

**Application and Interpretation of Multiple Locus Variable Number Tandem
Repeat Analysis for *Escherichia coli* O157:H7 Laboratory Surveillance and
Outbreak Response in Canada, 2008-2012**

by

Jillian Rumore

**A Thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of**

MASTER OF SCIENCE

Department of Medical Microbiology

University of Manitoba

Winnipeg

Copyright © by Jillian Rumore

Abstract

Foodborne disease outbreaks are a major threat to public health and an increasing problem worldwide. The O157:H7 serotype of Shiga toxin-producing *Escherichia coli* is established in Canada as an enteric pathogen capable of causing large outbreaks of gastrointestinal and systemic disease. Foods contaminated with *E. coli* O157:H7 are often nationally distributed, resulting in these outbreaks, and investigations into the source and the assessment of outbreak cases requires laboratory tests to distinguish between outbreak and sporadic clinical isolates and any food isolates. To enhance these investigations, PulseNet Canada has recently applied Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) as a supplemental subtyping tool in combination with the gold standard subtyping method Pulsed-field Gel Electrophoresis (PFGE). MLVA was developed for the purposes of further resolving isolates that demonstrate indistinguishable/highly similar PFGE patterns, and the objective of the study was to assess the overall discriminatory power of MLVA and to evaluate the level of specificity MLVA offers for outbreak detection and response. A total of 2,285 *E. coli* O157:H7 isolates uploaded to the National PulseNet Canada database from 2008-2012 were selected to calculate the Simpson's Index of Diversity and Adjusted Wallace coefficient for subtype congruence compared to PFGE. Sixty-three PFGE clusters were selected to evaluate the sensitivity and specificity of adding MLVA as a secondary test. Clusters for which there were known major outbreak investigations with epidemiological links (n= 16 clusters) were selected for detailed analysis as a proxy for estimating epidemiological congruence and to recalculate the sensitivity and specificity. The results of this study demonstrate that MLVA provides a statistically significant increase in discriminatory power for cluster identification and

outbreak investigations (0.998) in comparison to PFGE alone (0.993). The AW congruence coefficients obtained between PFGE (0.227) and MLVA (0.696) were statistically different and therefore indicated these techniques are unlikely to be congruent methods and the combination of their results is likely to offer additional subtype information and discriminatory power during analyses of clusters and outbreak investigations of *E. coli* O157:H7. MLVA was able to provide additional resolution over PFGE analysis and generally agreed with PFGE when isolates were identical and epidemiologically linked. The addition of MLVA retrospectively changed the overall laboratory subtype conclusions for 60 percent of the clusters analyzed in this study and resulted in: the inclusion of isolates that were previously missed by PFGE alone in 16 percent of clusters; the exclusion of isolates that were likely erroneously included by PFGE in 29 percent of clusters, and both (16 percent of clusters). No change was observed in 25 out of the 63 clusters. Although the sensitivity (79 percent) and specificity (97 percent) obtained for the analysis of clusters was less than the sensitivity (95 percent) and specificity (99 percent) obtained for the 16 outbreaks, the values still provide strong statistical support that MLVA (when used in combination with PFGE) offers enhanced specificity to cluster/outbreak investigations without significantly impacting the sensitivity. The MLVA assay shows great promise as a molecular epidemiological tool to complement PFGE, as it improves case categorization during investigations of *E. coli* O157:H7 in Canada. The greatest benefits of utilizing MLVA as a supplemental test may be realized during routine surveillance, when epidemiological information is not available.

Acknowledgments

First and foremost I would like to express my sincerest gratitude to my supervisor, Dr. Celine Nadon, for not only providing me with the opportunity to obtain a Master of Science in Medical Microbiology, but also for her expertise, professionalism, valuable guidance, financial support, and leadership over the past two years. Her exemplary mentorship skills have greatly contributed to my personal development as a researcher. The completion of this thesis would not have been possible without her continual support and intellectual contributions, and for that, I am forever grateful. Celine has truly made this a rewarding graduate school experience. I would also like to thank my co-advisor, Dr. Matthew Gilmour and committee members, Dr. John Wylie, and Dr. Lawrence Elliott for their encouragement and insightful comments, which were integral in the success of this study. I must also express my profound gratitude to Lorelee Tschetter, the acting head of PulseNet Canada, for her extreme patience, inspiring words, and constant support throughout my graduate studies. Despite her demanding schedule, she never failed to provide an answer to my countless questions or lend a helping hand no matter the task or circumstance. I do not have the words to thank her enough for all that she has done for me. I cannot forget to thank the PulseNet Canada team for their knowledge and expertise, as well as for providing me with the necessary training to conduct my research. I'd like to give a special thanks to Ashley Kearney, Keri Trout-Yakel and Chelsey Goodman for their tremendous assistance and most importantly for keeping the laboratory environment fun and enjoyable. Finally, I must thank my family and friends for providing continuous encouragement throughout the process of researching and writing this thesis. This accomplishment would not have been possible without their support. Thank you.

Table of Contents

Chapter 1: Introduction.....	1
1.1 The Global Concern of Foodborne Illness.....	2
1.2 Epidemiology of <i>Escherichia coli</i> O157:H7 Infections in Humans.....	6
1.2.1 Burden of Illness Attributed to <i>Escherichia coli</i> O157:H7 in Canada.....	8
1.2.2 Reservoirs of <i>Escherichia coli</i> O157:H7.....	10
1.2.3 Routes of Transmission for <i>Escherichia coli</i> O157:H7.....	11
1.3 Features of <i>Escherichia coli</i> O157:H7 – Biology and Disease.....	17
1.3.1 Clinical Manifestations.....	17
1.3.2 Pathogenesis.....	19
1.3.3 Treatment and Prevention.....	20
1.3.4 Factors Contributing to Virulence.....	21
1.3.4.1 The Locus of Enterocyte Effacement.....	21
1.3.4.2 pO157 plasmid.....	22
1.3.4.3 Shiga Toxins.....	23
1.4 Isolation and Identification of <i>Escherichia coli</i> O157:H7.....	24
1.4.1 Selective Media used for the Isolation of <i>Escherichia coli</i> O157:H7.....	25
1.4.2 Biochemical Tests.....	27
1.4.3 Serotyping.....	27
1.4.4 Phagotyping.....	29
1.5 Characterization of <i>Escherichia coli</i> O157:H7 Beyond the Phenotypic Level.....	29
1.5.1 Detection of Shiga Toxins.....	30
1.5.2 Pulsed-field Gel Electrophoresis.....	35
1.5.3 Alternative DNA-based Molecular Subtyping Techniques.....	38
1.5.4 Multiple Locus Variable Number Tandem Repeat Analysis.....	40
1.6 <i>Escherichia coli</i> O157:H7 Surveillance and Outbreak Response in Canada.....	45
1.6.1 National Enteric Surveillance Program.....	46
1.6.2 PulseNet Canada.....	47
1.6.3 Investigations of Outbreaks.....	51
1.6.4 Limitations of the Current System.....	56
1.7 Current Study.....	59
Chapter 2: Materials and Methods.....	60
2.1 Isolate Collection.....	61
2.2 Multiple Locus Variable Tandem Repeat Analysis.....	61
2.3 Evaluation of the Discriminatory Power and Congruence of PFGE and MLVA.....	65
2.4 Selection of <i>Escherichia coli</i> O157:H7 Clusters for MLVA-Enhanced Retrospective Analysis.....	67
2.5 MLVA-Enhanced Retrospective Analysis.....	67
2.6 Evaluation of the Sensitivity and Specificity of MLVA as a Supplemental Test for Outbreak Investigations.....	71

Chapter 3: Results	72
3.1 Coverage of <i>Escherichia coli</i> O157:H7 Isolates Obtained for Analysis	73
3.2 Evaluation of Simpson's Index of Diversity and the Adjusted Wallace Coefficient	73
3.3 MLVA-Enhanced Retrospective Analysis of <i>Escherichia coli</i> O157:H7 Clusters	78
3.3.1 2008 Clusters	78
3.3.2 2009 Clusters	84
3.3.3 2010 Clusters	91
3.3.4 2011 Clusters	96
3.3.5 2012 Clusters	103
3.4 Variability in Related and Unrelated MLVA Profiles	111
3.5 Performance of MLVA as a Supplemental Method for Cluster and Outbreak Investigations	111
3.6 Sensitivity and Specificity of MLVA When Used in Combination with PFGE	116
Chapter 4: Discussion	118
4.1 Discriminatory Power of MLVA	119
4.2 Level of Congruency between PFGE and MLVA	120
4.3 Impact of MLVA on Cluster Detection	120
4.4 MLVA Offers Enhanced Specificity to Cluster Detection and Outbreak Investigations	122
4.5 Conclusion	125
References	127

List of Tables

Table 1. Molecular Based Laboratory Techniques used to Genotypically Characterize <i>Escherichia coli</i> O157:H7 (with the exception of Vero Cell Assay).....	32
Table 2. Characteristics of Eight Variable-Number Tandem Repeats (VNTRs) used in the Standardized Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) Protocol for <i>Escherichia coli</i> O157:H7 at PulseNet Canada.....	43
Table 3. PulseNet Canada Interpretation Guidelines for Determining the Level of Relatedness between Food and Clinical Isolates based on Pulsed-field Gel Electrophoresis (PFGE) Results.....	52
Table 4. PCR primers utilized in the PulseNet Canada/International <i>Escherichia coli</i> O157:H7 standardized MLVA Protocol.....	63
Table 5. Comparative analysis of the Simpson's Index of Diversity for two subtyping methods, Pulsed-field Gel Electrophoresis and Multiple-locus Variable Number Tandem Repeat Analysis, utilized to characterize <i>Escherichia coli</i> O157:H7 isolates for laboratory surveillance and outbreaks in Canada.....	74
Table 6. Analysis of the Adjusted Wallace coefficient for two subtyping methods, Pulsed-field Gel Electrophoresis and Multiple-locus Variable Number Tandem Repeat Analysis, used to characterize <i>Escherichia coli</i> O157:H7 isolates for laboratory surveillance and outbreaks in Canada.....	76
Table 7. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 10 <i>Escherichia coli</i> O157:H7 clusters that occurred in 2008 as reported by PulseNet Canada.....	81
Table 8. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 12 <i>Escherichia coli</i> O157:H7 clusters that occurred in 2009 as reported by PulseNet Canada.....	87
Table 9. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 10 <i>Escherichia coli</i> O157:H7 clusters that occurred in 2010 as reported by PulseNet Canada.....	94
Table 10. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 15 <i>Escherichia coli</i> O157:H7 clusters that occurred in 2011 as reported by PulseNet Canada.....	99

Table 11. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 16 <i>Escherichia coli</i> O157:H7 clusters that occurred in 2012 as reported by PulseNet Canada.....	106
Table 12. Summary of the isolates included in 63 clusters of <i>Escherichia coli</i> O157:H7 based on the original real-time findings (PFGE alone) in comparison to those isolates identified as cluster related following MLVA-enhanced retrospective analysis (PFGE and MLVA).....	112
Table 13. Comparative analysis of the sensitivity and specificity of MLVA when used as a supplemental method in combination with the current gold standard PFGE for 63 clusters (11a) and 16 well-characterized outbreaks (11b) of <i>Escherichia coli</i> O157:H7 that occurred in Canada from 2008-2012.....	114

List of Figures and Illustrations

Figure 1. Model depicting possible routes for transmission of <i>Escherichia coli</i> O157:H7 from the bovine host to humans.....	12
Figure 2. Flowchart detailing the inclusion criteria and characteristics of <i>Escherichia coli</i> O157:H7 clusters selected for MLVA-enhanced retrospective analysis.....	68
Figure 3. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 142 <i>Escherichia coli</i> O157:H7 isolates (clinical and food) associated with 10 retrospective clusters that occurred in 2008 as reported by PulseNet Canada.....	79
Figure 4. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 94 <i>Escherichia coli</i> O157:H7 isolates (clinical and food) associated with 12 retrospective clusters that occurred in 2009 as reported by PulseNet Canada.....	85
Figure 5. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 47 <i>Escherichia coli</i> O157:H7 isolates (clinical and food) associated with 10 retrospective clusters that occurred in 2010 as reported by PulseNet Canada.....	92
Figure 6. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 108 <i>Escherichia coli</i> O157:H7 isolates (clinical and food) associated with 15 retrospective clusters that occurred in 2011 as reported by PulseNet Canada.....	97
Figure 7. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 183 <i>Escherichia coli</i> O157:H7 isolates (clinical and food) associated with 16 retrospective clusters that occurred in 2012 as reported by PulseNet Canada.....	104

Chapter 1: Introduction

1.1 The Global Concern of Foodborne Illness

According to the Foodborne Illness Outbreak Response Protocol (FIORP), a foodborne illness outbreak is defined as “an incident in which two or more persons, from different households and therefore not linked, experience similar illness after a common source of exposure. An outbreak is often identified through laboratory surveillance or other surveillance mechanism demonstrating an increase in illness that is unusual in terms of time and/or place. An outbreak is confirmed through laboratory, food safety and/or epidemiological evidence” (Public Health Agency of Canada, 2010). The explosive increase in worldwide trade of raw and processed foods, in addition to travel across international borders have made it possible for human foodborne infections to originate in a region different from where the illness is observed (Barrett *et al.*, 2006). Based on reports from the World Health Organization (WHO), it was estimated that 1.8 million people died worldwide in 2005 from acute gastrointestinal illness (AGI) largely attributed to contaminated food and drinking water (Newell *et al.*, 2010).

A broad spectrum of foodborne pathogens exist, including bacteria, parasites, viruses, and prions, all of which have the potential to contaminate food and water supplies and subsequently, cause illness following ingestion (Tauxe, 2002). In regards to the developed world, although the clinical course tends to be generally milder, and the mortality rate is considerably lower, the magnitude of illness is significant and the resulting morbidity-associated burden and costs are remarkably high (Majowicz *et al.*, 2006). In Canada, it is estimated that approximately 4.0 million cases of domestically acquired foodborne illnesses occur each year and are attributed to at least 30 known pathogens (Thomas *et al.*, 2013). Based on the Hamilton study, AGI is estimated to attribute to an annual per capita cost of 115

dollars related to healthcare and lost productivity in Canada (Thomas *et al.*, 2008). If this estimate is applied to the entire Canadian population, the annual cost of AGI in Canada would be predicted at 3.7 billion dollars; however, due to potential confounding variables (i.e. geographical differences), additional studies are necessary to provide a more reliable estimate of the costs associated with the burden of AGI in Canada (Thomas *et al.*, 2006). Factors contributing to the occurrence of foodborne illness typically include: (1) large-scale production and widespread distribution of food, (2) globalization of food and food-related products, (3) increase in food consumption outside the home setting, and (4) genomic diversification of microbes and emergence of new pathogens (Nyachuba *et al.*, 2010).

As a direct result of the growing human population worldwide, industrial-scale and centralized production systems, including large-scale farming, food processing and distribution practises have drastically increased over the past several decades in order to meet the high demands for food (Nyachuba *et al.*, 2010). Although large-scale food processing has certain advantages such as increasing product value, and lengthening the shelf life of products, it is also suspected to contribute largely to food contamination and the spread of foodborne pathogens (Nyachuba *et al.*, 2010). Increases in contamination are likely attributable to the fact that food and/or food ingredients are typically handled by several individuals using various pieces of equipment throughout the course of processing. A characteristic example highlighting the severe consequences of a contaminated food processing environment is the 2009 *Salmonella* Typhimurium outbreak associated with a peanut butter manufacturing plant in the United States that resulted in 714 cases of illness and presumably 9 deaths (Maki, 2009). The contamination incident triggered a recall of the peanut butter, in addition to several other

manufacturers' products containing the affected peanut butter as an ingredient, such as cookies, crackers, and cereals among others, due to the possibility of contamination with *Salmonella* Typhimurium (Maki, 2009). Therefore, preventing contamination in the food processing environment is of critical importance for ensuring food safety.

Globalization of the food supply is another factor contributing to the burden of foodborne illnesses. According to the United States Department of Agriculture Economic Research Service (USDA-ERS), 15 percent of all food consumed in 2005 in the United States was imported, with a majority of those imports including fresh produce, tree nuts, fish and shellfish (Nyachuba *et al.*, 2010; Bravaro, 2012). Importation of food and food related products can increase the risk of foodborne illness as safety standards in some countries may be inferior; in addition, water contamination tends to be a significant public health issue in developing nations due to the use of untreated human waste or animal manure as fertilizer (Bravaro, 2012). Lifestyle, social, and economic changes, as well as increased travel and tourism have all contributed to an increase in the number of consumers eating outside of the home (Creel *et al.*, 2008). According to the Centers for Disease Control and Prevention (CDC), it is estimated that food consumed outside of the home contributes to about 50% of foodborne illnesses in the United States (Bravaro, 2012). Based on this data, the risk of foodborne illness associated with eating out is inherently greater than that of consuming food that is prepared in the home (Nyachuba *et al.*, 2010). Additionally, the demand for more ready-to-eat and refrigerated or frozen products has dramatically increased among consumers. Since processed foods require more stages of handling in comparison to unprocessed foods, the opportunity for contamination is significantly higher, and thus, the risk of acquiring a foodborne illness

following consumption is much greater (Nyachuba *et al.*, 2010). Although a large proportion of infections are contributed to eating outside of the home, numerous cases of foodborne illness have also been reported to occur within the home and most often arise due to poor time-temperature control, cross-contamination, and poor personal hygiene (Scott, 1996). Despite intensified efforts directed towards prevention and control, foodborne illness continues to be a significant public health threat.

Microbial evolution and diversification, in addition to the emergence of new pathogens, is well recognized as a major factor contributing to the overall burden of foodborne illness. A significant proportion of the genetic diversity reported among enteric bacteria has been obtained from DNA sequences acquired from distantly related organisms via horizontal gene transfer (Ochman *et al.*, 2000). These lateral transfers may result in enhanced virulence, and/or antimicrobial resistance, which may have serious implications for public health (Ochman *et al.*, 2000; Nyachuba *et al.*, 2010). To date, bacterial foodborne infectious agents are the most well characterized, investigated and monitored cases of intestinal disease (Newell *et al.*, 2010). Despite considerable efforts aimed at prevention and control, three major bacterial species (*Salmonella* spp., *Campylobacter* spp., and *Escherichia coli*) have persisted throughout the 1990's to present day (Newell *et al.*, 2010). Collectively, these bacterial pathogens constitute the greatest burden of foodborne illness for which etiology is known, and therefore have demanded the most research and surveillance attention from government agencies, in addition to awareness from the food industry (Newell *et al.*, 2010).

Although outbreaks only account for a minority of cases, however, these investigations contribute greatly to the understanding of the epidemiology of foodborne pathogens with respect to transmission routes, vehicles, and mechanisms of contamination (Rangel *et al.*, 2005). As well, it is the driving force that obliges regulatory and public health agencies and industry to evaluate prevention and control measures in order to prevent similar outbreaks from occurring in the future (Rangel *et al.*, 2005). Due to the severity of disease and high cost of illness, the Shiga-toxin producing *Escherichia coli* O157:H7 is considered to be one of the most significant foodborne pathogens from the public health standpoint, and therefore additional efforts to control this pathogen are warranted (Buchanan *et al.*, 1997; Law *et al.*, 2000; Lim *et al.*, 2010).

1.2 Epidemiology of *Escherichia coli* O157:H7 Infections in Humans

First described by Theodor Escherich in 1885, *Escherichia coli* (*E. coli*) is a catalase-positive, oxidase-negative, motile, Gram-negative, rod-shaped, facultative anaerobe belonging to the family *Enterobacteriaceae* (Buchanan *et al.*, 1997; Lim *et al.*, 2010). The bacterial species is a genetically heterogeneous group of typically non-pathogenic bacteria that naturally colonize the gastrointestinal tract (GIT) of both humans and animals as part of the normal microflora (Buchanan *et al.*, 1997; Lim *et al.*, 2010). However, some strains of *E. coli* have evolved into pathogenic bacterium through the acquisition of virulence factors present on plasmids, transposons, bacteriophages and/or pathogenicity islands (Lim *et al.*, 2010). Six classes of pathogenic *E. coli* are recognized based on their virulence factors and mechanisms by which they cause disease and include, enterohemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EaggEC), enteropathogenic *E. coli*

(EPEC), and diffusely adherent *E. coli* (DAEC) (Buchanan *et al.*, 1997). Characterized by the production of Shiga toxins, which are the leading cause of hemorrhagic colitis (HC) and hemolytic uremic syndrome (HUS), EHEC comprise a group of zoonotic pathogens frequently associated with diarrheal illnesses in humans (Lim *et al.*, 2010).

