.C. Lazzaroni.** 2001. Energy-dependent conformational change in the TolA protein of *Escherichia coli* involves its N-terminal domain, TolQ, and TolR. J. Bacteriol. **183:** 4110-4114.

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

6. **Karlsson, F., C.A. Borrebaeck, N. Nilsson, A.C. Malmborg-Hager.** 2003. The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. J. Bacteriol. **185:** 2628-2634.

7. **Lazzaroni, J.C., P. Germon, M.C. Ray, and A. Vianney.** 1999. The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. FEMS Microbiol. Lett. **177:** 191-197.

8. **Levengood, S.K., W.F. Beyer, Jr., R.E. Webster.** 1991. TolA: A membrane protein involved in colicin uptake contains an extended helical region. Proc. Natl. Acad. Sci. USA. **88:**5939-5943.

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

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

11. **Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink,**

A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Ananthara-man, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.

12. **Sambrook, J. and D.W. Russell**. 2001. Working with bacteriophage M13 vectors. molecular cloning: A laboratory manual. J. Argentine. Cold Spring Harbor, Cold Spring Harbor Laboratory Press. **1**: 3.1-3.49.

13. **Richards, R.I. and G.R. Sutherland.** 1997. Dynamic mutation: possible mMechanisms and significance in human disease. **22:** 432-436.

14. **van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh.** 1998. Short-sequences DNA repeats in prokaryotic genomes. Micro. Mol. Biol. Rev. **62:** 275-293.

4. DISCUSSION

Escherichia coli O157:H7 was first detected in 1982 in association with a food-borne outbreak (5). Now, this bacterium is thought to cause an estimated 74,000 illnesses a year in the United States alone (2). With the high number of cases and the association with outbreaks, the detection and genetic relatedness of this organism became a high priority. The use of PFGE for *E. coli* O157:H7 outbreak detection greatly enhanced the surveillance of this organism along with the traditional epidemiological methods (1).

In Specific Aim I, we wanted to create a molecular subtyping technique for detecting outbreaks that was as discriminatory as PFGE, if not better, but did not have the drawbacks of PFGE: time consuming protocol, subjective interpretations, etc. MLST and MLVA both have the advantage of being PCR-based. As we saw, MLST did not provide enough variation to be an effective submolecular typing scheme most likely due to *E. coli* O157:H7 being a clonal species and MLST targeting housekeeping genes (3). Conversely, MLVA utilizes tandem repeats, which are one of the most rapidly evolving genetic elements in the bacterial genome. We found that MLVA was able to discriminate outbreak strains from sporadic strains and strains from different outbreaks (4). Additionally, MLVA discriminated some strains that PFGE had been unable to which becomes important if an outbreak investigation would begin because of the molecular subtyping results.

MLVA can become a powerful tool in the surveillance and detection of *E. coli* O157:H7 outbreaks with its fast, reproducible, and objective benefits. We are hoping to create a website where investigations can import the MLVA types of their isolates of interest. They can then

compare their isolates to other isolates in the database to look for potential outbreaks. As the MLVA protocol takes 2-3 days and the Internet comparison takes a matter of moments, our ability to detect potential outbreaks will be enhanced greatly. The next step will be how we involve public health departments into our MLVA website so that our MLVA results have consequences in the real world.

MLVA does have a major limitation: we use expensive dyes and sequencing machines to obtain the results. For those laboratories with limited funds or equipment, our current MLVA protocol is not appropriate, but an alternative does exist. While determining which VNTRs were appropriate for our MLVA protocol, we chose to ignore those VNTRs with large repeat sizes (>50bp). These alleles caused a problem with running them on a sequencing gel because the range could become too large and could overlap with multiple other loci. These larger VNTRs could be ideal for an agarose gel-based system where the VNTR PCR products could be run and loci visually determined using an adequate ladder. No work has begun on this protocol, but this would offer an alternative method, but still discriminatory, that would be more feasible in costs for some facilities.

With the establishment of our MLVA protocol, we needed to apply some basic guidelines for the calling of isolates' genetic relatedness. The basis of MLVA is that these TR loci are in fact VNTRs resulting in multiple alleles at each locus. What is important to understand is how often do these VNTRs mutate from one allele to the next. In Specific Aim II, we attempted to answer this question by analyzing our 7 VNTRs after multiple generations of growth. We found that TR2 was the most hypervariable locus and that there were more single, additions of a repeat than

any other addition or deletion. Additionally, TR1 and TR5 had a moderately higher rate of mutation compared to the other 4 loci. These preliminary mutation results will help us to evaluate the MLVA data. A group of suspected, outbreak isolates will typically have the same MLVA type; but as we have seen before, single locus variants can occur and at any locus (4). Our mutation data suggest that a single change at the TR2 locus during an outbreak will be very common but a double or triple repeat change could also occur. Additionally, this TR2 hypermutability could result in a potential double locus variant as another locus could mutate.