The most frequently isolated EHEC serotype from bloody stool samples is *E. coli* O157:H7, and is defined by the presence of the O antigen, which is a somatic antigen localized to the external part of the cell wall lipopolysaccharide of Gram-negative bacteria and the flagellar H antigen, which is found in the flagella of motile bacteria (Doyle *et al.*, 2006). Unlike the majority of *E. coli* species, *E. coli* O157:H7 is incapable of fermenting D-sorbitol and lacks β -glucuronidase (GUD) activity (Lim *et al.*, 2010). *E. coli* O157:H7 was first recognized as a human pathogen in 1982 following two outbreaks of bloody diarrhea associated with the consumption of undercooked hamburgers from a fast food chain in the United States (Buchanan *et al.*, 1997; Lim *et al.*, 2010). Since then, multitudes of food related outbreaks and sporadic cases have been reported, in which *E. coli* O157:H7 has been implicated as the main source of infection (Doyle *et al.*, 2006). *E. coli* O157:H7 infections occur worldwide and have been reported on every continent with the highest frequencies of disease transpiring in both Canada and the United States (Boerlin *et al.*, 1999). While *E. coli* O157:H7 is traditionally the most common serotype associated with bloody diarrhea and HUS, non-O157 serotypes producing Shiga toxins have also been reported to cause severe illness in humans (Bravaro, 2012).

The most commonly associated non-O157 serotypes with clinical manifestations of HUS are O111, O26, O103, and O145 (Brooks *et al.*, 2005; Bravaro, 2012). Infections associated with non-O157 serotypes are more common than infections with *E. coli* O157:H7 in Europe,

Australia, and Argentina (Brooks *et al.*, 2005; Bravaro, 2012). As a result of diagnostic limitations and inadequate surveillance, the burden of illness attributable to non-O157 serotypes is largely under recognized in both Canada and the United States (Brooks *et al.*, 2005; Bravaro, 2012). However, *E. coli* O157:H7 remains to be the most important cause of acute diarrhoeal illness, contributing to approximately 12,827 cases (per 100,000) of domestically acquired foodborne illness in Canada, circa 2006 (Thomas *et al.*, 2013).

1.2.1 Burden of Illness Attributed to *Escherichia coli* O157:H7 in Canada

According to the National Enteric Surveillance Program (NESP), 2,560 cases of laboratory confirmed *E. coli* O157:H7 were reported from 2008-2012 in Canada (Public Health Agency of Canada, 2012). Despite a decrease in the incidence rate from 1.98 cases per 100,000 in 2008 to 1.18 cases per 100,000 cases in 2010, the rate slightly increased to 1.39 in 2011 and 2012, which may be attributable to the occurrence of several multi-provincial outbreaks during this period of time in Canada (Public Health Agency of Canada, 2012).

Multi-jurisdictional (involving cases from multiple provinces and involving several federal agencies) *E. coli* O157:H7 outbreaks are investigated in a coordinated manner by both federal and provincial partners through an activated Outbreak Investigation Coordinating Committee (OICC) (Health Canada, 2011). Several high profile multi-jurisdictional *E. coli* O157:H7 outbreaks have occurred in Canada in recent years. For example, prompted by a multi-provincial outbreak of *E. coli* O157:H7 illness in 2011 (14 confirmed cases), the Canadian Food Inspection Agency (CFIA) issued a recall of raw shelled walnut products due to the suspected contamination with *E. coli* O157:H7 (The Canadian Press, 2011). Although not a large outbreak, this outbreak caused substantial morbidity, resulting in the hospitalization of 79

percent of cases. Of those hospitalized, 3 cases developed Hemolytic Uremic Syndrome (HUS), of which contributed to the death of one individual with underlying medical conditions (News Desk, 2011). This was the first time walnuts were associated with an outbreak of *E. coli* O157:H7 (there were no previously published reports), providing a clear demonstration of the versatility of this bacterial pathogen. Another multi-provincial and high profile outbreak occurring in 2012 that may have contributed to the increase in incidence was linked to the consumption of beef products from XL Foods Inc. A total of 18 laboratory confirmed cases were included in the outbreak, of which 6 of those cases required hospitalization (Lewis *et al.*, 2013).

E. coli O157:H7 outbreaks occur most typically as single strain events, and it can be possible (or even common) to see highly similar strain variants within a single outbreak, due to genetic changes that can occur during transmission throughout the outbreak. It is also possible for multiple strains to be involved in a single outbreak, which likely relates to the nature of the contamination event and the type of food product (Barrett *et al.*, 2006). For example, a multi-strain *E. coli* O157:H7 outbreak (multi-provincial) occurred in the late fall/early winter of 2011 through to early February of 2012 and was associated with the consumption of beef burger/steakette products, which resulted in 15 laboratory confirmed cases, 6 hospitalizations and 1 case of HUS (L. Tschetter, personal communication, June 10, 2014). Another multi-strain outbreak, which was also multi-provincial and epidemiologically linked to frozen beef burgers, extended over a 7 month period and resulted in 8 confirmed cases of illness, and 1 probable case. Of those reported to be ill, 2 hospitalizations and 1 death occurred among the probable case (L. Tschetter, personal communication, June 10, 2014).

1.2.2 Reservoirs of *Escherichia coli* O157:H7

Understanding the epidemiology of *E. coli* O157:H7 requires a widespread knowledge of where this microbe exists, how it flourishes in the environment, and all the ways in which humans can come into contact with it. *E. coli* O157:H7 naturally colonizes the gastrointestinal tract of ruminants, particularly those of the bovine species (i.e. cattle), of which have been considered to be the most important source of human infections (Doyle *et al.*, 2006; Lim *et al.*, 2010). STEC O157:H7 do not typically cause illness in the host reservoir, with the exception of a few rare cases of diarrhea reported in calves (Doyle *et al.*, 2006).

Examination of both naturally and experimentally infected cattle and calves indicated that *E. coli* O157:H7 adhere to the lymphoid dense mucosa located in the terminal rectum near the recto-anal junction (RAJ) (Naylor *et al.*, 2003). Studies conducted by Fitzhenry and colleagues have shown that this selectivity in colonization within the bovine host may be attributed to an increased affinity of intimin- γ binding to the follicle associated epithelium (FAE) of ileal Peyer's patches, which are known to be abundant in the RAJ (Fitzhenry *et al.*, 2002). Experimental evidence suggests that the prevalence of *E. coli* O157:H7 carriage in cattle and their excreta not only varies on a seasonal basis, with higher fecal shedding occurring in the warmer months, but also with age, diet, and immunity of the animal host (Fitzhenry *et al.*, 2002). Other ruminants such as sheep, goats, and deer have also been found to actively shed *E. coli* O157:H7 in their feces, however, these animals are considered to be secondary reservoirs and are not a principal source of infection (Buchanan *et al.*, 1997; Lim *et al.*, 2010).

In rare instances, STEC O157:H7 has been isolated from other animal sources including, swine, poultry, and shellfish, but it is speculated the pathogenic bacteria was transient in nature

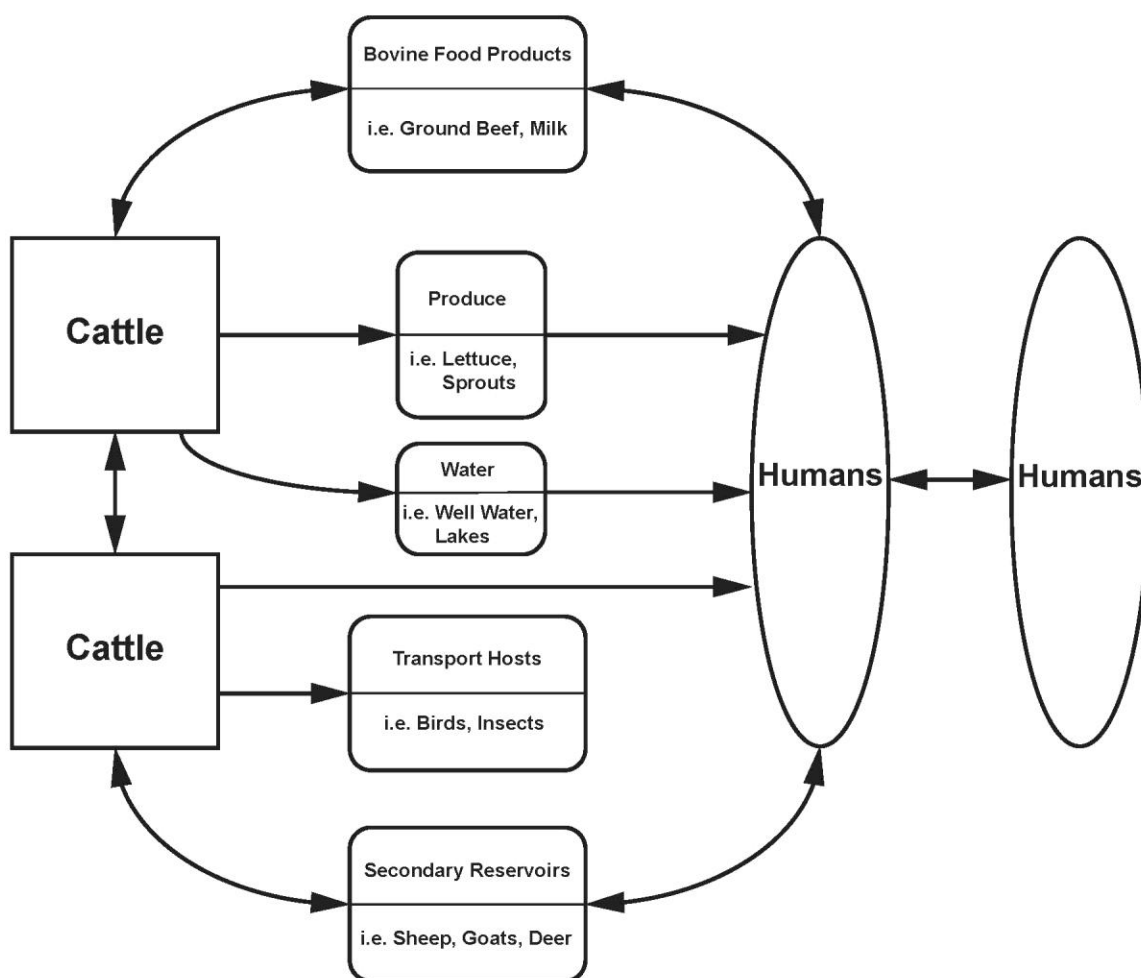
and was most likely acquired through contact with meat, vegetation or water contaminated with fecal material from ruminants (Buchanan *et al.*, 1997). In addition, birds, insects (i.e. flies and beetles), and slugs have been described as transport hosts and have been hypothesized to contribute to the dissemination of the bacterial pathogen on farms (Cizak *et al.*, 2000; Sproston *et al.*, 2006; Talley *et al.*, 2009).

1.2.3 Routes of Transmission for *Escherichia coli* O157:H7

There are various routes for human infection with *E. coli* O157:H7, of which are all variations of fecal-oral transmission (Figure 1). Studies in both Canada and France have demonstrated a correlation between the incidence of STEC infections in humans with indicators of cattle density (Haus-Cheymol *et al.*, 2006; Valcour *et al.* 2002). *E. coli* O157:H7 shed by colonised ruminants may be spread amongst various other animals and surfaces present in the surrounding area, either through direct contact or via contaminated water troughs, feed, pastures and/or other environmental sources (Islam *et al.*, 2004; Solomon *et al.*, 2002). Based on a survey of 29 feedlots conducted in the Midwestern United States, 38 percent of cattle were found to have positive hide samples for *E. coli* O157:H7 (Elder *et al.*, 2000).

Under favourable environmental conditions (cooler temperatures and humid conditions), pathogenic *E. coli* O157:H7 has also been found to persist for considerable amounts of time on a variety of farmyard surfaces such as, gates and stiles; therefore, these surfaces may serve as a potential source in the infection pathway, particularly when the risk of direct human contact is high (i.e. petting zoos) (Williams *et al.*, 2005). Although humans have not been reported as a maintenance host for STEC O157:H7, person-to-person transmission has

Figure 1. Model depicting possible routes for transmission of *Escherichia coli* O157:H7 from the bovine host to humans. The figure is constructed based on data compiled from several studies and illustrates examples of the major sources of *Escherichia coli* O157:H7 infections that have been reported in the scientific literature. Arrows in the figure indicate direct transmission (based on data presented by Gansheroff and O'Brien, 2000).



been considered the primary mode of infection in many outbreaks involving daycares, schools, hospitals and other institutions, where lapses in hygiene have occurred (Belongia *et al.*, 1993). Contaminated food and food-related products continue to be the principal vehicle for transmission of *E. coli* O157:H7 to humans, accounting for approximately 52 percent of STEC outbreaks that occurred in the United States from 1982 to 2002 (Erickson *et al.*, 2007).

Since the ultimate reservoir of *E. coli* O157:H7 is the bovine host, there is ample opportunity for contamination to ensue during food-processing (Price *et al.*, 2004). In a study conducted in the Midwestern United States, 158 out of 341 (46.3%) cattle carcasses tested positive at one or more sites for *E. coli* O157:H7 (Elder *et al.*, 2000). Based on these findings, contamination of beef and other meat-related products is posited to occur most frequently during the slaughter and grinding process as it often results in internalization of the pathogen within the meat source (Lim *et al.*, 2010). Interestingly, *E. coli* O157:H7 has been documented to persist for extensive periods of time on both stainless steel and plastic; therefore, these surfaces could potentially serve as intermediate sources of contamination during food processing operations, greatly enhancing the possibility for cross-contamination (Erickson *et al.*, 2007).

Cross-contamination is one of the most common avenues of contamination and is defined as the transfer of microorganisms from one food or surface to another (Nyachuba *et al.*, 2010). Although a large proportion of cross-contamination occurs within the food-processing environment, these events are not limited to one particular setting and have been documented to occur in both food service operations and the home (Erickson *et al.*, 2007). Hence, to reduce the risk of acquiring an *E. coli* O157:H7 infection via consumption of

contaminated beef and/or meat products, it is crucial to prepare these products to the appropriate internal cooking temperature (71°C for ground beef). Throughout the years, there has been more of a shift from meat related outbreaks to produce related sources of infection, which has been increasingly recognized as a major contributor to foodborne outbreaks associated with *E. coli* O157:H7 in developed nations (Lynch *et al.*, 2009). This shift may be attributed to society's desire for a healthier diet that includes more fruits and vegetables (Bravaro, 2012).

Globalization of the food supply, in addition to large-scale production and distribution are other major factors contributing to the ongoing issue associated with foodborne diseases linked to produce (Bravaro, 2012). One of the largest outbreaks was linked to the consumption of raw produce (radish sprouts) in July 1996, in Sakai City, Japan and resulted in 12,680 infections and 12 deaths among affected individuals (Michino *et al.*, 1999). Several large-scale outbreaks have occurred since then, including the 2006 multinational (United States and Canada) bagged spinach outbreak, where 204 individuals were affected and 3 deaths occurred, thus highlighting the potential for widespread outbreaks of human infection caused by contaminated produce (Wendel *et al.*, 2009). The desire for fresh fruits and vegetables year round, especially during the winter season ultimately results in a higher demand for globally sourced produce (Lynch *et al.*, 2009). Although contamination of produce can occur at several steps during processing, the risk is highest in the field, during initial processing, and during preparation (Bravaro, 2012). As agriculture becomes more intensive, produce fields tend to become situated in closer proximity to animal production zones, minimizing the ecological connection between wild animals, farm animals and fresh produce (Lynch *et al.*, 2009).

The main sources implicated in contaminating fresh produce in the field include, improperly treated manure applied as fertilizer, contaminated surface run-off from cattle production operations, and/or contaminated irrigation water (Lim *et al.*, 2010). An environmental investigation following a large-scale spinach outbreak associated with *E. coli* O157:H7 contamination suggested feral swine may have played a role in contaminating the field where the produce was grown (Lynch *et al.*, 2009). Since *E. coli* O157:H7 has the ability to survive for extended periods of time (>7 months) in soil, especially in the presence of manure, edible portions of plants may become contaminated without direct exposure to the pathogen (Solomon *et al.*, 2002). Water contaminated with *E. coli* O157:H7 may be actively taken up by the root system and internalized within the leaf structures 20 to 100 µm below the surface, thereby preventing exposure of the pathogen to traditional external sanitizing treatments (Islam *et al.*, 2004; Solomon *et al.*, 2002).

During processing, contaminated water may be used for washing or in ice used to preserve produce for shipping (Bravaro, 2012). Contamination can also transpire during food preparation and typically occurs through improper cleaning of produce or cross-contamination with utensils or boards that had been previously used in the preparation of meat (Bravaro, 2012). It is important to note that vehicles for transmission of *E. coli* O157:H7 to humans are not solely limited to animal products or fresh produce, and can include a wide variety of consumable products. For example, the consumption of ready-to-bake prepackaged cookie dough was identified as the source of infection during a multistate outbreak occurring in 2009 in the United States (Neil *et al.*, 2011). Drinking and recreational water contaminated with fecal matter from infected animals and/or humans has also been recognized as a source of human

infection, and has been linked to several outbreaks associated with *E. coli* O157:H7. In 2000, an estimated 2,300 illnesses and 7 deaths occurred from exposure to contaminated drinking water in the town of Walkerton, Ontario, Canada (Hrudey *et al.*, 2003). The primary pathogens identified as being responsible for the outbreak were *E. coli* O157:H7 and *Campylobacter jejuni*, although other pathogens were likely to have been present (Hrudey *et al.*, 2003). Hence, the ability of *E. coli* O157:H7 to survive and flourish in the environment is crucial for the persistence and dissemination of the pathogen (Islam *et al.*, 2004).

1.3 Features of *Escherichia coli* O157:H7 - Biology and Disease

1.3.1 Clinical Manifestations

The incubation period for *E. coli* O157:H7 ranges from 1 to 16 days, with most signs of an infection becoming apparent after 3 to 4 days (Su *et al.*, 1995; Buchanan *et al.*, 1997). During an *E. coli* O157:H7 infection, a spectrum of clinical manifestations can arise, ranging from asymptomatically infected individuals, to those experiencing mild diarrhea, severe bloody diarrhea, hemorrhagic colitis (HC) and in the most severe cases, hemolytic uremic syndrome (HUS) (Su *et al.*, 1995). Most symptomatic cases begin with the onset of watery diarrhea and may resolve within a weeks' time independent of treatment (Su *et al.*, 1995). However, most infections often progress to HC, and are characterized by inflammation of the large bowel, severe bloody diarrhea, abdominal tenderness and most often, severe abdominal cramps (Su *et al.*, 1995). As well, some patients may experience a mild fever, in addition to nausea, vomiting and dehydration (Su *et al.*, 1995).

Manifestations of HC are self-limiting and tend to resolve in approximately one week (Su *et al.*, 1995). In the weeks following the onset of diarrhea, HUS develops in approximately 5

percent to 10 percent of patients and is most common in children, the elderly and immunocompromised individuals (Su *et al.*, 1995; Buchanan *et al.*, 1997). The disease is characterized by acute renal failure, hemolytic anemia and thrombocytopenia; however, the relative importance of these signs during disease progression varies (Su *et al.*, 1995; Buchanan *et al.*, 1997). Although not part of the classic triad of conditions associated with HUS, extra-renal disease involving central nervous system (CNS) complications such as lethargy, irritability and seizures may develop in 30 percent to 50 percent of patients (Su *et al.*, 1995; Buchanan *et al.*, 1997). *E. coli* O157:H7 accounts for 85 to 95 percent of cases of HUS in North America (Su *et al.*, 1995).

In extremely rare instances, HUS may progress to Thrombotic Thrombocytopenic Purpura (TTP), which is defined by a pentad of conditions including thrombocytopenia, microangiopathic hemolytic anemia, fever, renal failure and neurological symptoms (Su *et al.*, 1995; Buchanan *et al.*, 1997). All documented cases of TTP associated with *E. coli* O157:H7 have occurred exclusively in adults and the elderly, and progression from HUS to TTP in children has largely not been reported (Buchanan *et al.*, 1997). Approximately 65 percent to 85 percent of children diagnosed with HUS recover after supportive treatment without exhibiting permanent damage; however, long term renal complications may occur (Su *et al.*, 1995). The case-fatality rate of children with HUS is 3 percent to 10 percent, and is significantly lower than the case-fatality rate observed in elderly patients with TTP, which can be up to 50 percent (Su *et al.*, 1995; Buchanan *et al.*, 1997).

1.3.2 Pathogenesis

Initial attachment of the bacterium to the gastrointestinal epithelium occurs via flagella, fimbriae, and other surface adhesion proteins (Tree *et al.*, 2009). The formation of A/E lesions, which are characterized by intimate bacterial adherence, destruction of microvilli and pedestal formation (Law, 2000; Garmendia *et al.*, 2005), is a major mechanism employed by *E. coli* O157:H7 and other EHEC in the establishment of infections (Garmendia *et al.*, 2005). While remaining extracellular, the pathogen triggers A/E lesion formation through architectural alterations in the cytoskeletal microtubule, intermediate filament (IF), and actin microfilaments of infected host cells via translocation and integration of specific effector proteins (Garmendia *et al.*, 2005). The effector protein Tir is essential for A/E lesion formation and is translocated into the host epithelial cell via TTSS where it localizes to the plasma membrane (Garmendia *et al.*, 2005). Forming a hair-pin like structure within the membrane, Tir interacts with the adhesion factor intimin displayed on the outer portion of the bacterial membrane, resulting in intimate attachment of the bacterium (Garmendia *et al.*, 2005; Lim *et al.*, 2010). The C-terminus of the Tir protein recruits several focal adhesion proteins to the site of bacterial attachment in order to stimulate actin polymerization and pedestal formation (Garmendia *et al.*, 2005; Lim *et al.*, 2010). Pedestals form directly beneath adherent bacterial cells as a result of actin accumulation and may extend up to 10 micrometers outwards from the host cell (forms a pseudopod-like structure) (Garmendia *et al.*, 2005; Lim *et al.*, 2010).

Additional alterations to the host cell epithelium resulting from the integration of effector proteins in *E. coli* O157:H7 infections include; (i) microvilli brush border deficiency, (ii) disruption of the intestinal barrier function through loss of trans-epithelial resistance (TER) and

increased tight junction permeability, (iii) reduction in resting membrane potential leading to accumulation of sodium (Na^{+1}) and water in the lumen, (iv) loss of mitochondrial membrane potential, (v) inhibition of the G₂/M cell cycle phase transition, and (vi) cellular apoptosis (Garmendia *et al.*, 2005; Tree *et al.*, 2009). The binding of Shiga-toxins to Gb₃ induces inhibition of protein synthesis, cell apoptosis and the destruction of red blood cells, which leads to further destruction of intestinal epithelium (Caprioli *et al.*, 2005). These modifications not only elucidate to persistent bloody diarrhea, but also to severe inflammation of the large bowel as a result of the production of interleukin (IL)-8 and successive transmigration of acute inflammatory cells to the site of infection (Garmendia *et al.*, 2005; Lim *et al.*, 2010).

1.3.3 Treatment and Prevention

Currently, no specific regimen exists for infections with *E. coli* O157:H7 and treatment is largely limited to supportive care, including electrolyte fluid therapy and management of complications such as anemia and renal failure (Su *et al.*, 1995; Wong *et al.*, 2000). The use of antibiotics is controversial and often not recommended as they may worsen the clinical course of the disease through elimination of competing bowel flora (leading to overgrowth of *E. coli* O157:H7) and liberation of Shiga-toxins (induced by lysis or sublethal damage to the pathogen) (Goldwater *et al.*, 2012). In addition, treatment with antibiotics does not improve *E. coli* O157:H7 infections as they have demonstrated no substantial effect on reduction of symptoms, prevention of complications or decrease in bacterial shedding (Su *et al.*, 1995).

Patients with severe complications often require intensive care, dialysis, transfusion and/or platelet infusion, and in most extreme cases, kidney transplantation may be necessary (Su *et al.*, 1995; Wong *et al.*, 2000). Since the treatment of *E. coli* O157:H7 within the human

host remains problematic, many researchers are focusing on prevention/intervention strategies aimed at decreasing/eliminating bacterial carriage in ruminants (Gu *et al.*, 2009; Potter *et al.*, 2004; Yetka *et al.* 2011). Therefore, at the present time, the main strategies to reduce the risk of *E. coli* O157:H7 infections include; (i) frequent hand washing, especially before food preparation and consumption (Buchanan *et al.*, 1997), (ii) screening and control programs to reduce microbial contamination during food-processing events (Buchanan *et al.*, 1997), (iii) thorough washing of vegetables and preparation of meat products to a minimum internal temperature of 71°C (Buchanan *et al.*, 1997), (iv) standard water treatment procedures for public water supplies and irrigation systems (Buchanan *et al.*, 1997), (v) adequate treatment of manure applied as fertilizer to eliminate pathogenic microbes (Islam *et al.*, 2004; Solomon *et al.*, 2002), and (vi) implementation of buffer zones positioned between animal grazing areas and vegetable production fields (Islam *et al.*, 2004).

1.3.4 Factors Contributing to Virulence

1.3.4.1 The Locus of Enterocyte Effacement

Colonization of the intestinal mucosa by *E. coli* O157:H7 induces the formation of histopathological lesions, otherwise referred to as attaching and effacing (A/E) lesions, on the follicle associated epithelium (FAE) of ileal Peyer's patches, of which is considered a major mechanism of infection (Garmendia *et al.*, 2005; Law, 2000). The genes necessary for conferring the A/E phenotype are located on locus of enterocyte effacement (LEE) pathogenicity island (PAI) and include: (i) a type three secretion system (TTSS) required for the translocation of several effector molecules into the host cell to induce cytoskeletal alterations in epithelial cells that result in A/E lesions, (ii) genes encoding secreted proteins (Esp) of which

modify host cell signal transduction pathways, (iii) an *eae* gene encoding the outer membrane protein intimin that becomes inserted into the outer membrane of the bacterial cell to promote adhesion and (iv) a *Tir* gene encoding for the protein Tir (translocated intimin receptor), which is translocated into the host cell via TTSS (Garmendia *et al.*, 2004; Garmendia *et al.*, 2005; Law, 2000). Tir inserts into the epithelial cell membrane where it serves as both a receptor for intimin and as a nucleator of proteins involved in actin assembly (Garmendia *et al.*, 2005; Law, 2000).

1.3.4.2 pO157 plasmid

E. coli O157:H7 clinical isolates possess a conserved, 92 kilobase putative virulence plasmid referred to as pO157 (Law, 2000; Lim *et al.*, 2010). Due to its heterogeneous composition, experimental evidence suggests that the pO157 plasmid is formed by the integration of several mobile genetic elements such as transposons, prophages, and insertion sequences from evolutionally diverse species (Law, 2000). According to several studies, the highly conserved nature of the plasmid amongst clinical isolates strongly suggests that pO157 encodes essential virulence factors and/or elements necessary for efficient colonization and/or survival in the host (Lim *et al.*, 2010). Additionally, the presence of pO157 was found to increase the level of *E. coli* O157:H7 adherence to the RAJ mucosa in the bovine GIT and therefore, was vital for long term colonization (Sheng *et al.*, 2006; Lim *et al.*, 2010).

In contrast, strains lacking the pO157 plasmid were recovered in significantly lower numbers and were cleared faster than the wild type (Sheng *et al.*, 2006). These findings indicate pO157 may be necessary for optimal survival and persistence of *E. coli* O157:H7 in the bovine reservoir both directly and through regulation of chromosomal genes (Sheng *et al.*,

2006). Among the 100 open reading frames (ORFs) identified, only 19 genes have been characterized and therefore, the biological significance of pO157 in *E. coli* O157:H7 pathogenesis is not yet fully understood (Lim *et al.*, 2010; Law, 2000).

1.3.4.3 Shiga Toxins

The production of Shiga toxin (Stx) or Verocytotoxin is another defining characteristic of *E. coli* O157:H7 and is a contributing factor to the manifestations of hemorrhagic colitis and hemolytic uremic syndrome (Lim *et al.*, 2010). Two antigenically distinct forms of Stx may be produced by *E. coli* O157:H7 including Stx1, which demonstrates significant sequence homology to the toxin of *Shigella dysenteriae* serotype 1, and/or Stx2, which shares less than 60% amino acid sequence with Stx1 (Fraser *et al.*, 2004). While Stx1 is relatively homogeneous, several variants of Stx2 have been described (Caprioli *et al.*, 2005). The Stx2c variant of Stx has been found to most commonly associate with *E. coli* O157:H7 (Fraser *et al.*, 2004).

Belonging to the AB₅ toxin family, these compound toxins are composed of a single A (active) subunit noncovalently associated with a pentameric ring of B (binding) subunits (Law, 2000). The B-pentamer contains the binding sites for glycolipid receptors located on the membrane surface of target eukaryotic cells and therefore mediates the entry of the catalytically active A subunit into these cells (Fraser *et al.*, 2004; Law, 2000). These receptors have been identified as globotriaosylceramide (Gb₃), which preferentially bind Stx1, Stx2, and Stx2 variants with the exception of Stx2e, which typically binds to globotetraosylceramide (Gb₄) (Fraser *et al.*, 2004; Law, 2000). Once bound, the A subunit is internalized by clathrin-dependent endocytosis and processed through the Golgi apparatus and endoplasmic reticulum (ER) where subsequent cleaving of the A subunit by a host protease generates a catalytically

active A1 fragment and an A2 fragment (Fraser *et al.*, 2004). The A1 subunit is translocated to the cytosol where it utilizes N-glycosidase activity to remove a single adenosine residue from the 28S ribosomal RNA (rRNA) of the 60S ribosomal subunit (Fraser *et al.*, 2004; Law, 2000). Alteration of the 28S rRNA prevents binding of elongation factor, which leads to inhibition of protein synthesis and ultimately death of the intoxicated cell by apoptosis (Law, 2000).

Despite the similar structures and modes of action by both Stx1 and Stx2, the toxicities are distinct with the latter exhibiting 1000 fold greater cytotoxicity than the former (Law, 2000). Epidemiological evidence indicates that *E. coli* O157:H7 isolates producing Stx2 only are more commonly associated with severe disease than isolates producing Stx1 or Stx1 and Stx2 (Manning *et al.*, 2008). Due to their extreme toxicity, Shiga toxins have been classified as category B bioterrorism agents by the Center for Disease Control and Prevention (CDC) (Parma *et al.*, 2011).

1.4 Isolation and Identification of *Escherichia coli* O157:H7

Accurate and definitive identification of *E. coli* O157:H7 is essential for both correct disease diagnosis and treatment of the infection and fundamentally relies on subtyping techniques that generate strain/clone-specific fingerprints below the species/subspecies level (van Belkum *et al.*, 2007). The data generated from subtyping is also useful for detecting or ruling out cross-infections, elucidating transmission pathways and most importantly, identifying reservoirs or sources of infection in humans in order to prevent further illness (van Belkum *et al.*, 2007; Shelton *et al.*, 2008). There are several conventional laboratory techniques for phenotypically analyzing *E. coli* O157:H7, which results in the grouping of these microorganisms based on the expression of similar characteristics (Park *et al.*, 1999; van Belkum *et al.*, 2007).

Due to the low infectious dose (approximately 10 organisms by ingestion) and relatively low prevalence of the pathogenic microbe in food when compared to other competing microflora, screening and identification methods must be both sensitive and specific for *E. coli* O157:H7 strains (Vernozy-Rozand, 1997). Identification of *E. coli* O157:H7 is largely facilitated by classical microbiological procedures that target characteristic phenotypic features of this bacterial pathogen including: the inability to ferment sorbitol, lack of Beta (β)-glucuronidase activity, toxin production, phage-type susceptibility, and the presence of the O157 (somatic) and H7 (flagellar) antigens (Sanchez *et al.*, 2010).

1.4.1 Selective Media used for the Isolation of *Escherichia coli* O157:H7

Introduced as a valuable tool in 1986 by March and Ratnam, sorbitol-MacConkey agar (SMAC) was the first selective medium to be routinely implemented in the screening and isolation process of *E. coli* O157:H7 (March and Ratnam, 1986). This selective media has been proven to be useful in discriminating between O157 and non-O157 *E. coli* isolates based on the phenotypic characteristic of sorbitol fermentation (March and Ratnam, 1986). Unlike most *E. coli* strains, *E. coli* O157:H7 does not typically ferment sorbitol and therefore, will form distinctive white or colorless colonies on the surface of the agar following incubation (Karch *et al.*, 1999). These isolates are distinguished from sorbitol fermenting (SF) intestinal microorganisms present in stool samples, such as other strains of *E. coli*, as these microorganisms will form pink or red colonies (March and Ratnam, 1986; Karch *et al.*, 1999). However, other non-sorbitol fermenting (NSF) intestinal *Enterobacteriaceae* that may be present in human stool, including *Proteus* spp., *Providencia* spp., *Hafnia* spp., *Enterobacter* spp., *Morganella* spp., and *Escherichia hermannii* (*E. hermannii*), can also form colorless colonies on

SMAC agar, potentially leading to false positive results (Feng, 1995; Karch *et al.*, 1999). Due to the prevalence of phenotypic variations within the serogroup, the exclusive use of sorbitol medium is inadequate in detecting *E. coli* O157:H7 and all NSF isolates present on the medium must be considered presumptive until serologically confirmed (Feng, 1995; Jordan and Maher, 2006).

Another distinctive phenotypic feature of *E. coli* O157:H7 that differentiates this pathogenic microorganism from most other strains of *E. coli* is the lack of Beta (β)-glucuronidase (GUD) activity (Chapman, 2000). The fluorogenic substrate 4-methylumbelliferyl β -D-glucuronide (MUG) may be incorporated into SMAC to provide increased discriminatory power for *E. coli* O157:H7 identification (Chapman, 2000). When MUG is cleaved by β -D-glucuronidase, a fluorescent product, 4-methylumbelliferone, is produced and is detectable under long wave ultraviolet (UV) light (Park *et al.*, 1999). Since *E. coli* O157:H7 lacks GUD activity in addition to SF, these colorless colonies will not emit fluorescence and therefore can be enumerated and further characterized (Chapman, 2000; Park *et al.*, 1999). A chromogenic substrate 5-bromo-4-chloro-3-indoxyl- β -D-glucuronide (BCIG) may also be added to SMAC as a substrate for detecting β -glucuronidase activity (Chapman, 2000; Park *et al.*, 1999). Microorganisms possessing β -glucuronidase activity will cleave the substrate BCIG and release indoxyl, which is then rapidly oxidized to the insoluble compound indigo, leading to the formation of distinct blue to purple colored colonies (Chapman, 2000; Park *et al.*, 1999). In contrast, the non-sorbitol-fermenting and β -glucuronidase-negative *E. coli* O157:H7 will appear as colorless colonies (Chapman, 2000; Park *et al.*, 1999). Despite the selectivity and efficiency

of fluorogenic and chromogenic media, the high cost of these substrates would most likely prevent routine use in a majority of laboratories (Chapman, 2000).

1.4.2 Biochemical Tests

Colonies suspected to be *E. coli* O157:H7 that have been isolated from selective plating methods can be further characterized using biochemical tests. The IMViC (Indole, Methyl Red, Voges-Proskauer and Citrate) test is usually employed as a primary biochemical test in the initial screening of *E. coli* O157:H7 (Deisingh and Thompson, 2004). The use of well-established commercial identification systems such as, the API 20 E and ID 32E systems to detect biochemical features of presumptive *E. coli* O157:H7 strains, offer the advantage of confirming isolates at the genus and/or species level in addition to possibly identifying profiles that are unique to those strains (Leclercq *et al.*, 2001). Despite the relatively high level of reproducibility for commercial biochemical systems, phenotypic variants of *E. coli* O157:H7 may cause identification problems in a number of these systems (Leclercq *et al.*, 2001). Therefore, further phenotypic characterization is required to confirm the presence of *E. coli* O157:H7.

1.4.3 Serotyping

Non-sorbitol fermenting colonies confirmed as *E. coli* from biochemical tests must be serologically identified as O157:H7 (Karch *et al.*, 1999). Since over 700 serotypes of *E. coli* are known to exist, consisting of both pathogenic and commensal bacteria, accurate determination of the serotype is crucial for providing the foundation of pathogenic *E. coli* diagnostics (Karch *et al.*, 1999). Briefly, serological identification is conducted via determination of the somatic (O) antigen, which is located in the external portion of the cell wall lipopolysaccharide, and the

flagellar (H) antigen, which is contained within the flagella of motile bacteria (Karch *et al.*, 1999; Sowers *et al.*, 1996). Serotyping is typically performed using tube (or slide) agglutination with specific antiserum or commercially available anti-O157 latex reagents and standard biochemical tests (Karch *et al.*, 1999; Sowers *et al.*, 1996). A positive test is indicated by clumping of the culture with specific O157 antiserum (Sowers *et al.*, 1996; Yang *et al.*, 2007). However, the O157 antigen has been found to be present in other non-H7 *E. coli* serotypes of which are typically non-pathogenic, and therefore, high rates of false positive reactions may occur due to antigenic mimicry (Feng, 1995; Rozand and Feng, 2009). Only those isolates producing a positive agglutination reaction with anti-O157 serum are further serotyped for the H7 antigen (Feng *et al.*, 1998).

Identification of 'H' antigens from *E. coli* O157:H7 cultures are conducted using broth cultures of actively motile bacteria (Sowers *et al.*, 1996; Yang *et al.*, 2007). A well isolated colony suspected to be *E. coli* O157:H7 is transferred into the inner portion of a Craigie tube containing semisolid medium and incubated until complete migration (from inner tube to the outer area) occurs, typically within 24 hours (Sowers *et al.*, 1996; Yang *et al.*, 2007). At least three subcultures through the semisolid medium are necessary to achieve optimal motile ability of the flagellar antigen (Sowers *et al.*, 1996; Yang *et al.*, 2007). The motile culture is inoculated into Ewing's broth and screened for standardized antigens in pooled 'H' antisera using the agglutination technique (Sowers *et al.*, 1996; Yang *et al.*, 2007). Clumping of the culture with specific H7 monoclonal antiserum is indicative of a positive test for *E. coli* O157:H7 (Sowers *et al.*, 1996; Yang *et al.*, 2007).

1.4.4 Phagetyping

Phagetyping is a classic phenotypic method utilized in the screening and identification process of *E. coli* O157:H7 (Ahmed *et al.*, 1987). The phage is a virus that is capable of infecting bacteria (bacteriophage) and ultimately results in cell death (lysis); however, different strains of bacteria demonstrate variation in their susceptibility to certain phage types (Ahmed *et al.*, 1987). The phagetyping scheme for *E. coli* O157:H7 employs a set of 19 standard phages provided by the Enteric Disease Program at the National Microbiology Laboratory and allows for the differentiation of approximately 90 phage types (Ahmed *et al.*, 1987; Khakhria *et al.*, 1990). The standard set of 19 phages is applied at a routine test dilution (RTD) to a Difco phage agar plate containing presumptive *E. coli* O157:H7 culture using a multiple inoculum syringe and is examined for lytic reactions following incubation (Ahmed *et al.*, 1987; Khakhria *et al.*, 1990). Phage types most frequently associated with outbreaks of *E. coli* O157:H7 in Canadian provinces are 1, 4, 8, 14, and 31, with types 8 and 14a being the most prominent (Ahmed *et al.*, 1987; Khakhria *et al.*, 1990). Although phagetyping is considered to be a rapid and reliable high throughput method, phage types have the potential to change during an outbreak (or passage through a host), resulting in mixed or atypical phage types (Khakhria *et al.*, 1990).

1.5 Characterization and Subtyping of *Escherichia coli* O157:H7 beyond the Phenotypic Level

Phenotypic methods generally have low discriminatory power and are primarily implemented as first level subtyping methods (Ahmed *et al.*, 1987). These techniques are not only prone to false-positive results but also are labour-intensive (require lengthy enrichment processes), time consuming, involve substantial subjective interpretation, and require strict standardisation of experimental conditions (Brunt *et al.*, 2010). Additionally, the emergence of

phenotypic variants, extremely low infectious dose, and presence of characteristically similar competitor microorganisms make detection of *E. coli* O157:H7 extremely difficult based on phenotype alone (Deisingh and Thompson, 2004). More importantly, laboratory data obtained from phenotypic methods does not always accurately reflect the genotype of a microorganism, and thus, may not provide a reliable marker for epidemiological studies (van Belkum *et al.*, 2007). For example, two isolates demonstrating indistinguishable phenotypes may be completely unrelated; while conversely, two isolates with vastly different phenotypes may actually be related. The ability to react in a timely manner to potential foodborne outbreaks of *E. coli* O157:H7 and to trace the contaminated sources contributing to infection are of paramount importance and depend largely on the availability of highly discriminatory and reliable subtyping methods (Lindstedt *et al.*, 2003). Since epidemiological relationships cannot be established based on phenotypic testing alone, more specific methods that assess the genetic variation must be used in conjunction with current phenotypic techniques to effectively analyze *E. coli* O157:H7. As sequences of the *E. coli* O157: H7 genome have become increasingly available, a wide variety of genotypic methods targeting different aspects of the bacterial DNA, have been developed (Table 1) (van Belkum *et al.*, 2007).