While the mutation experiment provided us with preliminary guidelines on how to call isolates related or not, much more work needs to be done. This experiment was performed using a single isolate, which had close to the median allele for each locus. To more fully understand how each locus mutates and how that affects the calling of outbreaks, more mutation studies need to be performed. Alleles at the extremes of a range should be analyzed to understand how these extremes affect the mutation rate for each locus. More analysis will only strengthen our definitions of highly related versus sporadic isolates defined by their MLVA types.

Finally, Specific Aim III began the work of examining the functional roles of specific VNTRs. The original goal was to look at the MLVA loci and determine if there was a functional role that could potentially affect how or why they mutate, which could in turn affect their epidemiologic role. Unfortunately, the 7 MLVA are all within or between unknown or hypothetical open reading frames. The discovery of the function or lack of function of these 7 loci was beyond the scope of our study; therefore we chose to focus on a few VNTRs with known genes. We saw that one of the VNTR loci could be a hotspot for recombination events with an insertion of a

toxin pseudogene. There is a potential for this locus to recombine with an active gene that could make changes to the phenotype of the *E. coli* O157:H7 host. If VNTRs can become hotspots for recombination, there is a chance that some of the MLVA loci could also become targets for an insertion. Obviously, this would greatly change the size of the VNTR, which would in turn change the MLVA type. Researchers using the MLVA protocol should be mindful of an isolate with an allele outside the normal distribution and perhaps should sequence this allele to be sure the aberrant size is due only to the number of repeats and not some other event.

The other VNTR of interest was the one found in the gene *tol*A. Three alleles were found with PHIDL #60 containing the shortest allele; PHIDL #61, the middle allele; and PHIDL #62, the longest allele representing the 3 alleles found in our entire isolate bank. The indirect assays using DOC and an antimicrobial peptide were used to show tolA's role in cell stability. Two of the 3 assays suggested that PHIDL #60 with the shortest allele was the most resistant to DOC and the peptide. This finding contrasts our hypothesis that the shorter tolA allele would be the most susceptible to these agents. The direct measure of tolA was looking at the susceptibility of the 3 tolA alleles to the M13 phage. After analyzing the data 2 different ways, the results conflicted in that the short tolA was the most susceptible to its own modified phage, but overall was the least susceptible to all the modified phages. This conflicting result once again shows that PHIDL #60, the shortest allele, is different from the other 2 alleles. Perhaps the difference between the short allele and the other 2 is due to a conformational difference between the short tolA and the other 2 alleles. PHIDL #60's central domain containing the VNTR may end in a way that causes the C-terminus to interact differently with the outer membrane than that of the other 2 alleles. But much more research needs to be done to answer this question including

analyzing the 3 *tol*A alleles in an isogenic strain. This may allow a more conclusive analysis if this VNTR locus has a functional role. Additionally, structural analysis of the 3 tolA alleles would help to determine if the conformation of the protein changes depending on the length of the tandem repeat, which then may affect the functionality of the protein. Other researchers have found functional purposes for VNTRs in other bacteria, and it is only a matter of time and more experiments to discover the role of specific VNTRs in *E. coli* O157:H7.

Overall, this project has improved our outbreak detection capabilities for *Escherichia coli* O157:H7 and has begun to understand how variable-number tandem repeats impact their genotypic and phenotypic functions of the genome. We have greatly enhanced the ability to detect outbreaks, which will affect any person who could come in contact with *E. coli* O157:H7. Our faster and more sensitive protocol has the potential to identify outbreaks much earlier possible resulting in fewer illnesses and protecting the public's health. The continuing investigation of the behavior of these VNTRs will help to develop our guidelines for the optimal interpretation of MLVA data obtained during outbreak investigations.

4.1. Literature Cited

1. **Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm.** 1997. Surveillance for *Escherichia coli* O157:H7 infections in Minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.