1.5.1 Detection of Shiga-toxins

Several methods are available to laboratories for toxin detection including cell culture techniques, selective media plating methods, commercial enzyme immunoassays, and PCR; however, the most common techniques implemented in routine laboratory screening of *E. coli* O157:H7 toxins are Paton multiplex PCR and cell culture (Rahn *et al.*, 1996; Paton and Paton, 1998). To determine the presence of the *stx1* and *stx2* genes, Paton multiplex PCR is conducted

using primer sequences specific to each toxin gene, in addition to primer pairs specific for putative virulence genes *eaeA* and EHEC-*hlyA* (Paton and Paton, 1998). Since STEC is comprised of a large group of microorganisms, putative virulence genes are included in the protocol to distinguish pathogenic *E. coli* O157 from less virulent serotypes, as these strains do not typically produce accessory virulence factors such as intimin (encoded by *eaeA*) and the plasmid-encoded enterohemolysin (encoded by EHEC-*hlyA*), and therefore, can be easily recognized during analysis (Paton and Paton, 1998). At the National Microbiology Laboratory (NML) in Winnipeg, Manitoba, Canada, amplification products are analyzed by submarine gel electrophoresis following Paton multiplex PCR, and stained with GelRed according to SOP # E-TT-PR-012 (Public Health Agency of Canada, 2013). If *stx1*, *stx2*, *eaeA*, and EHEC-*hlyA* are present, amplification products of 180, 255, 384, and 534 base pairs, respectively, will be detected (Paton and Paton, 1998). Further subtyping at the NML is performed if the sample is positive for the *stx2* gene to determine the *stx2* subtype and is carried out in accordance with SOP # E-TT-PR-011 (Public Health Agency of Canada, 2014).

Although Paton PCR is both a rapid and sensitive method for genotyping of pure cultures obtained from human samples, the assay offers limited information when applied directly to food/environmental samples due to the widespread presence of these genes in other strains of *E. coli* that may not be pathogenic to humans (Paton and Paton, 1998). Therefore, delays in human illness investigations often occur due to the lengthy enrichment procedures required to obtain a pure culture (Paton and Paton, 1998). Since genotype is not always reflective of phenotype, cell culture is often utilized to confirm the results obtained by Paton PCR.

Table 1. Molecular Based Laboratory Techniques used to Genotypically Characterize *Escherichia coli* O157:H7 (with the exception of Vero Cell Assay). Information included in the table is referenced in the text.

Subtyping Method	Brief Description	Advantages	Disadvantages
Toxin Testing	• Multiplex PCR assay that utilizes four PCR primer pairs and detects the presence of <i>stx1</i>, <i>stx2</i> (including variants of <i>stx2</i>), <i>eaeA</i>, and <i>EHEC-hlyA</i> • Vero cell assay for toxin detection, used to confirm the results obtained from PCR, and to determine whether toxin is being expressed	• PCR is both a rapid and sensitive method • Genotypic characterization of the toxin • Confirmation of the results by two methods	• Time-consuming • Labour-intensive (Vero cell assay) • Results do not always agree (prone to subjectivity)
Pulsed-field Gel Electrophoresis (PFGE)	• Based upon the variable size-dependent migration of large DNA restriction fragments in a periodically pulsing electric field of alternating polarities • Gold Standard among molecular typing methods	• Good discriminatory power (new or rare patterns) • High epidemiological concordance • Relatively inexpensive • Universal typeability • Reproducible • Large historical database	• Technically demanding • Labour-intensive • Time-consuming • Cannot handle large sample sets • Results prone to subjectivity • Does not always provide information of value to the epidemiologists (i.e. highly common patterns)
Multilocus Sequence Typing (MLST)	• Characterisation of bacterial isolates via examination of the sequences of internal fragments (approximately 450-500 bp) of seven house-keeping genes using PCR and sequence analysis	• Unambiguous data • Highly reproducible • Large historical database	• Expensive • Labour-intensive • Time-consuming • Insufficient discriminatory power among <i>E. coli</i> O157:H7 isolates
Multiple Locus Variable Number Tandem Repeat Analysis (MLVA)	• Utilizes the naturally occurring variation in the number of repeating DNA sequences present in the <i>E. coli</i> O157:H7 genome known as VNTRs • Based on the determination of the number of repeat copy units in specific VNTR loci using multiplex PCR amplification of the repeat regions followed by sizing of the PCR product using conserved primers and high resolution capillary electrophoresis	• High-throughput • Rapid-turnaround time • Cost effective • Automation of results • Offers increased discriminatory power through the use of highly variable target sequences (VNTRs)	• Not a universal typing method • Limited historical data

Based on the fact that Shiga toxins produce characteristic cytotoxic effects on Vero cells, which are derived from African green monkey kidney epithelial cells, these cells are utilized as the primary cell line in cytotoxicity cell culture assays (Chapman, 2000; Park *et al.*, 1996). The assay is performed in a flat-bottomed 96-well microtiter plate in which presumptive *E. coli* O157:H7 culture filtrates grown in Brain Heart Infusion (BHI) broth are added to a suspension of Vero cells (Rahn *et al.*, 1996). The microtiter plates are examined at 24 and 48 hours to estimate the degree of cytotoxicity, which is a measurement of the percent cell death (Park *et al.*, 1999; Rahn *et al.*, 1996). When compared to the standard control well, those wells displaying 50 percent or greater cytotoxicity (rounding and clumping of cells in the monolayer) are considered to be positive for Shiga-toxin (ST) production (Rahn *et al.*, 1996).

To determine which Shiga toxins are present, neutralization assays using commercially available monoclonal antibodies specific for Stx1 and Stx2 are utilized. In accordance with PulseNet Canada SOP # E-TT-PR-004, if isolates produce only one Shiga toxin, neutralization will be observed for both the single antisera and for the mixture of the antibodies (Public Health Agency of Canada, 2013). Those samples that are not neutralized by the antisera may require repeat testing with a more dilute sample; however, if results remain unchanged, the *E. coli* O157:H7 sample may be producing some other type of cytotoxin. If cell culture produces a positive result for a toxin, but however, the PCR is negative, the toxin gene sequence may contain a mutation that prevents it from being detected using the DNA-based method. In contrast, if the cell culture is negative for toxin, but the PCR results indicate the gene is present, a gene mutation or repressor may be preventing expression of the toxin.

Although the Vero cell culture has proven to be useful in detecting ST, this technique has numerous disadvantages; the process takes approximately 5 days to obtain a result, the assay is prone to bacterial contamination, and a number of other toxins that may be present in the fecal sample such as *Clostridium perfringens* enterotoxin, *Clostridium difficile* cytotoxin and *E. coli* heat-labile enterotoxin could affect cultured Vero cells (Chapman, 2000).

1.5.2 Pulsed-field Gel Electrophoresis (PFGE)

Pulsed-field Gel Electrophoresis (PFGE) is a universal, comparative, subtyping method and the current gold standard method implemented by PulseNet Canada and partner laboratories for surveillance and outbreak detection of *E. coli* O157:H7 in Canada (Goering, 2010). PFGE was selected as the molecular subtyping method for enteric pathogens associated with foodborne illnesses after its utility for outbreak investigations was demonstrated in a study of *E. coli* O157:H7 isolates associated with one of the largest outbreaks in the Western United States in 1993 by effectively discriminating outbreak from sporadic cases (Barrett *et al.*, 1994; Ribot *et al.*, 2006). Since then, the method has provided high discriminating power and epidemiological concordance, and has revolutionized public health practice (Barrett *et al.*, 2006; Tauxe, 2006).

The foundational principle behind PFGE is based upon the variable size-dependent migration of large DNA restriction fragments in a periodically pulsing electric field of alternating polarities (Goering, 2010). Since data reproducibility and comparability are vital in the successful implementation of any molecular subtyping system, the PFGE protocol, in addition to the data analysis system, are rigorously standardized at PulseNet Canada to ensure the results generated are comparable and of the highest quality possible (Ribot *et al.*, 2006). The

standardized PFGE protocol (SOP # PNC-PR-210) for *E. coli* O157:H7 has been developed with two restriction enzymes that identify specific hexanucleotide recognition sequences in the genomic DNA (Public Health Agency of Canada, 2014). The primary enzyme, *XbaI* (isolated from *Xanthomonas badrii*), is used to screen for relatedness of isolates and the secondary enzyme, *BlnI* (isolated from *Brevibacterium linens*), is used to confirm the results obtained by the first enzyme and to provide supplemental discrimination.

Another essential feature of PulseNet Canada's standardized PFGE operating procedure is the implementation of a universal DNA fragment size standard, which is typically run in the first, fifth and last lanes in all gels (Gerner-Smidt *et al.*, 2006). The reference standard, which consists of *XbaI* restriction fragments of *Salmonella* Braenderup strain H9812 in the approximate range of 1,135 kb to 20.5 kb, contains 15 bands that are sufficiently evenly distributed to allow for consistent and reproducible normalization of all PFGE images run according to the PulseNet Canada protocol for *E. coli* O157:H7. The electrophoresis running conditions are the same for both primary and secondary restriction enzymes; however, depending on the size of the gel, complete migration of the macrorestricted DNA can range from 18 hours to 21 hours using the CHEF (contour-clamped homogenous electric field electrophoresis) DR III PFGE system.

Following electrophoresis, the gel is stained with GelRed and then de-stained with deionized water and the banding pattern is observed under ultraviolet (UV) light illumination in order to obtain a digital image of the gel (Gerner-Smidt *et al.*, 2006; Ribot *et al.*, 2006). PFGE patterns are analyzed using BioNumerics software and then uploaded to the National PulseNet

Canada EcoliAdmin database to generate a comparative analysis for the purpose of designating pattern names according to the standard nomenclature system (Hyytia-Trees *et al.*, 2006).

A change in the molecular size of the macrorestricted genomic *E. coli* O157:H7 DNA will ultimately result in a corresponding change in electrophoretic migration rate, which may or may not be visually detected depending upon the magnitude of the difference (Goering, 2010). Changes in fragment size result in different PFGE pattern combinations and are likely attributable to genetic events such as insertions, deletions, rearrangements and substitutions (Goering, 2010). It is hypothesized that PFGE pattern differences among strains of *E. coli* O157:H7 are most often caused by discrete insertions or deletions containing the restriction enzyme 'cut site' as opposed to single-nucleotide polymorphisms in the sites themselves (Kudva *et al.*, 2002). For example, a deletion or insertion within a restriction site would increase or decrease the migration of the associated fragment within the PFGE pattern, respectively, giving rise to a unique banding pattern (Goering, 2010). Conversely, rearrangements or substitutions within the genomic DNA that do not directly affect a restriction site would essentially not produce any alterations in the PFGE pattern (Goering, 2010).

If DNA macrorestriction is incomplete, the PFGE pattern may contain fragments of expected size as well as larger fragments possessing one or more restriction sites as a result of unsuccessful cleavage (Barrett *et al.*, 2006). These fragments typically appear as additional bands near the top of the PFGE gel that are typically of a weaker intensity than the surrounding bands and therefore, are often referred to as 'ghost' bands or artifacts (Barrett *et al.*, 2006). PFGE patterns differing by a single band are quite common and are posited to occur as a result

of the incorporation of an extra plasmid within one of the *E. coli* O157:H7 isolates (Barrett *et al.*, 2006).

Although the success of PFGE as the current gold standard for subtyping *E. coli* O157:H7 is well documented, the method does suffer from some limitations (Gerner-Smidt *et al.*, 2006; Noller *et al.*, 2003). While PFGE is relatively simple and inexpensive, it is quite labour intensive (strict DNA preparation requirements), the process takes several days to complete (typically four days), it cannot easily handle large sample sets, pattern designations may be biased due to the subjective interpretation of banding patterns, and the subtyping data does not always provide information of value to the epidemiologists, especially when the PFGE pattern is highly common (Sabat *et al.*, 2013). The lack of pattern diversity among PFGE patterns can make it difficult to separate potential outbreak-related cases from geographically and temporally associated sporadic cases, often leading to complications during the epidemiological investigation (Gerner-Smidt *et al.*, 2006).

1.5.3 Alternative DNA-based Molecular Subtyping Techniques

To overcome these limitations, alternative DNA-sequence based methods have been explored. These methods are appealing as they have the advantage of being easily standardized and automated unlike a majority of fingerprinting-based methods. Multilocus sequence typing (MLST) examines the internal sequences (of approximately 450-500 bp) of (usually) seven housekeeping genes using PCR and sequence analysis (Sabat *et al.*, 2013). The unique sequence (allele) associated with each housekeeping gene is assigned an arbitrary number, and based on the combination of identified alleles (i.e. allelic profile), the sequence type (ST) is determined (Sabat *et al.*, 2013). Due to the instilment of an internationally

standardized nomenclature, the data generated by this method is not only unambiguous, but also highly reproducible (Sabat *et al.*, 2013). As well, the availability of large databases containing allele sequences and ST profiles allows researchers to determine the genetic relatedness between bacterial strains within a species and to track isolates associated with specific STs (Sabat *et al.*, 2013).

Although MLST has been successfully used for the differentiation of a number of organisms, such as *Streptococcus pneumoniae*, and *Neisseria meningitides*, in reference to a scientific paper published in 2003, the researchers found that MLST showed a striking lack of diversity among *E. coli* O157:H7 isolates that were found to be diverse by PFGE (Noller, McEllistrem and Stine *et al.*, 2003). In addition to the poor discriminatory power of MLST for *E. coli* O157:H7, the method is also expensive, labour-intensive, and time-consuming and therefore, alternative targets for strain differentiation have been evaluated, such as variable-number tandem repeat (VNTR) loci (Noller *et al.*, 2003).

Identified in many microbial genes and intergenic regions in essentially all pro- and eukaryotic species, VNTR loci consist of repetitive DNA sequences or motifs organized in tandem that exhibit a high degree of polymorphism with respect to the number of repeat units present (Noller *et al.*, 2006, Hyytia-Trees *et al.*, 2010). Highly variable VNTR loci exist in the *E. coli* O157:H7 genome and are considered to be excellent estimators of genetic relationships, particularly for closely related isolates (Keys *et al.*, 2005). Since the variation generated at VNTR loci is known to provide an increased level of subtyping discriminatory power, VNTRs have the potential to serve as useful epidemiological markers (Vogler *et al.*, 2006). VNTR analysis may only include a single locus, however, the use of multiple loci to estimate

evolutionary relationships tends to provide a higher analysis confidence by lessening the complications caused by convergent evolution at a particular locus (Keys *et al.*, 2005). Hence, the study of multiple VNTR loci is known as Multiple Locus Variable Number Tandem Repeat Analysis (MLVA), and is currently in pilot implementation as a secondary subtyping tool to be used in combination with PFGE in order to provide further discrimination during analysis of clusters and outbreak investigations of *E. coli* O157:H7 in Canada.

1.5.4 Multiple Locus Variable Number Tandem Analysis (MLVA)

Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) is a molecular subtyping method that utilizes the naturally occurring variation in the number of repeating DNA sequences present in the *E. coli* O157:H7 genome known as VNTRs (Noller *et al.*, 2003). The foundational principle of MLVA is based on the determination of the number of repeat copy units in specific VNTR loci using multiplex PCR amplification of the repeat regions followed by sizing of the PCR product using conserved primers and high resolution capillary electrophoresis (Hyytia-Trees *et al.*, 2006). The use of different fluorescent dyes and amplicon size ranges allows multiple loci to be multiplexed and analyzed simultaneously (Vogler *et al.*, 2006). Differences in amplicon size observed among loci is attributed to the variation in repeat copy number at that particular locus and the peak patterns are scored accordingly using the appropriate scientific software (Noller *et al.*, 2006).

Once elucidated, the VNTR repeat copy number occurring at each locus is combined into a discrete dataset of integers to produce an MLVA profile (Hyytia-Trees *et al.*, 2006). As illustrated by Keys and colleagues, there are thousands of potential VNTR loci in the *E. coli* O157:H7 genome (Keys *et al.*, 2005). Based on the findings of their study, researchers

developed an MLVA scheme targeting 29 polymorphic VNTR regions that were found to discriminate among closely related *E. coli* O157:H7 isolates demonstrating highly similar PFGE patterns (Keys *et al.*, 2005). To render the typing scheme suitable for routine use by public health laboratories, the protocol was modified and optimized to include only 9 VNTRs, which was determined to be the minimum number of loci needed to obtain maximum discrimination among unrelated isolates (Hyytia-Trees *et al.*, 2006). As reported by Hyytia-Trees and participating researchers, adding one or more of the remaining 20 loci did not provide a noticeable improvement to the discriminatory power of the subtyping assay (Hyytia-Trees *et al.*, 2006).

The overall discriminatory power of the nine locus MLVA system appeared to be equal to that of PFGE when two restriction enzymes were utilized (Hyytia-Trees *et al.*, 2006). To keep the protocol simple and cost-effective, eight VNTRs, which have been declared sufficiently discriminatory for *E. coli* O157:H7, have been incorporated into a standardized MLVA protocol and include VNTR 3, 34, 9, 25, 17, 19, 36 and 37, all of which have been arbitrarily named by researchers (Table 2) (Noller *et al.*, 2003; Keys *et al.*, 2005; Hyytia-Trees *et al.*, 2006). Of noteworthy importance, it is presumed VNTR 10 was removed from the protocol as it was found to offer limited information to strain differentiation due to the highly variable nature of this loci (Vogler *et al.*, 2006). The VNTRs included in the final PulseNet/PulseNet Canada protocol (SOP # PNC-PR-404) are relatively short, possessing repeat units of 6 or 7 base pairs with the exception of VNTR 34, which is 18 base pairs long (Public Health Agency of Canada, 2014). Short VNTRs were chosen to be included in the final MLVA protocol for two reasons; (1) shorter VNTRs are often associated with an increased potential of 'replication slippage', which

results in either the loss or gain of a repeat motif, (2) shorter repeats are thought to facilitate automation by reducing the potential overlap of different loci (Noller *et al.*, 2003). Since VNTRs are rapidly evolving genetic elements, multiple MLVA profiles tend to emerge during an outbreak initially caused by a single clone of *E. coli* O157:H7 (Noller *et al.*, 2003). Multiple factors influence the frequency and type of mutation occurring at VNTR loci such as the number of repeating motifs, the nucleotide length of the copy units and the composition of the VNTR array (Noller *et al.*, 2006).

Variation in the number of repeating DNA sequences at VNTR loci is attributed to insertion and/or deletion mutations that result in the gain or loss of repeat motifs leading to the generation of highly variable VNTR loci (Vogler *et al.*, 2006). These mutations are thought to arise through slippage and mispairing during DNA replication due to the occasional dissociation of DNA polymerase (Noller *et al.*, 2006). During the replication process, as DNA polymerase copies the template, the enzyme may stutter in areas where tandem repeats are located, resulting in an increase in the number of repetitive sequences (Noller *et al.*, 2006). This process is often referred to as backward replication slippage. If the tandem repeat region contains numerous repeats, loops will often form when the DNA is single stranded thereby hindering effective replication by DNA; however, deletion events are statistically more likely to take place when a mutation involves greater than 4 to 5 repeat units at a VNTR locus (Vogler *et al.*, 2006). It is important to note that although a majority of VNTR loci are located in protein coding regions, variations in repeat units have not been linked with documented biological or phenotypic effects thus far, and therefore are likely to be selectively neutral (Vogler *et al.*, 2006).

Table 2. Characteristics of Eight Variable-Number Tandem Repeats (VNTRs) used in the Standardized Multiple Locus Variable Number Tandem Repeat Analysis (MLVA) Protocol for *Escherichia coli* O157:H7 at PulseNet Canada (based on data presented by Hyytiä-Trees *et al.* 2006).

VNTR ^a	Repeat Size (bp)	Tandem Repeat Sequence ^b	Offset ^c (bp)	No. of Repeats Min. Max.	Average Mutation Rate ^d	Diversity ^e	Inside ORF ^f
3	6	AAGGTG	321	2 25	3.2 x 10 ⁻⁵	90.9	Yes (Hypothetical Protein)
34	18	TTAAATAATCTACAGAAG	100	4 12	< 3.4 x 10 ⁻⁶	71.9	Yes (Hypothetical Protein)
9	6	AAATAG	465	1 25	3.6 x 10 ⁻⁵	91.0	No
25	6	TGCAAA	110	1 13	1.0 x 10 ⁻⁵	58.9	No
17	6	TATCTT	120	2 21	2.0 x 10 ⁻⁵	78.3	Yes (Hypothetical Protein)
19	6	GACCAC	272	2 14	7.0 x 10 ⁻⁶	73.9	Yes (Hypothetical Protein)
36	7	ACCTCAC	102	3 20	2.3 x 10 ⁻⁵	86.9	No (on pO157 plasmid)
37	6	GCTACT	142	3 22	1.8 x 10 ⁻⁵	74.5	Yes (Hypothetical Protein on pO157 plasmid)

^aVNTR loci by Keys *et al.*, 2005.

^bRepeat sequences as defined by Noller *et al.*, 2003 and Lindstedt *et al.*, 2003.

^cThe sequence defined by the primer not including the VNTR array from Hyttia-Trees *et al.*, 2006.

^dMutation rate obtained from Vogler *et al.*, 2006.

^eSimpson's index of diversity [$D = 1 - \sum (\text{allele frequency})^2$] as calculated by Hyttia-Trees *et al.*, 2006.

^fORF identity obtained from Keys *et al.*, 2006.

VNTR, variable number tandem repeat; bp, base pair; ORF, open reading frame.

The high throughput, unambiguous data provided by MLVA is postulated to offer better discrimination through the use of highly variable target sequences (i.e. VNTRs). A major limitation of this method is that it is not a universal subtyping system, and therefore, primers must be designed specifically for each pathogenic species targeted (Sabat *et al.*, 2013). Also, it is possible to standardize the method, but this takes careful planning and coordination to ensure inter-laboratory results can be compared (Nadon *et al.*, 2013). For this reason, MLVA is highly unlikely to entirely replace PFGE in epidemiological investigations, but however, could potentially serve as a valuable supplemental tool during outbreak detection and response (Sabat *et al.*, 2013).

1.6 *Escherichia coli* O157:H7 Surveillance and Outbreak Response in Canada

Public health surveillance involves the systematic collection, analysis and interpretation of data relating to the morbidity and mortality associated with pathogenic microorganisms, such as *E. coli* O157:H7, to ensure timely dissemination of the information for public health action (World Health Organization, 2008). Essentially, outbreak detection relies on routine typing of isolates of foodborne pathogens in order to detect a surge in a particular subtype and link potentially related cases of illness (World Health Organization, 2008). Hence, the primary goal of surveillance is to promptly identify clusters of disease to prevent both ongoing transmission and similar outbreaks from occurring in the future (World Health Organization, 2008).

1.6.1 National Enteric Surveillance Program (NESP)

The National Enteric Surveillance Program, simply referred to as NESP, monitors the occurrence of pathogenic microorganisms by providing real-time detection and response through timely analysis and reporting of laboratory-confirmed isolations of bacterial, viral, and parasitic pathogens in Canada, at the species or serotype level (Public Health Agency of Canada, 2012). Provincial public health laboratories submit genus, species and serotype information pertaining to enteric microorganisms isolated from human patients either directly to the National Microbiology Laboratory (NML) at the Public Health Agency of Canada in Winnipeg, MB, or via a web-based application (webNESP) through the Canadian Network for Public Health Intelligence (CNPHI) (Public Health Agency of Canada, 2012). The collaborative laboratory data is summarized in weekly analyses that are reported to provincial/territorial and federal partners and provides alerts of any significant increases in the number of enteric illnesses in Canada. All reports are shared on CNPHI's Canadian Laboratory Surveillance Network (CLSN) between provincial public health laboratories, the Canadian Food Inspection Agency (CFIA), Health Canada (HC), PHAC, and provincial/territorial epidemiologists in order to permit interpretation of clinical microbiological evidence during multi-provincial epidemiologic investigations (Public Health Agency of Canada, 2012).

Annual reports consisting of the cumulative weekly data collected from all provincial public health laboratories are also produced in order to provide a comprehensive analysis of longer term national trends regarding the incidence of enteric pathogens in Canada (Public Health Agency of Canada, 2012). Despite the usefulness of the NESP data for surveillance, there are some inherent limitations of this system. Since isolates are not always referred from

the regional and local laboratories to the provincial public health laboratories (i.e. *Campylobacter*), the data provided by the provinces to NESP may not reflect all cases, resulting in an under-representation of the true incidence of disease in Canada (Public Health Agency of Canada, 2012). However, *Salmonella* and *E. coli* O157 isolates captured by NESP tend to be more representative of the true incidence of disease in Canada as the number of cases for both pathogens often coincide with those reported to the Canadian Notifiable Diseases and Surveillance System (CNDSS), which is a complimentary surveillance system that receives data collected on a mandatory basis by local health units (Public Health Agency of Canada, 2012).

As well, outbreaks and clusters reported to NESP do not necessarily represent all the enteric illness outbreaks that may have occurred nationally; additionally, the case counts reported to NESP are not typically representative of the actual final numbers of cases included in those clusters/outbreaks (Public Health Agency of Canada, 2012). Therefore, details regarding clusters/outbreaks are not included in the NESP report as they are more accurately tracked within PulseNet Canada.

1.6.2 PulseNet Canada

In response to the increasing demand for rapid, high discriminatory typing to accelerate the detection and investigation of outbreaks associated with *E. coli* O157:H7, PulseNet, which is the national molecular subtyping network for foodborne disease surveillance, was established in 1996 in the United States and in 2000 in Canada (Gerner-Smidt *et al.*, 2005). Implemented as an early warning system, PulseNet Canada provides real-time laboratory surveillance of all *E. coli* O157: H7 cases in Canada and is headquartered and coordinated by PHAC at the NML. PulseNet Canada has become a valuable tool for detection, investigation and subsequent

control of outbreaks of foodborne illnesses associated with *E. coli* O157:H7, in addition to other enteric pathogens, such as *Salmonella*, *Shigella*, and *Listeria* species (Gerner-Smidt *et al.*, 2005). Other members of PulseNet Canada include the Laboratory of Foodborne Zoonoses (LFZ) of PHAC, the Canadian Food Inspection Agency, Health Canada, and ten provincial public health laboratories located across the nation (Swaminathan *et al.*, 2006). The main objectives of PulseNet Canada are: (1) to detect clusters of cases (≥ 2) with matching DNA fingerprints, (2) facilitate early identification and investigation of foodborne disease outbreaks, (3) assist in differentiating outbreak from geographically and temporally associated sporadic cases in epidemiological investigations, and (4) to ultimately identify and confirm the source of the outbreak by providing a rapid communication platform that links both provincial and federal public health laboratories across the nation (Swaminathan *et al.*, 2006).

Through the electronic database, PulseNet Canada is able to accelerate the outbreak investigation by providing real-time subtyping for surveillance and outbreak detection by tracking DNA fingerprints of *E. coli* O157:H7, which is the single most critical component in human foodborne disease outbreak investigations (Gerner-Smidt *et al.*, 2005; Swaminathan *et al.*, 2006). To obtain these DNA fingerprints, PulseNet Canada utilizes genotypic methods, such as Pulsed-field Gel Electrophoresis (PFGE) to perform molecular subtyping on all *E. coli* O157:H7 isolated from clinical cases using standardized methodologies, equipment and software (Gerner-Smidt *et al.*, 2006; Swaminathan *et al.*, 2006). Briefly, subtyping results obtained from DNA-fingerprints are uploaded to the National EcoliAdmin database in the form of PFGE pattern combinations and are directly accessible to certified participating laboratory members through the connected database. Provincial labs independently post clusters of cases with matching

PFGE patterns in their jurisdictions to the Canadian Laboratory Surveillance Network (CLSN) discussion board on the Canadian Network for Public Health Intelligence (CNPHI), in addition to viewing and responding to posting on a daily basis. PulseNet Canada database managers perform weekly analyses of clusters with indistinguishable or highly similar patterns and post all multi-provincial clusters to the discussion board, in addition to reporting the results to the participating lab(s) and PHAC epidemiologists (Gerner-Smidt *et al.*, 2006).

For all clusters identified, both local laboratory and national databases are scanned by the respective database managers for clinical cases, as well as for any food isolates with matching PFGE patterns within the last 60 days of when the first cluster isolate was documented. A 60 day timeframe is recognized as the gold standard amount of time permissible for clinical cases to be considered related to a particular cluster or outbreak as it provides ample time for crucial events to occur such as, exposure, clinical manifestations, medical treatment, isolation and laboratory testing of the pathogen (Gerner-Smidt *et al.*, 2006). It is important to note that the results generated in PulseNet Canada cannot solely determine whether a cluster or outbreak is occurring, and therefore, the laboratory results must be confirmed by epidemiological data (Gerner-Schmidt *et al.*, 2005).

Decisions regarding relatedness of *E. coli* O157:H7 isolates in surveillance and outbreak detection is based on the ability to compare restriction fragment banding patterns; therefore, the gel must be of sufficient quality to be interpreted thoroughly and accurately (Barrett *et al.*, 2006; Goering, 2010). A PFGE gel that contains partial digestions and other artifacts or bands that are not sharp and clear are typically re-run since interpreting a poor quality gel can often be misleading (Barrett *et al.*, 2006). If any distinguishable differences in the banding patterns

are observed, the PFGE patterns are assigned different designations; similarly, patterns that are indistinguishable from one another are reported as such (Barrett *et al.*, 2006). All communication occurring between PulseNet Canada and other members of the molecular subtyping network including epidemiologists and other provincial and federal laboratories must be precise with respect to PFGE pattern designations (Barrett *et al.*, 2006). To ensure accurate delivery of information, PulseNet Canada utilizes a standardized PFGE pattern naming system for all *E. coli* O157:H7 isolates analyzed, which consists of nine characters in the format 'XXYYY.####' (Barrett *et al.*, 2006). The first two characters (XX) represent the microorganism (e.g. EC is code for *E. coli* O157:H7); the next three characters (YYY) indicate the enzyme used for macrorestriction of the genomic DNA (e.g. XAI represents the enzyme *Xba*I and BNI represents the enzyme *Bln*I); the last four characters to the right of the decimal place (####) are consecutive numbers allocated to PFGE patterns as they are detected (Barrett *et al.*, 2006).

Once PFGE pattern names have been assigned, the results are interpreted based on four main criteria: diversity of the microorganism, degree of differentiability by two restriction enzymes *Xba*I and *Bln*I, historical frequency of the pattern combination and consistency with additional subtyping results (Table 3) (Health Canada, 2011). However, the results generated by PFGE alone cannot adequately determine whether an outbreak is occurring and therefore, must be supported by strong epidemiological evidence (Barrett *et al.*, 2006). To provide additional discriminatory power to analyses of clusters/outbreaks of *E. coli* O157:H7 in Canada, PulseNet Canada has recently explored the use of MLVA as a secondary subtyping method in combination with PFGE. Essentially, the key to MLVA is centered on the fact that different loci demonstrate differing rates of evolution. The use of fast evolving loci provides discrimination

among closely related isolates, but however, offers limited value between distant strains; conversely, the use of more slowly evolving loci overcomes the convergent evolution complication but does not provide optimal discrimination between closely related isolates (Keys *et al.*, 2005).

It is hypothesized that VNTR loci with low to medium diversities are most likely contributing to strain differentiation at the PFGE level, while few loci with higher diversities provide discrimination among closely related isolates (Keys *et al.*, 2005; Vogler *et al.*, 2006). Based on the interpretation guidelines developed and implemented by PulseNet Canada, isolates demonstrating profiles differing by up to 3 repeats at a single locus or 1 repeat at 3 loci and no variation at VNTR 34 are considered to be genetically related (Nadon and Rumore, 2013; Rumore *et al.*, 2013).

1.6.3 Investigation of Outbreaks

When a foodborne illness outbreak is suspected, ranging from a local outbreak of a small number of linked cases with mild disease to a nationwide/international outbreak of severe disease, follow-up procedures are initiated and may involve various levels of public health resources (municipal, provincial, federal) in order to link the illness to a common source (Health Canada, 2011). To validate or refute the existence of an outbreak, all available information obtained from the case interviews, must be promptly assessed (World Health Organization, 2008). Case interviews are critical for providing a clear picture of the clinical and epidemiological features of the affected group; therefore, interviews are required to be conducted in a timely manner as delays often lead to recall bias or to the inability to remember food consumption history, which can obscure the investigation (World Health Organization,

Table 3. PulseNet Canada Interpretation Guidelines for Determining the Level of Relatedness between Food and Clinical Isolates based on Pulsed-field Gel Electrophoresis (PFGE) Results (based on data presented by Health Canada *et al.*, 2011).

Criteria	Nature of Evidence	Weight
Diversity by PFGE	Based on historical data, sporadic cases display diverse PFGE patterns and therefore, are suitably diverse by PFGE	Strong
	Little or no historical data exists for the microorganism	Weak
Level of distinguishability by two restriction enzymes	Based on historical data, sporadic cases display indistinguishable or highly similar PFGE patterns, and therefore are not suitably diverse by PFGE PFGE patterns have indistinguishable primary and secondary restriction enzymes	Strong
	PFGE patterns are indistinguishable for the primary restriction enzyme and distinguishable for the secondary restriction enzyme	
	PFGE patterns are distinguishable for primary and secondary restriction enzymes, with minor differences in the lower molecular weight region of the banding pattern	Weak
Historical frequency of the PFGE pattern combination	PFGE pattern combinations do not match by multiple bands, particularly in the higher molecular weight region of the banding pattern The PFGE pattern or pattern combination is new	Strong
	Based on historical frequencies, the PFGE pattern is not common	
	Based on historical frequencies, the PFGE pattern is common	Weak
Consistency of PFGE results with additional subtyping data	Additional subtyping data are in agreement with PFGE results Additional subtyping data are not available Additional subtyping data are not in agreement with PFGE results	Strong Weak

2008). During the investigation, it is imperative to demonstrate that the isolate causing the human illness is indistinguishable from the isolate collected from the implicated food source (obtained from case interviews) as it is essential to guide further investigation (Health Canada, 2011). To achieve this, detailed comparisons between all isolates collected are typically performed using molecular typing techniques such as PFGE, which is the current gold standard method for isolate comparison during surveillance and outbreak detection and response (Health Canada, 2011).

Once a suspected food source has been linked to clinical cases, further investigations are initiated to determine the origin of the food product (traceback) and/or where the food product was distributed (traceforward) in order to inform a risk management decision and generate preliminary hypothesis about the potential cause of the human illness (World Health Organization, 2008). To ensure successful investigation and control of foodborne disease outbreaks, an Outbreak Investigation Coordinating Committee (OICC) is often activated to ensure all members involved in the investigation clearly understand the roles and responsibilities to be undertaken by each sector of public health during the investigation (Public Health Agency of Canada, 2010). Following the identification of an outbreak, a case definition is established, which is used as an epidemiological tool for counting cases as it outlines a set of criteria for determining whether an ill person should be included in the investigation (World Health Organization, 2008).

Ideally, a case definition should include all outbreak related cases (high sensitivity) and exclude all non-outbreak related cases (high specificity) (World Health Organization, 2008). Although a highly sensitive case definition will essentially detect all individuals who truly have

disease associated with the outbreak, it would also be more likely to include individuals who have a foodborne illness but are not associated with the outbreak; while conversely, a more specific case definition will capture only those who truly are part of the outbreak, but however, is more likely to miss out on some cases (World Health Organization, 2008). During the initial stages of an outbreak investigation, the aim is to detect as many cases as possible and therefore, requires a more sensitive case definition (World Health Organization, 2008). As epidemiological evidence and laboratory-confirmed data become more readily available, the overall clarity of the outbreak is significantly improved, thereby allowing investigators to implement a more specific case definition (World Health Organization, 2008). However, some argue the reverse.

During the early stages of hypothesis generation and testing, a very specific case definition might be the most useful for detecting common exposures; whereas later in the investigation, for example, after a source has been identified, it would be appropriate to expand to a highly sensitive case definition in order to measure the true scope of the outbreak (Besser, 2011; Hyytia-Trees, 2009). Since a single case definition that satisfies all needs is rare, it is not uncommon for case definitions to change during an outbreak investigation or for different case definitions to be applied for different purposes (World Health Organization, 2008). Typically, investigators use the following case definitions in parallel: (1) confirmed cases; demonstrate a laboratory-confirmed positive result, (2) probable cases; demonstrate the clinical features of the disease but do not have laboratory confirmation, (3) possible cases; demonstrate fewer or atypical clinical features (World Health Organization, 2008). For foodborne disease outbreak in Canada, the case definitions typically include laboratory data

including molecular subtype results. Once investigators have identified an association between a particular food source and transmission of the suspected pathogen, a recall of the implicated product is conducted and public health alerts are issued informing consumers of the contaminated products and actions that should be taken to prevent further exposure and illness (World Health Organization, 2008). Following the implementation of control measures, the OICC must decide when an outbreak is over and a structured review of the investigation should be conducted in a formal debriefing (Who 2008).

The main goals of the debriefing are to ensure the instituted control measures were effective for the outbreak, discuss needs to optimize future outbreak response, identify factors that may have compromised the investigation and seek solutions, discuss any legal issues that may have arisen, and arrange for completion of the final outbreak report (World Health Organization, 2008). The final outbreak report/summary, which includes a statement about the effectiveness of the investigation, the control measures taken, and recommendations for the future, is circulated to the appropriate public health authorities and institutions for the primary purpose of providing a comprehensive review of the outbreak (World Health Organization, 2008).

1.6.4 Limitations of the Current System

PFGE has been used with great success by PulseNet Canada for detecting and responding to outbreaks of *E. coli* O157:H7 for over 10 years (Gerner-Smidt *et al.*, 2006). The robust historical PFGE data contained within the PulseNet Canada database from greater than 10,000 surveillance, outbreak and food isolates has provided the underlying foundation of the molecular subtyping network (Gerner-Smidt *et al.*, 2006). New or rare PFGE patterns provide

the strongest evidence that an outbreak may be occurring and therefore, additional subtyping data is not typically required for further identification of outbreak-related cases or to confirm the source of contamination (Hyytia-Trees *et al.*, 2006). However, PFGE does not always provide optimal discrimination for closely related isolates and therefore, differentiating sporadic cases from outbreak cases demonstrating highly similar/indistinguishable pattern combinations can become difficult.

In the early 2000s, reports began to emerge that indicated MLVA could be useful in discriminating among epidemiologically unrelated isolates that were indistinguishable by PFGE (Keys *et al.*, 2005). In one study, 80 isolates, including 21 isolates from 5 epidemiologically well-characterized outbreaks from Pennsylvania and Minnesota, were analyzed by both PFGE and MLVA (Noller *et al.*, 2003). The results indicated that MLVA was able to effectively distinguish outbreak from sporadic cases, as well as differentiate unrelated isolates that may have been falsely clustered together by PFGE (Noller *et al.*, 2003). The ability of MLVA to distinguish among *E. coli* O157:H7 isolates that appeared to be highly related by PFGE infers MLVA is a more discriminatory subtyping assay than PFGE. It was postulated that use of MLVA during routine public health surveillance may lead to fewer false-positive signals suggestive of an outbreak (Noller *et al.*, 2003). In a separate study conducted by Hyytia-Trees and fellow researchers, the utility of MLVA was shown to be particularly useful in discriminating among epidemiologically unrelated isolates, which were otherwise indistinguishable by PFGE (Hyytia-Trees *et al.*, 2006). These findings further indicated the usefulness of MLVA as a complement to PFGE due to the ability of the assay to discriminate among closely related isolates that cannot be distinguished by PFGE alone (Hyytia-Trees *et al.*, 2006).

The easily standardized, high throughput, unambiguous data provided by MLVA offers better discrimination through the use of highly variable target sequences such as VNTRs (Hyytia-Trees *et al.*, 2006). The lack of pattern diversity among PFGE pattern combinations often makes it difficult to separate potential outbreak related cases from geographically and temporally associated sporadic cases. However, *E. coli* O157:H7 isolates demonstrating new or rare PFGE patterns often may not typically require additional subtyping results since these uncommonly encountered patterns may provide a significant amount of evidence autonomously (Hyytia-Trees *et al.*, 2006; Keys *et al.*, 2005). In such instances, it was suggested that MLVA may be useful as a secondary subtyping method by providing further confirmation of the results obtained by PFGE, but however may not add further information to the outbreak investigation per se (van Belkum, 2007). Inclusion of MLVA results in the case definition is not typically expected to delay confirmation of outbreak-related cases as this molecular subtyping technique is faster than PFGE by at least one day (Hyytia-Trees *et al.*, 2006).

Despite the usefulness of the subtyping technique, historical MLVA data is lacking in Canada, and multiple clusters exist in the database for which no epidemiological relationships have been established (Hyytia-Trees *et al.*, 2006). While some limited data exist on the performance of the method, this work was primarily conducted in other countries and had not been evaluated or validated on strains of *E. coli* O157:H7 that occur in Canada. The shortage of historical MLVA data highly limits the ability of sound interpretation of these results and therefore, a more robust database is required for a full understanding of MLVA performance in order to further improve detection and response for *E. coli* O157:H7 outbreaks in Canada.

1.7 Current Study

The ideal molecular subtyping method would be 100 percent sensitive and specific, meaning all epidemiologically related isolates would be included in the cluster/outbreak, and all epidemiologically unrelated isolates would be excluded from the investigation (Ribot *et al.*, 2006). Although no currently available subtyping method meets these stringent criteria, well-established subtyping methods, such as PFGE, have been proven effective as an outbreak detection and response tool by providing high levels of sensitivity to investigations, but however, often lack in specificity, especially for closely related *E. coli* O157:H7 isolates. To enhance outbreak detection and response, DNA sequence-based methods, such as MLVA, have been explored and appear to provide increased discriminatory power for *E. coli* O157:H7 isolates demonstrating indistinguishable or highly similar PFGE pattern combinations. The hypothesis of the current study is that MLVA provides increased discriminatory power over PFGE, and when used as a secondary method in conjunction with PFGE, MLVA provides increased specificity for outbreak detection and response without significantly impacting the sensitivity. Hence, the main objectives of the study are to; (1) determine the overall discriminatory power of MLVA and in comparison to PFGE, (2) determine the level of congruency between PFGE and MLVA, (3) determine the expected variability among outbreak-related MLVA profiles, and (4) characterize the ability of MLVA to provide specificity to outbreak detection and response when used as a supplemental subtyping tool in conjunction with PFGE.