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

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

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

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

APPENDIX

3100 DNA ANALYZER SCREEN SHOWING MLVA ISOLATES

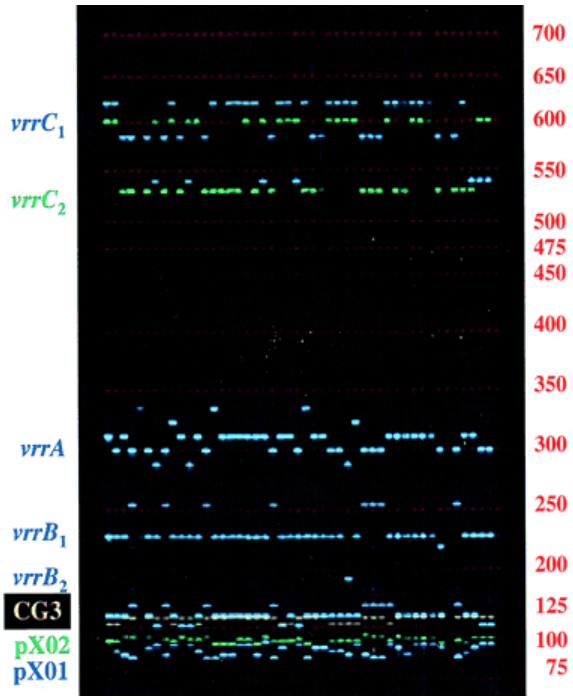


Figure 17. **The figure above represents 48 isolates of *Bacillus anthracis* that have been examined at 8 TR loci.** The labeled PCR products were run on a 3100 DNA Analyzer (Keim 2000). Our MLVA analysis differs in that we used a 3700 DNA Analyzer. Fluorescent-labeling of PCR products and creation of discrete ranges (if possible) allows each TR locus to have its own unique color and range. This allows for simple, objective analysis of the data.

BIBLIOGRAPHY

Ballous, F. and N. Lugon-Moulin. 2002. The estimation of population differentiation with microsatellite markers. Mol. Evol. **11:**155-165

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

Bender, J. B., C. W. Hedberg, J. M. Besser, D. J. Boxrud, K. L. MacDonald, and M. T. Osterholm. 1997. Surveillance for *Escherichia coli* O157:H7 infections in minnesota by molecular subtyping. N. Engl. J. Med. **337:**388–394.

Benson, G. 1999. Tandem Repeats Finder: A program to analyze DNA sequences. Nucleic Acids Res. **27:**573–580.

Besser, R. E., P. M. Griffin, and L. Slutsker. 1999. *Escherichia coli* O157:H7 gastroenteritis and the hemolytic uremic syndrome: An emerging infectious disease. Annu. Rev. Med. **50:**355–367.

Bettelheim, K.A. 2003. Non-O157 verotoxin-producing *Escherichia coli*: A problem, paradox, and paradigm. Exp. Biol. Med. **228**:333-344.

Bouveret, E., **A. Rigal, C. Lazdunski, and H. Benedetti.** 1998. Distinct regions of the colicin a translocation domain are involved in the interaction with TolA and TolB proteins upon import into *Escherichia coli.* Mol. Microbiol. **27:** 143-157.

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

Brooks, J.T., D. Bergmire-Sweat, M. Kennedy, K. Hendricks, M. Garcia et al. 2004. Outbreak of shiga toxin-producing *Escherichia coli* O111:H8 infections among attendees of a high school cheerleading camp. Clin. Inf. Dis. **38**:190-198.

Centers for Disease Control and Prevention. 1993. Update: multistate outbreak of *Escherichia coli* O157:H7 infections from hamburgers—Western United States, 1992–1993. Morb. Mortal. Wkly. Rep. **42:**258–263.

Cozzolino, S., D. Cafasso, G. Pellegrino, A. Musacchio, and A. Widmer. 2003. Molecular evolution of a plastid tandem repeat locus in an orchid lineage. J. Mol. Evol. **57:**S41-S49.

Di Rienzo, A., A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, and N.B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA. **91:**3166-3170.

Eckert, K.A. and G. Yan. 2000. Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: Influence of sequence on expansion mutagenesis. Nucl. Acids Res. **28:**2831-2838.

Enright, M. C., N. P. J. Day, C. E. Davies, S. J. Peacock, and B. G. Spratt. 2000. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus.* J. Clin. Microbiol. **38:**1008–1015.

Ewing, B., and P. Green. 1998. Base-calling of automated sequencer traces using Phred. II. Error Probabilities. Genome Res. **8:**186–194.

Ewing, B., L. Hillier, M. C. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. Genome Res. **8:**175–185.

Farlow, J., K.L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim. 2001. *Francisella tularensis* strain typing using multiple-locus variable-number tandem repeat analysis. J.Clin. Micr. **39:**3186-3192.

Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt. 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics. **154:**1439–1450.

Field, D. and C. Wills. 1998. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellite in eight prokaryotes and *S. cerevisiae* result from strong mutation pressures and a variety of selective forces. Proc. Natl. Acad. Sci. USA. **95:**1647-1652.

Freudenreich, C.H., J.B. Stavenhagen, and V.A. Zakian. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. Mol. Cell. Bio. **17:**2090-2098.

Frothingham, R., and W. A. Meeker-O'Connell. 1998. Genetic diversity in the *Mycobacterium tuberculosis* complex based on variable numbers of tandem DNA repeats. Microbiology **144**(Pt 5)**:**1189–1196.

Germon, P., M.C. Ray, A. Vianney, J.C. Lazzaroni. 2001. Energy-dependent conformational change in the TolA protein of *Escherichia coli* involves its N-terminal domain, TolQ, and TolR. J. Bacteriol. **183:** 4110-4114.

Gordon, D., C. Abajian, and P. Green. 1998. Consed: A graphical tool for sequence finishing. Genome Res. **8:**195–202.

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

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. J. Comput. Graph. Stat. **5:**299–314.

Izumiya, H., J. Terajima, A. Wada, Y. Inagaki, K.-I. Itoh, K. Tamura, and H. Watanabe. 1997. Molecular typing of enterohemorrhagic *Escherichia coli* O157:H7 isolates in Japan by using pulsed-field gel electrophoresis. J. Clin. Microbiol. **35:**1675–1680.

Jeanmougin, F., J. D. Thompson, M. Gouy, D. G. Higgins, and T. J. Gibson. 1998. Multiple sequence alignment with Clustal X. Trends Biochem. Sci. **23:**403–405.

Kang, S., A. Jaworski, K. Ohshima, and R.D. Wells. 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. Nature Gen. **10:**213-218.

Karlsson, F., C.A. Borrebaeck, N. Nilsson, A.C. Malmborg-Hager. 2003. The mechanism of bacterial infection by filamentous phages involves molecular interactions between TolA and phage protein 3 domains. J. Bacteriol. **185:** 2628-2634.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond, D. M. Adair, M. Hugh-Jones, C. R. Kuske, and P.Jackson. 1997. Molecular evolution and diversity in *Bacillus anthracis* as detected by amplified fragment length polymophism Markers. J. Bacteriol. **179:**818-824.

Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones. 2000. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. J. Bacteriol. **182:**2928–2936.

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

Kim, J., J. Nietfeldt, J. Ju, J. Wise, N. Fegan, P. Desmarchelier, and A. K. Benson. 2001. Ancestral divergence, genome diversification, and phylogeographic variation in subpopulations of sorbitol-negative, β-glucuronidase-negative enterohemorrhagic *Escherichia coli* O157. J. Bacteriol. **183:**6885–6897.

Kimura, A.C., K. Johnson, M.S. Palumbo, J. Hopkins, J.C. Boase, R. Reporter, M. Goldoft, K.R. Stefonek, J.A. Farrar, T.J. Van Giler, & D.J. Vugia. 2004. Multistate Shigellosis outbreak and commercially prepared food, United States. Emerg. Inf. Dis. **10:**1147-1149.

Kimura, M. and T. Ohtia. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite pPopulation. Proc. Natl. Acad. Sci. USA. **75:**2868-2872.

Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim. 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pesti*s genome. J. Clin. Microbiol. **39:**3179–3185.

Kotetishvili, M., O. C. Stine, A. Kreger, J. G. Morris, Jr., and A. Sulakvelidze. 2002. Multilocus sequence typing for characterization of clinical and environmental *Salmonella* strains. J. Clin. Microbiol. **40:**1626–1635.

Kudva, I. T., P. S. Evans, N. T. Perna, T. J. Barrett, F. M. Ausubel, F. R. Blattner, and S. B. Calderwood. 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. J. Bacteriol. **184:**1873–1879.

Lai Y. & F. Sun. 2003. The relationship between microsatellite slippage mutation rate and the number of repeat units. Mol Biol. Evol. **20:** 2123-2131.

Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. Proc. Natl. Acad. Sci. USA **95:**9413–9417.

Lazzaroni, J.C., P. Germon, M.C. Ray, and A. Vianney. 1999. The Tol proteins of *Escherichia coli* and their involvement in the uptake of biomolecules and outer membrane stability. FEMS Microbiol. Lett. **177:** 191-197.