Chapter 2: Materials and Methods

2.1 Isolate Collection

A list of all *E. coli* isolates uploaded to the PulseNet Canada National EcoliAdmin PFGE database in BioNumerics v6.01 (Applied Maths, Belgium) from 2008 to 2012 (inclusive) was extrapolated for consideration against the sample selection criteria (n= 3,341). A total of 1,056 isolates were excluded: 747 were not available for further analysis, 133 were duplicates and the remaining 176 had been previously subtyped as non-O157 *E. coli* following their initial upload. In total, 2,285 *E. coli* O157:H7 isolates were readily available at the National Microbiology Laboratory (NML) for retrospective MLVA, of which 2,104 were clinical isolates, 166 were food isolates, 14 were non-human isolates, and one isolate was collected from an environmental sample.

2.2 Multiple Locus Variable Number Tandem Repeat Analysis (MLVA)

Three hundred forty nine *E. coli* O157:H7 isolates out of the total available isolates (n= 2,285), had already been previously subtyped using MLVA by the PulseNet Canada team. Thus, MLVA was performed on 1,936 isolates according to the standardized PulseNet Canada/International protocol as outlined in SOP # PNC-PR-404 (Public Health Agency of Canada, 2014). Briefly, *E. coli* O157:H7 isolates were cultivated on Columbia Blood Agar Base (CBAB) with 5 percent sheep blood (Thermo Scientific, Nepean, Ontario) and incubated at 37°C overnight. Boil cell preparations were prepared by transferring 3 medium sized, well isolated colonies into 100 microliters (µl) of UltraPure DNase/RNase free distilled water (Life Technologies, Burlington, Ontario) and boiled for 10 minutes at 99°C. Cell lysates were centrifuged at 16,000 rcf at room temperature for 2 minutes to sediment the cellular debris. The supernatant was extracted and used as the DNA template for the assay. All DNA templates

were stored at -20°C for future use. The eight VNTR loci were amplified in two multiplex PCR reactions (R1 and R2; Table 4). The PCR amplification was performed in a final volume of 10 µl containing 1.0 µl of 10X PCR buffer without magnesium (Life Technologies), 2.0 mM of magnesium chloride (MgCl₂) (Life Technologies), 0.2 mM of PCR Nucleotide Mix (Roche Applied Science, Indianapolis, Indiana), 1.0 U of Platinum *Taq* polymerase (Life Technologies) and 1.0 µl of DNA template. The concentrations for the fluorescently labelled forward and unlabelled reverse PCR primers (Biosearch Technologies, Novato, California) are listed in Table 4. The amplification conditions were as follows: an initial denaturation step at 95°C for 5 minutes, PCR amplification with 35 cycles of 94°C for 20 seconds, 65°C for 20 seconds and 72°C for 20 seconds and a final extension step at 72°C for 5 minutes.

All samples were run in duplicate. Positive control EDL933 and Internal Ladder strains EC04PN0139 and EC04PN0570 (Centers for Disease Control and Prevention, Atlanta, Georgia) were also run in duplicate for R1 and R2, in addition to a negative control (no template) for both R1 and R2. Following amplification, the PCR products were diluted 1:19 with UltraPure DNase/RNase free water. A 1.0 µl aliquot of each diluted PCR product was mixed with 8.0 µl of HiDi Formamide (Life Technologies) and 1.0 µl of GeneFlo 625-base pair DNA size standard (Chimerx, Madison, Wisconsin) and then denatured at 95°C for 3 minutes. PCR products were sized using either the Applied Biosystems 3130xl Genetic Analyzer (Life Technologies) or the Applied Biosystems 3730xl DNA Analyzer (Life Technologies). Data analysis was conducted using GeneMapper software (version 5), and the fragment size list generated by either the Applied Biosystems 3130xl Genetic Analyzer or the Applied Biosystems 3730xl DNA Analyzer

Table 4. PCR primers utilized in the PulseNet Canada/International *Escherichia coli* O157:H7 standardized MLVA Protocol (based on data presented by Hyytia-Trees *et al.*, 2006).

Reaction	Locus	Dye ^a	Forward Primer (5' to 3') ^b	Reverse Primer (5' to 3') ^b	Concentration (μM)
R1	VNTR-3	CalRed590	GG CGG TAA GGA CAA CGG GGT GTT TGA ATT G	GAA CAA CCT AAA ACC CGA CTC GCC ATC G	0.150
	VNTR-34	FAM	GA CAA GGT TCT GGC GTG TTA CCA ACG G	GTT ACA ACT CAC CTG CGA ATT TTT TAA GTC CC	0.066
	VNTR-9	FAM	GC GCT GGT TTA GCC ATC GCC TTC C	GTG TCA GGT GAG CTA CAG CCC GCT TAC GCT C	0.050
	VNTR-25	HEX	GC CGG AGG AGG GTG ATG AGC GGT TAT ATT TAG TG	GCG CTG AAA AGA CAT TCT CTG TTT GGT TTA CAC GAC	0.045
R2	VNTR-17	CalRed590	GC AGT TGC TCG GTT TTA ACA TTG CAG TGA TGA	GGA AAT GGT TTA CAT GAG TTT GAC GAT GGC GAT C	0.120
	VNTR-19	FAM	GC AGT GAT CAT TAT TAG CAC CGC TTT CTG GAT GTT C	GGG GCA GGG AAT AAG GCC ACC TGT TAA GC	0.018
	VNTR-36	FAM	GG CGT CCT TCA TCG GCC TGT CCG TTA AAC	GCC GCT GAA AGC CCA CAC CAT GC	0.013
	VNTR-37	HEX	GC CGC CCC TTA CAT TAC GCG GAC ATT C	GCA GGA GAA CAA CAA AAC AGA CAG TAA TCA GAG CAG C	0.028

^aThe 5' end of the forward primer is fluorescently labelled.

^bAll forward and reverse primer sequences are from Keys *et al.*, 2005.

PCR, polymerase chain reaction; VNTR, variable number tandem repeat.

was imported into BioNumerics v6.01 (Applied Maths, Belgium) O157 Ecoli MLVA-Admin database (version 6.01) for further analysis using a customized script. The script recorded all detectable fragments, in addition to the type of dye and the level of fluorescence associated with each fragment. The import script also created database entries for each isolate included in the fragment list and linked the appropriate fragment data to the corresponding entries. A second script was used to calculate the repeat copy number for each locus based on the size of fragment detected. If no amplification product was detected (null allele), the repeat copy number for that particular locus was designated as a -2.0 in the database. To ensure ease of interpretation of the results, null alleles were defined by a '0' in the MLVA profiles presented in this study.

2.3 Evaluation of the Discriminatory Power and Congruence of PFGE and MLVA

A quantitative comparison of the results obtained from each typing method under investigation was performed. To conduct the statistical analysis, the comparing partitions website was utilized and can be found at the following link (<http://darwin.phyloviz.net/ComparingPartitions/index>). The framework uses the Simpson's Index of Diversity (SID) and Adjusted Wallace (AW) coefficients as a quantitative way to evaluate the discriminatory power and congruence of laboratory results among different subtyping methods (Carrico *et al.*, 2006). As defined by Hunter and Gaston, the SID indicates the probability of two strains sampled randomly from a population belonging to two different types (Hunter and Gaston, 1988). Since the measured coefficients are dependent on the particular sample taken from the population, there is a certain level of variability inherent in the estimated values relative to those of the true populations (Carrico *et al.*, 2006). Therefore,

to improve the objective assessment of the discriminatory power of different subtyping methods, Grundmann *et al.* proposed a method for calculating confidence intervals (CIs) of the Simpson's index (Grundmann *et al.*, 2001). If the CIs overlap, one cannot exclude the hypothesis that both methods have similar discriminatory power (Hunter and Gaston, 1988). To calculate the level of congruency among subtyping techniques, the AW, which indicates the probability that if two isolates share the same classification by method A, they will also share the same classification by method B, corrected by chance agreement, was utilized (Carrico *et al.*, 2006; Severiano *et al.*, 2011). Using the website, the SID and AW for each of the subtyping methods was calculated.

PFGE pattern combinations and MLVA profiles pertaining to all *E. coli* O157:H7 isolates available for MLVA-enhanced retrospective analysis (n= 2,285) were extrapolated from the O157 Ecoli MLVA-Admin database and placed into Excel spreadsheets (Microsoft Corp, 2010). The data was formatted as tab separated values, with the first row containing the column titles, 'PFGE Pattern Combination', and 'MLVA Profile'. The data, and corresponding subtyping method name was copied and pasted into the text area in the Online Tool tab on the website and the SID and AW coefficients were selected for calculation. To effectively compare the coefficients obtained for each of the subtyping methods, the analytical 95 percent confidence interval (CI) was calculated. Additionally, to provide a better understanding of the statistical significance of the data set, and the effect that sampling can have on the results, the jackknife pseudo-values resampling approach was utilized to estimate the 95 percent CI as well as to calculate the p-values for both the SID and AW. Following the same process as detailed above, the SID for the combined PFGE and MLVA laboratory data was calculated and analyzed in

comparison to the SID obtained for PFGE in order to determine the overall discriminatory power of MLVA, when used in combination with PFGE.

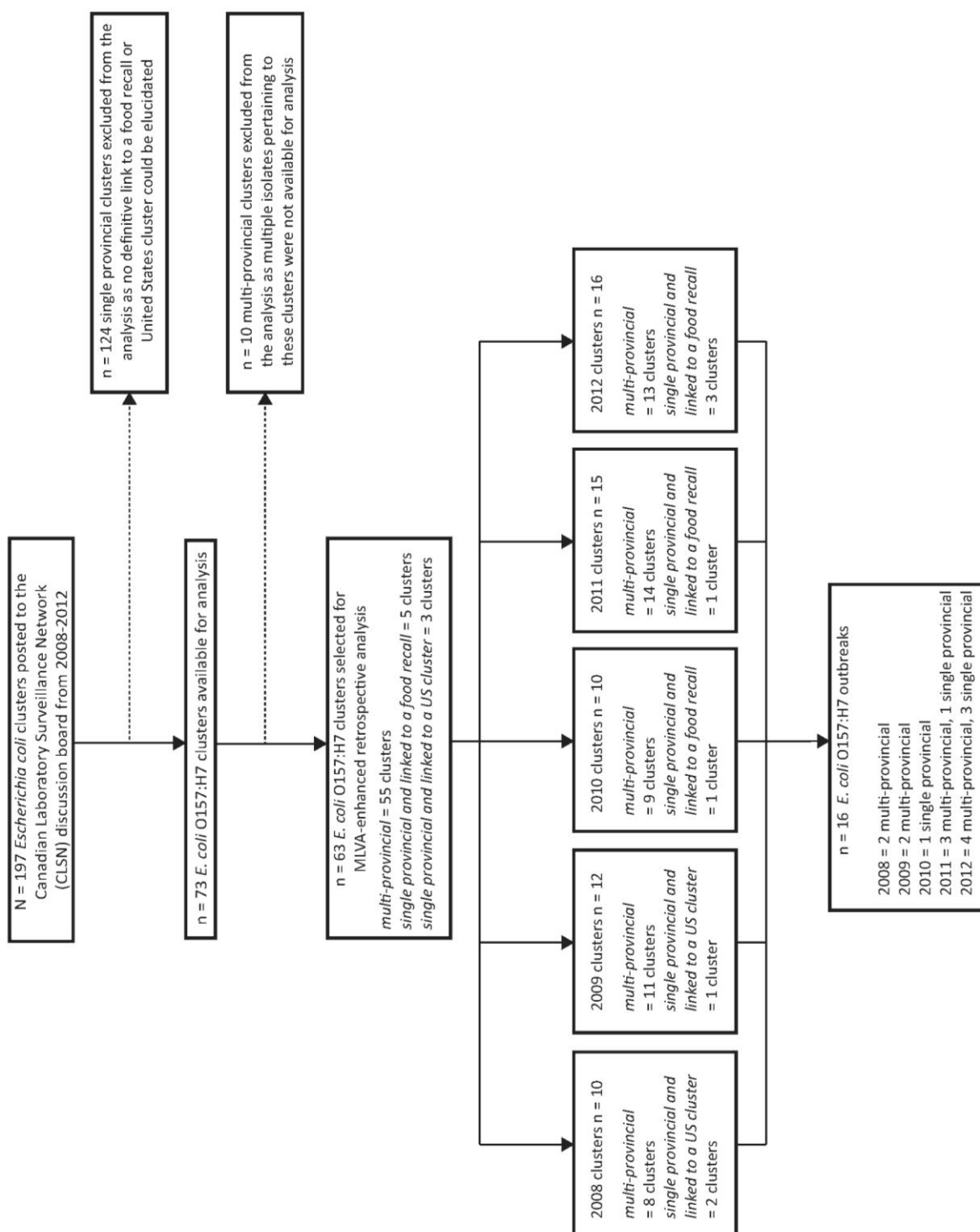
2.4 Selection of *Escherichia coli* O157:H7 Clusters for MLVA-Enhanced Retrospective Analysis

Information posted to the Canadian Laboratory Surveillance Network (CLSN) discussion board on the Canadian Network for Public Health Intelligence (CNPHI) was collected from 2008 to 2012. A total of 197 clusters of *E. coli* O157:H7 were posted to the CLSN discussion board during the five year time span. To clarify, the CLSN discussion board is generally referred to as the PulseNet Canada discussion board and is used as a means of communication between PulseNet Canada and provincial health laboratories. For a cluster to be included in this study, it must have satisfied at least one of the following inclusion criteria: (i) multi-provincial, or (ii) single provincial and linked to food recall, or (iii) single provincial and a PFGE match to concurrent cluster in the United States. A total of 63 *E. coli* O157:H7 clusters satisfied the selection criteria and thus, were selected for retrospective analysis (Figure 2). These 63 clusters comprise approximately 69 percent (n= 1576 isolates) of the total isolates in the study. For the purposes of simplicity, outbreaks were denoted as multi-strain if isolates demonstrated multiple non-similar PFGE patterns but had the same epidemiological link or were isolated from recalled food products.

2.5 MLVA-Enhanced Retrospective Analysis

All isolates identified in original real-time findings for each of the 63 selected *E. coli* O157:H7 clusters were arranged into linelists based on the information acquired from the CLSN discussion board and from epidemiological summaries (where available). To conduct the

Figure 2. Flowchart detailing the inclusion criteria and characteristics of *Escherichia coli* O157:H7 clusters selected for MLVA-enhanced retrospective analysis.



MLVA-enhanced retrospective analyses, comparisons containing isolates identified in original real-time findings were generated in the BioNumerics O157 Ecoli MLVA-Admin database for each of the 63 clusters. For every comparison, the MLVA database was scanned for isolates demonstrating an upload date equivalent to that of the first cluster isolate identified all the way through to 60 days after the last cluster isolate was uploaded. Upload date is typically the only date variable used in PulseNet Canada laboratory analyses; illness onset date is usually not available and collection/isolation date are not consistently provided. The limitation of using the upload date is that it may not be a true reflection of when the cluster began/ended and therefore, some cases may have been missed due to potential delays in uploading PFGE subtyping information to the PulseNet Canada National database. Using the BioNumerics software, similarity among MLVA profiles was measured using the unweighted pair group method with arithmetic mean (UPGMA) and minimum spanning trees were constructed using the advanced cluster analysis tool for categorical data with single and double locus variance priority rules. Of note, relatedness of isolates was solely based on laboratory evidence. In accordance with the current PulseNet Canada interpretation guidelines for relatedness of MLVA profiles, isolates demonstrating profiles that differed by no more than 1 repeat at 3 loci, or 3 repeats at one locus and showed no variation at VNTR 34 in comparison to the main MLVA cluster profile were considered to be retrospectively related (Nadon and Rumore, 2013; Rumore *et al.*, 2013). The main cluster profile was defined as the most common MLVA profile observed among the clinical isolates within each of the retrospective clusters. Variability in MLVA profiles was determined by calculating the number of repeats occurring at each VNTR locus in comparison to the main cluster profile.

2.6 Evaluation of the Sensitivity and Specificity of MLVA as a Supplemental Test for Outbreak Investigations

Using the methods described by Galen and Gambino, the sensitivity and specificity of MLVA as a supplemental subtyping method for outbreak detection and response was calculated using only the clinical isolates identified for each of the 63 *E. coli* O157:H7 clusters (Galen and Gambino, 1977). For each cluster, true negatives, representing valid non-outbreak cases, were defined as those isolates demonstrating upload dates that fell within the specified range but did not group by either PFGE or by PFGE and MLVA collectively. As well, the positive predictive value, negative predictive value and accuracy were also calculated to further elucidate the overall validity of the conjunctional use of PFGE and MLVA in comparison to the gold standard (Galen and Gambino, 1977; Lalkhen and McCluskey, 2008). Clusters for which there were known major outbreak investigations with epidemiological links were selected for detailed analysis as a proxy for estimating epidemiological congruence. Using only the isolates contained within those clusters (n=16 clusters), the statistical values were then recalculated and compared to the sensitivity and specificity obtained for all 63 retrospective clusters.

Chapter 3: Results

3.1 Coverage of *Escherichia coli* O157:H7 Isolates Obtained for Analysis

Based on the data obtained from the National Enteric Surveillance Program (NESP) annual summary, a total of 2,560 *E. coli* O157 isolations were reported to NESP from 2008 to 2012 (Public Health Agency of Canada, 2012). In comparison to the national total, 2,285 isolates were available for retrospective analysis, which provided coverage of approximately 89 percent of all *E. coli* O157 isolates reported in Canada over the 5 year time span.

3.2 Evaluation of Simpson's Index of Diversity and the Adjusted Wallace Coefficient

A total of 2,285 isolates uploaded to the National PulseNet Canada database from 2008 to 2012 (inclusive) were used to calculate both the SID and AW in order to obtain an assessment of the discriminatory power and level of congruency among the different subtyping methods examined in this study. Table 5 summarizes the SID coefficient obtained for PFGE, MLVA and a combination of both of these techniques. The number of partitions corresponds to the number of unique entities (patterns/profiles) identified for each typing method. Based on the data, MLVA demonstrated a higher SID (0.998) than that of PFGE (0.993) and the difference among the coefficients was statistically significant ($p < 0.001$) indicating MLVA was more discriminatory than PFGE. As well, there was a clear statistical difference ($p < 0.001$) between the SID calculated for the conjunctive use of PFGE and MLVA (0.998) and the SID calculated for PFGE alone (0.993) (Table 5). To quantify the degree of congruence, the AW coefficient, which avoids the overestimation of unidirectional concordance between subtyping methods, was calculated. If two isolates are in the same cluster by PFGE, they have approximately a 23 percent chance of having the same MLVA profile; while conversely, if two isolates are in the

Table 5. Comparative analysis of the Simpson's Index of Diversity for two subtyping methods, Pulsed-field Gel Electrophoresis and Multiple-locus Variable Number Tandem Repeat Analysis, utilized to characterize *Escherichia coli* O157:H7 isolates for laboratory surveillance and outbreaks in Canada.

Number of Isolates	Upload Year	Subtyping Method	Number of Partitions	Simpson's Index of Diversity	CI (95%)	Jackknife pseudo-values CI (95%)	p-value
2285	2008-2012	PFGE	1134	0.993	(0.992-0.994)	(0.992-0.994)	<0.001*
		MILVA	1222	0.998	(0.997-0.998)	(0.997-0.998)	
2285	2008-2012	PFGE	1134	0.993	(0.992-0.994)	(0.992-0.994)	<0.001*
		PFGE + MILVA	1541	0.998	(0.998-0.999)	(0.998-0.999)	

Table 6. Analysis of the Adjusted Wallace coefficient for two subtyping methods, Pulsed-field Gel Electrophoresis and Multiple-locus Variable Number Tandem Repeat Analysis, used to characterize *Escherichia coli* O157:H7 isolates for laboratory surveillance and outbreaks in Canada.

Number of Isolates	Upload Year	Clustering	Adjusted Wallace	CI (95%)	Jackknife pseudo-values CI (95%)	p-value
2285	2008-2012	PFGE → MLVA	0.227	(0.207-0.247)	(0.175-0.276)	<0.001*
		MLVA → PFGE	0.696	(0.653-0.739)	(0.641-0.752)	

same cluster by MLVA, they have about a 70 percent chance of having the same PFGE pattern combination (Table 6). The statistical significance of the results ($p < 0.001$) could potentially suggest that PFGE and MLVA are not likely to be congruent methods, and that performing MLVA along with PFGE is superior compared to PFGE alone.

3.3 MLVA-Enhanced Retrospective Analyses of *E. coli* O157:H7 Clusters

3.3.1 2008 Clusters

A minimum spanning tree (MST) of the MLVA profiles associated with 10 clusters, including 2 major outbreak investigations, of *E. coli* O157:H7 that occurred in 2008 was generated (Figure 3). The MST illustrates the relationship among isolates based on the number of VNTR loci that differ between 2 MLVA profiles. The detailed findings are summarized in Table 7. For clusters 1, 2, and 5, no PFGE or MLVA variability was observed among the isolates pertaining to each of the clusters. Interestingly, cluster 1 and 2, which occurred during the same time period, demonstrated identical MLVA profiles, in addition to highly similar PFGE pattern combinations (97 percent similarity), suggesting these isolates may have represented a single outbreak. However, epidemiological evidence was not available to further evaluate a potential link between these 2 clusters. Cluster 3 isolates demonstrated a PFGE pattern that was indistinguishable from the pattern combination associated with a concurrent United States cluster. Although no MLVA variability was observed among the isolates identified in original real-time findings for cluster 3, an additional 4 isolates demonstrating variant, yet highly similar PFGE patterns (≥ 89 percent similar), were identified via this MLVA-enhanced retrospective analysis. These isolates demonstrated comparable profiles to that of the main

Figure 3. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 142 *Escherichia coli* O157:H7 isolates (clinical and food) associated with 10 retrospective clusters that occurred in 2008 as reported by PulseNet Canada. Each node in the MST corresponds to a unique MLVA profile and is color coded based on the pulsed-field gel electrophoresis (PFGE) pattern. Node sizes are related to the number of isolates and the number indicated in the middle of each node corresponds to the arbitrarily assigned cluster code. The solid lines infer relatedness by stating the number of differing loci. Retrospectively related isolates not previously identified in original real-time findings are highlighted in yellow, and unrelated isolates demonstrating highly variable MLVA profiles are indicated by a dotted line outlining the node.

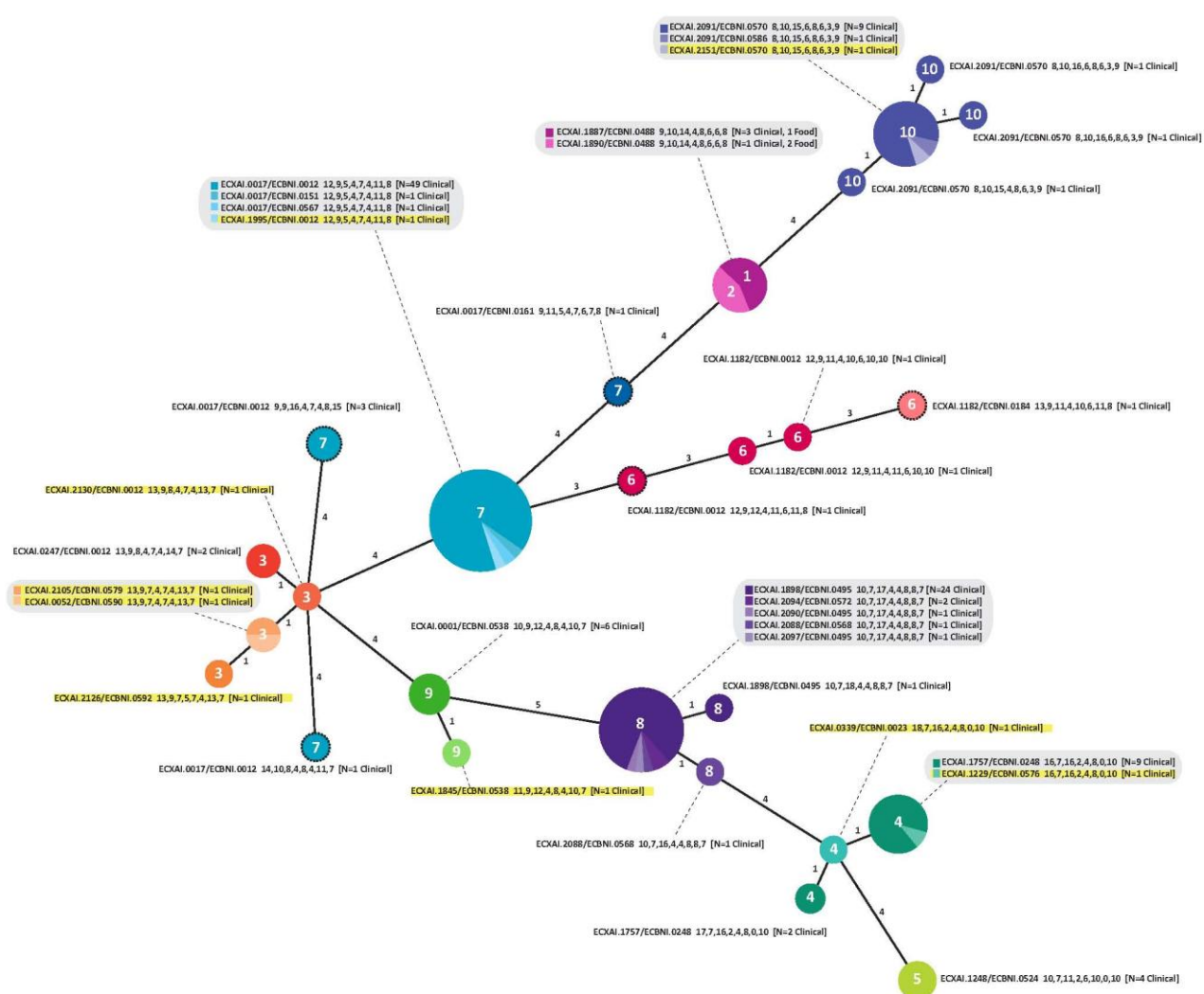


Table 7. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 10 *Escherichia coli* O157:H7 clusters that occurred in 2008 as reported by PulseNet Canada. Differences in repeat copy number in comparison to the main cluster profile are indicated in red.

Cluster (source)	Multi-Provincial?	Isolates identified in original real-time findings		Isolates identified via MLVA-enhanced retrospective analysis			
		n	PFGE Pattern Combination	n	PFGE Pattern Combination	n	MLVA Profile Variability in MLVA
1	Yes	4	ECXAI.1887/ECBNI.0488	4	ECXAI.1887/ECBNI.0488	4	9,10,14,4,8,6,6,8 No Variability
2	Yes	3	ECXAI.1890/ECBNI.0488	3	ECXAI.1890/ECBNI.0488	3	9,10,14,4,8,6,6,8 No Variability
3*	No	2	ECXAI.0247/ECBNI.0012	2	ECXAI.0247 / ECBNI.0012	2	13,9,8,4,7,4,14,7
				1	ECXAI.2105 / ECBNI.0579†	1	13,9,7,4,7,4,13,7 1 repeat at 2 loci
				1	ECXAI.0052 / ECBNI.0590†	1	13,9,7,4,7,4,13,7 1 repeat at 2 loci
				1	ECXAI.2126 / ECBNI.0592†	1	13,9,7,5,7,4,13,7 1 repeat at 3 loci
				1	ECXAI.2130 / ECBNI.0012†	1	13,9,8,4,7,4,13,7 1 repeat at 1 locus
4*	No	1	ECXAI.1757/ECBNI.0248	11	ECXAI.1757/ECBNI.0248	9	16,7,16,2,4,8,0,10
		1				2	17,7,16,2,4,8,0,10 1 repeat at 1 locus
				1	ECXAI.1229/ECBNI.0576†	1	16,7,16,2,4,8,0,10
				1	ECXAI.0339/ECBNI.0023†	1	18,7,16,2,4,8,0,10 1 repeat at 1 locus
5	Yes	4	ECXAI.1248 / ECBNI.0524	4	ECXAI.1248 / ECBNI.0524	4	10,7,11,2,6,10,0,10 No Variability
6	Yes	3	ECXAI.1182/ECBNI.0012	3	ECXAI.1182/ECBNI.0012	1	12,9,11,4,11,6,10,10
		1	ECXAI.1182/ECBNI.0184			1	12,9,11,4,10,6,10,10 1 repeat at 1 locus
				1	ECXAI.1182/ECBNI.0184	1	12,9,12,4,11,6,11,8 1-2 repeats at 3 loci*
						1	13,9,11,4,10,6,11,8 1-2 repeats at 3 loci*
7* (fast food restaurant)	Yes	5	ECXAI.0017 / ECBNI.0012	53*	ECXAI.0017 / ECBNI.0012	49	12,9,5,4,7,4,11,8
		6	ECXAI.0017 / ECBNI.0567			3	9,9,16,4,7,4,8,15 3-11 repeats at 4 loci*
		1	ECXAI.0017 / ECBNI.0161			1	14,10,8,4,8,4,11,7 1-3 repeats at 5 loci*
		1	ECXAI.0017 / ECBNI.0151	1	ECXAI.0017 / ECBNI.0567	1	12,9,5,4,7,4,11,8
				1	ECXAI.0017 / ECBNI.0161	1	9,11,5,4,7,6,7,8 2-5 repeats at 4 loci*
				1	ECXAI.0017 / ECBNI.0151	1	12,9,5,4,7,4,11,8
				1	ECXAI.1995 / ECBNI.0012†	1	12,9,5,4,7,4,11,8
8* (lettuce)	Yes	2	ECXAI.1898 / ECBNI.0495	25*	ECXAI.1898 / ECBNI.0495	24	10,7,17,4,4,8,8,7
		6	ECXAI.2094 / ECBNI.0572			1	10,7,18,4,4,8,8,7 1 repeat at 1 locus
		2	ECXAI.2088 / ECBNI.0568	2	ECXAI.2094 / ECBNI.0572	2	10,7,17,4,4,8,8,7
		2	ECXAI.2090 / ECBNI.0495	2	ECXAI.2088 / ECBNI.0568	1	10,7,17,4,4,8,8,7
		1	ECXAI.2097/ECBNI.0495			1	10,7,16,4,4,8,8,7 1 repeat at 1 locus
		1		1	ECXAI.2090 / ECBNI.0495	1	10,7,17,4,4,8,8,7
				1	ECXAI.2097 / ECBNI.0495	1	10,7,17,4,4,8,8,7
9	Yes	7	ECXAI.0001/ECBNI.0538	6*	ECXAI.0001/ECBNI.0538	6	10,9,12,4,8,4,10,7
					ECXAI.1845/ECBNI.0570†	1	11,9,12,4,8,4,10,7 1 repeat at 1 locus
10	Yes	1	ECXAI.2091/ECBNI.0570	12	ECXAI.2091/ECBNI.0570	9	8,10,15,6,8,6,3,9
		2	ECXAI.2091/ECBNI.0586			1	8,10,15,6,8,7,3,9 1 repeat at 1 locus
		1				1	8,10,15,4,8,6,3,9 2 repeats at 1 locus
						1	8,10,16,6,8,6,3,9 1 repeat at 1 locus
				1	ECXAI.2091/ECBNI.0586	1	8,10,15,6,8,6,3,9
				1	ECXAI.2151/ECBNI.0570†	1	8,10,15,6,8,6,3,9

▪ PFGE match to a concurrent United States cluster.

▫ Epidemiologically linked to a common source and/or food recall.

• Not all isolates were available for further study.

† Retrospectively related isolate.

* Not likely to be related due to the presence of a highly variable MLVA profile.

cluster MLVA profile, differing by a minimum of 1 repeat at 1 locus and a maximum of 1 repeat at 3 loci, and therefore, were considered to be retrospectively related to the cluster. Cluster 4 isolates demonstrated identical PFGE pattern combinations and very little MLVA variability (maximum of 1 repeat at 1 locus).

Following MLVA-enhanced retrospective analysis, 2 isolates with variant PFGE patterns that were 94-98 percent similar to the main cluster pattern also demonstrated highly similar MLVA profiles (maximum of 1 repeat at 1 locus), and therefore, were suspected to be cluster related. Additionally, this cluster was a PFGE match to a concurrent cluster in the United States. Among cluster 6 isolates, 2 highly similar PFGE patterns (97 percent similar *BlnI* patterns) demonstrating 4 different MLVA profiles were observed. For 2 isolates, little MLVA variability was observed (maximum of 1 repeat at 1 locus); while conversely, the remaining isolates differed by 1 to 2 repeats at 3 loci, indicating they may not have been related to the cluster. Based on the original real-time findings, 4 highly similar PFGE pattern combinations were included in cluster 7. Epidemiological evidence indicated a majority of the clinical cases were linked to a fast food restaurant. However, only 56 out of the 59 isolates identified in the original real-time findings were available for this study. These isolates demonstrated indistinguishable *XbaI* patterns and highly similar *BlnI* patterns (maximum of 1 band difference) and MLVA variability was observed for only 5 isolates (1 to 11 repeats at 4 to 5 loci). As a result of the highly variable nature of the MLVA profiles, these isolates were unlikely to be outbreak related.

In addition to discriminating among potentially unrelated outbreak isolates, MLVA-enhanced analysis also identified an extra isolate with a highly similar PFGE pattern to the main

cluster pattern (98 percent similar) and an identical MLVA profile to that of the main outbreak profile, indicating it may have been related. Only 31 out of the 32 isolates were available for analysis of cluster 8, which was linked to the consumption of lettuce. Among the 5 variant PFGE patterns ($\geq 92\%$ similar), no MLVA variability was observed, with the exception of 2 profiles which differed by 1 repeat at 1 locus. This low level of MLVA variability suggests these isolates were likely to be related. Of the 6 isolates that could be recovered for MLVA-enhanced retrospective analysis for cluster 9, they all appeared to be highly related as no PFGE or MLVA variability was observed. An additional isolate demonstrating a highly comparable PFGE pattern (95 percent similar) and MLVA profile (maximum of 1 repeat at 1 locus) was identified. Based on the high level of similarity to main cluster MLVA profile, this particular isolate was considered to be retrospectively related to the cluster. Very little MLVA variability was observed among the 2 highly similar PFGE pattern combinations associated with cluster 10 (1 band difference in *BlnI* patterns). Profiles differed by a minimum of 1 repeat at 1 locus and a maximum of 2 repeats at 1 locus indicating a high level of relatedness among isolates. One retrospectively related isolate demonstrating a 98 percent similar PFGE pattern and an MLVA profile identical to the main cluster profile was identified via MLVA-enhanced analysis.

3.3.2 2009 Clusters

Twelve clusters, including 3 major outbreak investigations from 2009 were selected for analysis (Figure 4). The comprehensive results pertaining to the MLVA-enhanced retrospective analysis are summarized in Table 8. According to the original real-time findings, 2 highly similar (*BlnI* differed by 1 band) and 1 variant PFGE pattern ($> 92\%$ similar) were included in cluster 11, which was linked to the consumption of beef products.

Figure 4. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 94 *Escherichia coli* O157:H7 isolates (clinical and food) associated with 12 retrospective clusters that occurred in 2009 as reported by PulseNet Canada. Each node in the MST corresponds to a unique MLVA profile and is color coded based on the pulsed-field gel electrophoresis (PFGE) pattern. Node sizes are related to the number of isolates and the number indicated in the middle of each node corresponds to the arbitrarily assigned cluster code. The solid lines infer relatedness by stating the number of differing loci. Retrospectively related isolates not previously identified in original real-time findings are highlighted in yellow, and unrelated isolates demonstrating highly variable MLVA profiles are indicated by a dotted line outlining the node.

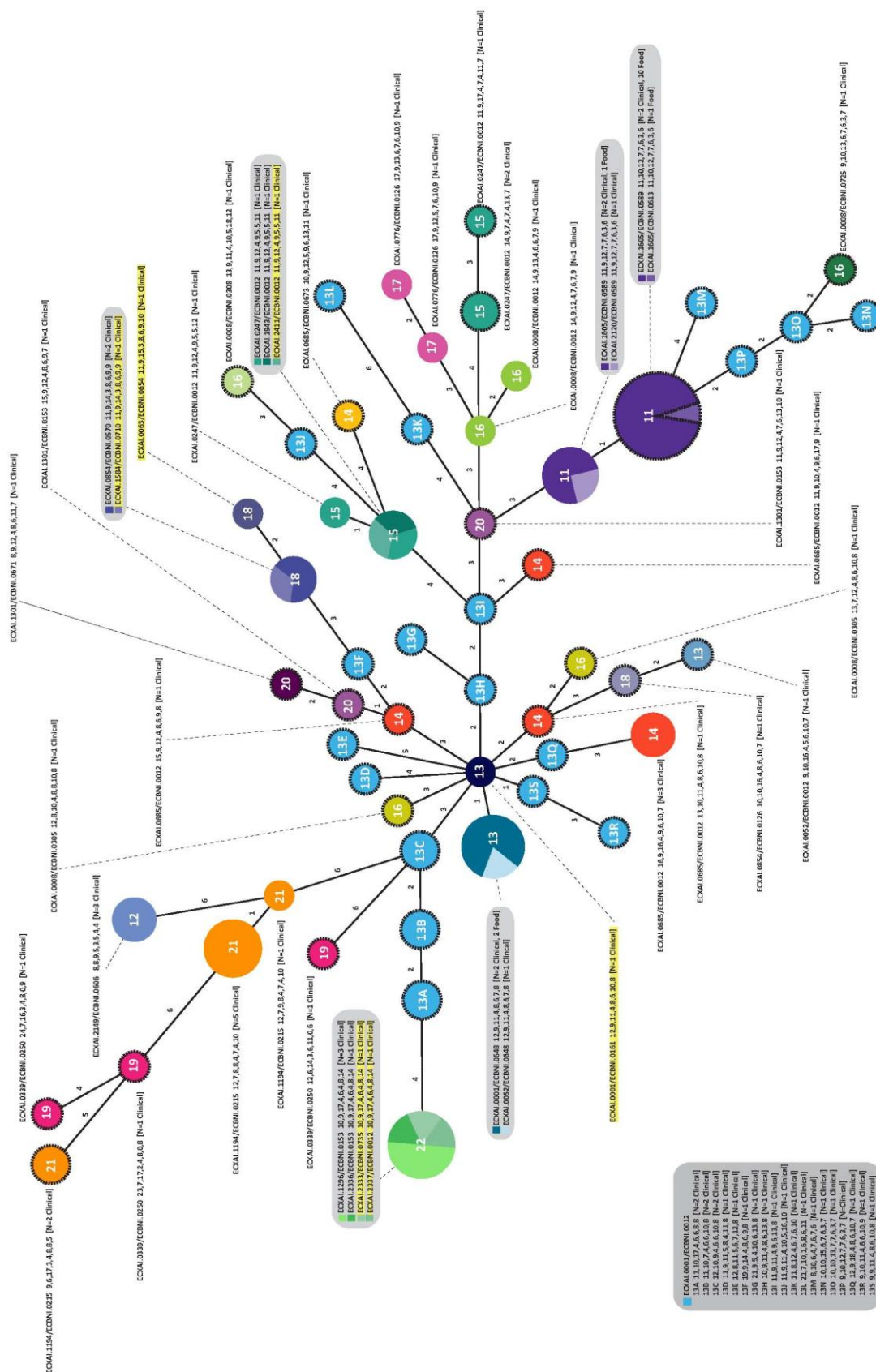


Table 8. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 12 *Escherichia coli* O157:H7 clusters that occurred in 2009 as reported by PulseNet Canada. Differences in repeat copy number in comparison to the main cluster profile are indicated in red.