Levengood, S.K., W.F. Beyer, Jr., R.E. Webster. 1991. TolA: A membrane protein involved in colicin uptake contains an extended helical region. Proc. Natl. Acad. Sci. USA. **88:**5939-5943.

Levinson G. & G.A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence eEvolution. Mol. Biol. Evol. **4:** 203-221.

Lindstedt, B., E. Heir, E. Gjernes, and G. Kapperud. 2003. DNA fingerprinting of *Salmonella enterica* subsp. *enterica* serovar Typhimurium with emphasis on phage type DT104 based on variable number of tandem repeat Loci. J. Clin. Microbiol. **41:**1469-1479.

Lovett, S.T. & V.V. Feschenko. 1996. Stabilization of diverged tandem repeats by mismatch repair: evidence for deletion formation via a misaligned replication intermediate. Proc. Natl. Acad. Sci. USA. **93:** 7120-7124.

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

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

Nallapareddy, S. R., R. W. Duh, K. V. Singh, and B. E. Murray. 2002. Molecular typing of selected *Enterococcus faecalis* isolates: pilot study using multilocus sequence typing and pulsed-field gel electrophoresis. J. Clin. Microbiol. **40:**868–876.

Nicolas, P., G. Raphenon, M. Guibourdenche, L. Decousset, R. Stor, andA. B. Gaye. 2000. The 1998 Senegal epidemic of meningitis was due to the clonal expansion of A:4:P1.9, clone III-1, sequence type 5 *Neisseria menin-gitidis* strains. J. Clin. Microbiol. **38:**198–200.

Noller, A.C., M.C. McEllistrem, and L.H. Harrison. 2004. Genotyping primers for the fully automated multi-locus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. J. Clin. Microbiol. **42:**3908.

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

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

Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Ananthara-man, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. Nature **409:**529–533.

Phillips, A D, Navabpour, S, Hicks, S, Dougan, G, Wallis, T, Frankel, G. 2000. Enterohaemorrhagic *Escherichia coli* O157:H7 target Peyer's patches in humans and cause attaching/effacing lesions in both human and bovine intestine. Gut. **47**: 377-381.

Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64–67.

Ribot, E. M., C. Fitzgerald, K. Kubota, B. Swaminathan, and T. J. Barrett. 2001. Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. J. Clin. Microbiol. **39:**1889–1894.

Richards, R.I. and G.R. Sutherland. 1997. Dynamic mutation: Possible mechanisms and significance in human disease. TIBS. **22:**432-436.

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

Samadpour, M., J. Stewart, K. Steingart, C. Addy, J. Louderback, M. McGinn, J. Ellington, and T. Newman. 2002. Laboratory investigation of an *E. coli* O157:H7 outbreak associated with swimming in Battle Ground Lake, Vancouver, Washington. J. Environ. Health **64:**16–20, 25, 26.

Sharma, R., S. Bhatti, M. Gomez, R.M. Clark, C. Murray, T. Ashizawa, and S.I. Bidichandani. 2002. The GAA triplet-repeat sequence in Friedreich Ataxia shows a high level of somatic instability *in vivo*, with a significant predilection for large contractions. Hum. Mol. Gen. **11:**2175-2187.

Stine, O. C., S. Sozhamannan, Q. Gou, S. Zheng, J. G. Morris, and J. A. Johnson. 2000. Phylogeny of *Vibrio cholerae* based on *recA* sequence. Infect. Immun. **69:**7180–7185.

Swaminathan, B., T. J. Barrett, S. B. Hunter, and R. V. Tauxe. 2001. PulseNet: The molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg. Infect. Dis. **7:**382–389.

Symonds V.V. & A.M. Lloyd. 2003. An analysis of microsatellite loci in *Arabidopsis thaliana*: mutational dynamics and aApplication. Genetics. **165:**1475-1488.

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

van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. **62:**275–293.

Viguera, E., D. Canceill, and S.D. Ehrlich. 2001. Replication slippage involves DNA polymerase pausing and dissociation. EMBO. **20:**2587-2595.

Zhou, J., M. C. Enright, and B. G. Spratt. 2000. Identification of the major spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing. J. Clin. Microbiol. **38:**977–986.

Zhu, P., A. van der Ende, D. Falush, N. Brieske, G. Morelli, B. Linz, T. Popovic, I. G. Schuurman, R. A. Adegbola, K. Zurth, S. Gagneux, A. E. Platonov, J. Y. Riou, D. A. Caugant, P. Nicholas, and M. Achtman. 2001. Fit genotypes and escape variant of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. Proc. Natl. Acad. Sci. USA **98:**5234–5239.