Cluster (source)	Multi-Provincial	Isolates identified in original real-time findings		Isolates identified via MLVA-enhanced retrospective analysis				
		n	PFGE Pattern Combination	n	PFGE Pattern Combination	n	MLVA Profile	Variability in MLVA
11 ^a (beef products)	Yes	15	ECXAI.1605/ECBNI.0589	15	ECXAI.1605/ECBNI.0589	12	11,10,12,7,7,6,3,6	1 repeat at 1 locus**
		1	ECXAI.1605/ECBNI.0613			3	11,9,12,7,7,6,3,6	
		1	ECXAI.2120/ECBNI.0589	1	ECXAI.1605/ECBNI.0613	1	11,10,12,7,7,6,3,6	1 repeat at 1 locus**
				1	ECXAI.2120/ECBNI.0589	1	11,9,12,7,7,6,3,6	
12*	No	3	ECXAI.2149/ECBNI.0606	3	ECXAI.2149/ECBNI.0606	3	8,8,9,5,3,5,4,4	No Variability
13	Yes	24	ECXAI.0001/ECBNI.0012	22*	ECXAI.0001/ECBNI.0012	19	Multiple MLVA Profiles	1-9 repeats at 2-8 loci*
		4	ECXAI.0001/ECBNI.0648	4	ECXAI.0001/ECBNI.0648	4	12,9,11,4,8,6,7,8	
		1	ECXAI.0052/ECBNI.0648	1	ECXAI.0052/ECBNI.0648	1	12,9,11,4,8,6,7,8	
		1	ECXAI.0052/ECBNI.0012	1	ECXAI.0052/ECBNI.0012	1	9,10,16,4,5,6,10,7	1-5 repeats at 6 loci*
				1	ECXAI.0001/ECBNI.0161+	1	12,9,11,4,8,6,10,8	3 repeats at 1 locus
14	Yes	8	ECXAI.0685/ECBNI.0012	6*	ECXAI.0685/ECBNI.0012	3	16,9,16,4,9,6,10,7	
		1	ECXAI.0685/ECNI.0673			1	11,9,10,4,9,6,17,9	2-7 repeats at 4 loci*
						1	15,9,12,4,8,6,9,8	1-4 repeats at 5 loci*
						1	13,10,11,4,8,6,10,8	1-5 repeats at 5 loci*
				1	ECXAI.0685/ECNI.0673	1	10,9,12,5,9,6,13,11	1-6 repeats at 5 loci*
15	Yes	5	ECXAI.0247/ECBNI.0012	5	ECXAI.0247/ECBNI.0012	2	14,9,7,4,7,4,13,7	1-8 repeats at 6 loci*
		1	ECXAI.1943/ECBNI.0012			1	11,9,12,4,9,5,5,11	
						1	11,9,12,4,9,5,5,12	1 repeat at 1 locus
						1	11,9,17,4,7,4,11,7	1-6 repeats at 5 loci*
				1	ECXAI.1943/ECBNI.0012	1	11,9,12,4,9,5,5,11	
				1	ECXAI.2411/ECBNI.0012+	1	11,9,12,4,9,5,5,11	
16	Yes	2	ECXAI.0008/ECBNI.0012	2	ECXAI.0008/ECBNI.0012	1	14,9,13,4,6,6,7,9	1 repeat at 2 loci
		2	ECXAI.0008/ECBNI.0308			1	14,9,12,4,7,6,7,9	
		2	ECXAI.0008/ECBNI.0305	2	ECXAI.0008/ECBNI.0305	1	12,8,10,4,8,8,10,8	1-3 repeats at 7 loci*
		1	ECXAI.0008/ECBNI.0725			1	13,7,12,4,8,6,10,8	1-3 repeats at 5 loci*
				1	ECXAI.0008/ECBNI.0308	1	13,9,11,4,10,5,18,12	1-11 repeats at 6 loci*
				1	ECXAI.0008/ECBNI.0725	1	9,10,13,6,7,6,3,7	1-5 repeats at 6 loci*
17	Yes	2	ECXAI.0776/ECBNI.0126	2	ECXAI.0776/ECBNI.0126	1	17,9,13,6,7,6,10,9	
						1	17,9,12,5,7,6,10,9	1 repeat at 2 loci
18 ^a (beef products)	Yes	2	ECXAI.0854/ECBNI.0570	2	ECXAI.0854/ECBNI.0570	2	11,9,14,3,8,6,9,9	
		1	ECXAI.0854/ECBNI.0126	1	ECXAI.0854/ECBNI.0126	1	10,10,16,4,8,6,10,7	1-2 repeats at 6 loci*
						1	11,9,14,3,8,6,9,9	
						1	11,9,15,3,8,6,9,10	1 repeat at 2 loci
19	Yes	3	ECXAI.0339/ECBNI.0250	3	ECXAI.0339/ECBNI.0250	1	24,7,16,3,4,8,0,9	
						1	23,7,17,2,4,8,0,8	1 repeat at 4 loci*
						1	12,6,14,3,6,11,0,6	1-12 repeats at 6 loci*
20	Yes	2	ECXAI.1301/ECBNI.0153	2	ECXAI.1301/ECBNI.0153	1	15,9,12,4,8,6,9,7	
		1	ECXAI.1301/ECBNI.0671			1	11,9,12,4,7,6,13,10	1-4 repeats at 4 loci*
				1	ECXAI.1301/ECBNI.0671	1	8,9,12,4,8,6,11,7	2-7 repeats at 2 loci*
21	Yes	8	ECXAI.1194/ECBNI.0215	8	ECXAI.1194/ECBNI.0215	5	12,7,8,8,4,7,4,10	
						1	12,7,9,8,4,7,4,10	1 repeat at 1 locus
						2	9,6,17,3,4,8,8,5	1-9 repeats at 7 loci*
22	Yes	3	ECXAI.1296/ECBNI.0153	3	ECXAI.1296/ECBNI.0153	3	10,9,17,4,6,4,8,14	No Variability
		1	ECXAI.2336/ECBNI.0153	1	ECXAI.2336/ECBNI.0153	1	10,9,17,4,6,4,8,14	
				1	ECXAI.2333/ECBNI.0735+	1	10,9,17,4,6,4,8,14	
				1	ECXAI.2337/ECBNI.0012+	1	10,9,17,4,6,4,8,14	

• PFGE match to a concurrent United States cluster.

◻ Epidemiologically linked to a common source and/or food recall.

• Not all isolates were available for further study.

† Retrospectively related isolate.

* Not likely to be related due to the presence of a highly variable MLVA profile.

** In retrospect, with variability at VNTR 34, it is possible this was not related to the cluster.

Among these 3 PFGE pattern groupings, 2 variable MLVA profiles were observed. One MLVA profile occurred among a majority of the clinical isolates (n=3), and therefore was considered to be the main outbreak profile. Even though the remaining MLVA profile differed by only 1 repeat at 1 locus, the variability occurred at VNTR 34 suggesting these isolates were not related to the main outbreak profile. With respect to cluster 12, which was a 2 enzyme match to a concurrent United States cluster, no PFGE or MLVA variability was observed, and therefore, these isolates were considered related. For cluster 13, a total of 30 isolates demonstrating 4 variant, yet highly similar PFGE patterns were identified during the original real-time findings; however, only 28 of these isolates were available for retrospective analysis. Among 2 of the PFGE pattern groupings, which differed by only 1 band with respect to the *BlnI* pattern, no MLVA variability was observed, and therefore, were considered likely to be related. Although the remaining isolates demonstrated indistinguishable PFGE pattern combinations, 19 highly variable MLVA profiles differing by 1-9 repeats at 2 to 8 loci from the main profile were observed. Therefore, these isolates were not likely to be related to the cluster, but rather, were likely representative of temporally and geographically associated sporadic cases, which would not be unexpected during investigations. MLVA-enhanced retrospective analysis identified an additional isolate demonstrating a highly similar PFGE pattern (> 97 percent similar) and MLVA profile (differed by 3 repeats at 1 locus) that may have potentially been cluster related.

With respect to cluster 14, only 6 out of the 9 isolates demonstrating 2 highly similar PFGE patterns could be obtained for MLVA-enhanced retrospective analysis. Three isolates demonstrating the main PFGE pattern had indistinguishable MLVA profiles, and thus, were

considered to be related. The remaining 6 isolates, of which one belonged to a highly similar PFGE variant (95 percent similar), were not likely to be cluster related as they were highly variable in comparison to the main cluster profile (differed by 1 to 7 repeats at 4 to 5 loci). For cluster 15, 2 variant PFGE patterns (> 97 percent similar) were observed. Three isolates demonstrated highly comparable MLVA profiles, with one isolate varying by a maximum of 1 repeat at 1 locus. This low level of MLVA variability suggests relatedness among these isolates. In contrast, the other 6 MLVA profiles were highly variable, differing by 1 to 8 repeats at 5 to 6 loci and therefore, were not assumed to be cluster related. MLVA-enhanced analysis identified an additional isolate with a highly comparable PFGE profile (94 percent similar) and an MLVA profile identical to that of the main cluster profile and therefore was considered to be retrospectively related. Based on the original real-time findings, 4 similar PFGE pattern combinations ($\geq$ 85 percent similar) were associated with cluster 16. However, only those isolates within PFGE pattern grouping ECXAI.0008/ECBNI.0012 were found to demonstrate highly similar MLVA profiles (maximum of 1 repeat at 2 loci), and therefore were most likely to be related. Due to the high level of MLVA variability, the remaining profiles, which differed by 1 to 11 repeats at 5 to 7 loci, did not appear cluster related.

Isolates included in cluster 17 are assumed to be related based on identical PFGE patterns and highly similar MLVA profiles (maximum of 1 repeat at 2 loci). In regards to cluster 18, which was defined as an outbreak based on the consumption of a common food source (ground beef), no MLVA variability was observed among the main PFGE pattern grouping ECXAI.0854/ECBNI.0126; while conversely, a high level of variability (1 to 2 repeats at 6 loci) was noted for the PFGE variant ECXAI.0854/ECBNI.0126 (89 percent similar), suggesting it may

not have been outbreak related. Two additional isolates demonstrating highly comparable PFGE patterns (> 92 percent similar) and MLVA profiles differing by a maximum of 1 repeat at 2 loci from the main outbreak profile were identified via MLVA-enhanced analysis. Based on the high level of similarity, these isolates may be potentially related to the cluster. A high level of MLVA variability was observed for both cluster 19 (1 to 12 repeats at 4 to 6 loci) and for alike PFGE patterns (> 97 percent similar) in cluster 20 (1 to 7 repeats at 2 to 4 loci) and therefore, these may not have been representative of true clusters. Although no PFGE variability was observed, only 6 of the isolates associated with cluster 21 demonstrated comparable MLVA profiles (maximum of 1 repeat at 1 locus). The remaining 2 isolates did not appear to be cluster related since the MLVA profiles varied by 1 to 9 repeats at 7 loci from the main cluster profile. Despite the occurrence of 2 variant PFGE patterns (> 97 percent similar) in cluster 22, there was no variability in the MLVA profiles. Following MLVA-enhanced analysis, 2 additional isolates with high comparable PFGE patterns (> 93 percent similar) and identical MLVA profiles were identified. In retrospect, these isolates were likely to be cluster related due to the high level of similarity to the main cluster MLVA profile.

3.3.3 2010 Clusters

A total of 9 clusters and 1 outbreak were selected for MLVA-enhanced retrospective analysis of 2010 *E. coli* O157:H7 isolates and are presented in Figure 5 and Table 9. With respect to clusters 23, 24 and cluster 29, which was linked to the consumption of frozen hamburgers, no PFGE or MLVA variability was observed within each grouping (although not all isolates were available for this study). For cluster 25, isolates were indistinguishable by PFGE, but however, only 2 of these demonstrated identical profiles and thus, appeared to be related.

Figure 5. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 47 *Escherichia coli* O157:H7 isolates (clinical and food) associated with 10 retrospective clusters that occurred in 2010 as reported by PulseNet Canada. Each node in the MST corresponds to a unique MLVA profile and is color coded based on the pulsed-field gel electrophoresis (PFGE) pattern. Node sizes are related to the number of isolates and the number indicated in the middle of each node corresponds to the arbitrarily assigned cluster code. The solid lines infer relatedness by stating the number of differing loci. Retrospectively related isolates not previously identified in original real-time findings are highlighted in yellow, and unrelated isolates demonstrating highly variable MLVA profiles are indicated by a dotted line outlining the node.

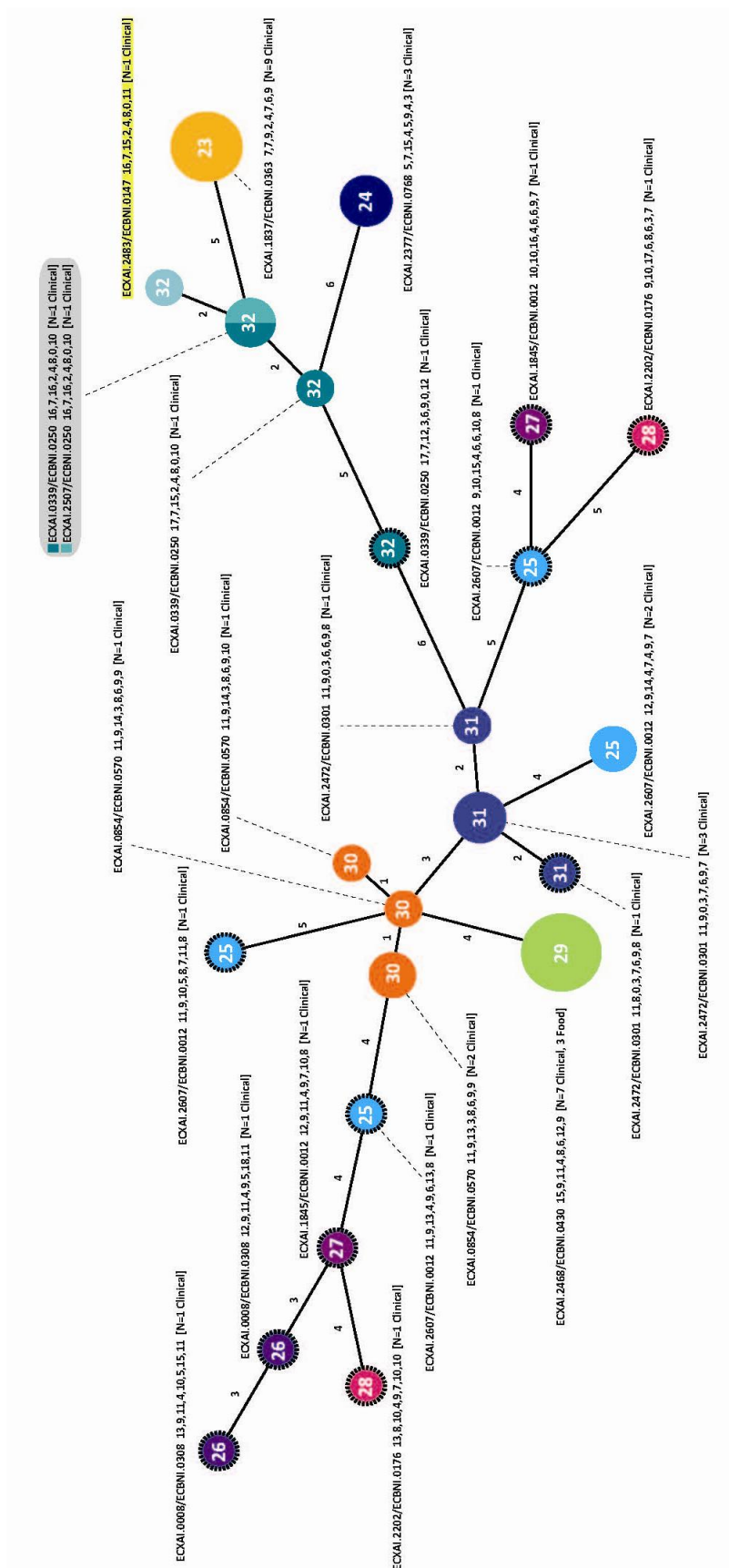


Table 9. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 10 *Escherichia coli* O157:H7 clusters that occurred in 2010 as reported by PulseNet Canada. Differences in repeat copy number in comparison to the main cluster profile are indicated in red.

Cluster (source)	Multi-Provincial	Isolates identified in original real-time findings		Isolates identified via MLVA-enhanced retrospective analysis						
		n	PFGE Pattern Combination	n	PFGE Pattern Combination	n	MLVA Profile	Variability in MLVA		
23	Yes	11	ECXAI.1837/ECBNI.0363	9*	ECXAI.1837/ECBNI.0363	9	7,7,9,2,4,7,6,9	No Variability		
24	Yes	3	ECXAI.2377/ECBNI.0768	3	ECXAI.2377/ECBNI.0768	3	5,7,15,4,5,9,4,3	No Variability		
25		5	ECXAI.2607/ECBNI.0012	5	ECXAI.2607/ECBNI.0012	2	12,9,14,4,7,4,9,7	1-4 repeats at 7 loci* 1-4 repeats at 6 loci* 1-3 repeats at 7 loci*		
						1	11,9,10,5,8,7,11,8			
						1	11,9,13,4,9,6,13,8			
						1	9,10,15,4,6,6,10,8			
26	Yes	2	ECXAI.0008/ECBNI.0308	2	ECXAI.0008/ECBNI.0308	1	13,9,11,4,10,5,15,11	1-3 repeats at 3 loci*		
						1	12,9,11,4,9,5,18,11			
27	Yes	2	ECXAI.1845/ECBNI.0012	2	ECXAI.1845/ECBNI.0012	1	12,9,11,4,9,7,10,8	1-5 repeats at 7 loci*		
						1	10,10,16,4,6,6,9,7			
28	Yes	2	ECXAI.2202/ECBNI.0176	2	ECXAI.2202/ECBNI.0176	1	9,10,17,6,8,6,3,7	1-7 repeats at 8 loci*		
						1	13,8,10,4,9,7,10,10			
29 ^a (beef products)	No	10	ECXAI.2468/ECBNI.0430	10	ECXAI.2468/ECBNI.0430	10	15,9,11,4,8,6,12,9	No Variability		
30	Yes	4	ECXAI.0854/ECBNI.0570	4	ECXAI.0854/ECBNI.0570	2	11,9,13,3,8,6,9,9	1 repeat at 1 locus 1 repeat at 2 loci		
						1	11,9,14,3,8,6,9,9			
						1	11,9,14,3,8,6,9,10			
31	Yes	5	ECXAI.2472/ECBNI.0301	5	ECXAI.2472/ECBNI.0301	3	11,9,0,3,7,6,9,7	1 repeat at 2 loci 1 repeat at 2 loci**		
						1	11,9,0,3,6,6,9,8			
						1	11,8,0,3,7,6,9,8			
						1				
32	Yes	5	ECXAI.0339/ECBNI.0250	3*	ECXAI.0339/ECBNI.0250	1	16,7,16,2,4,8,0,10	1 repeat at 2 loci 1-4 repeats at 6 loci*		
		1	ECXAI.2507/ECBNI.0250			1	17,7,15,2,4,8,0,10			
						1	17,7,12,3,6,9,0,12			
						1	16,7,16,2,4,8,0,10			
						1	16,7,15,2,4,8,0,11			
1	ECXAI.2483/ECBNI.0147+	1	16,7,16,2,4,8,0,10	1 repeat at 2 loci						

^a Epidemiologically linked to a common source and/or food recall.

• Not all isolates were available for further study.

† Retrospectively related isolate.

* Not likely to be related due to the presence of a highly variable MLVA profile.

** In retrospect, with variability at VNTR 34, it is possible this was not related to the cluster.

The remaining 3 isolates had highly variable profiles differing by 1 to 4 repeats at 6 to 7 loci indicating they were not likely to be cluster related. Although no PFGE variability was observed, the high level of MLVA variability demonstrated in cluster 26 (1 to 3 repeats at 3 loci), cluster 27 (1 to 5 repeats at 7 loci) and cluster 28 (1 to 7 repeats at 8 loci) suggests these are not likely to be true clusters. Isolates included in cluster 30, which demonstrated indistinguishable PFGE patterns, were considered to be highly related based on the observation that the MLVA profiles differed by a maximum of 1 repeat at 2 loci. Among indistinguishable PFGE patterns, very little MLVA variability was noted in cluster 31 (maximum of 1 repeat at 2 loci) with the exception of 1 isolate, which differed by 1 repeat at VNTR 34. Based on PulseNet Canada's interpretation guidelines for MLVA, isolates demonstrating variability at VNTR 34 are not considered related and therefore, this particular isolate was unlikely to be linked to cluster 9. Among the 2 PFGE variants in cluster 32, which were 91 percent similar, little variability was observed in the MLVA profiles (maximum of 1 repeat at 2 loci). However, 1 isolate demonstrated an MLVA profile that was highly variable, differing by 1 to 4 repeats at 6 loci, and therefore, was not considered to be related to the cluster. Following MLVA-enhanced analysis, an additional isolate with a highly similar PFGE pattern (> 91 percent similar) and MLVA profile that varied by only 1 repeat at 2 loci from the main cluster profile was identified and appeared to be related based on the low level of variability.

3.3.4 2011 Clusters

Figure 6 illustrates a MST of 11 clusters and 4 outbreaks selected for MLVA-enhanced retrospective analysis of 2011 isolates. Listed in Table 10 are the detailed findings of the 2011

Figure 6. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 108 *Escherichia coli* O157:H7 isolates (clinical and food) associated with 15 retrospective clusters that occurred in 2011 as reported by PulseNet Canada. Each node in the MST corresponds to a unique MLVA profile and is color coded based on the pulsed-field gel electrophoresis (PFGE) pattern. Node sizes are related to the number of isolates and the number indicated in the middle of each node corresponds to the arbitrarily assigned cluster code. The solid lines infer relatedness by stating the number of differing loci. Retrospectively related isolates not previously identified in original real-time findings are highlighted in yellow, and unrelated isolates demonstrating highly variable MLVA profiles are indicated by a dotted line outlining the node.

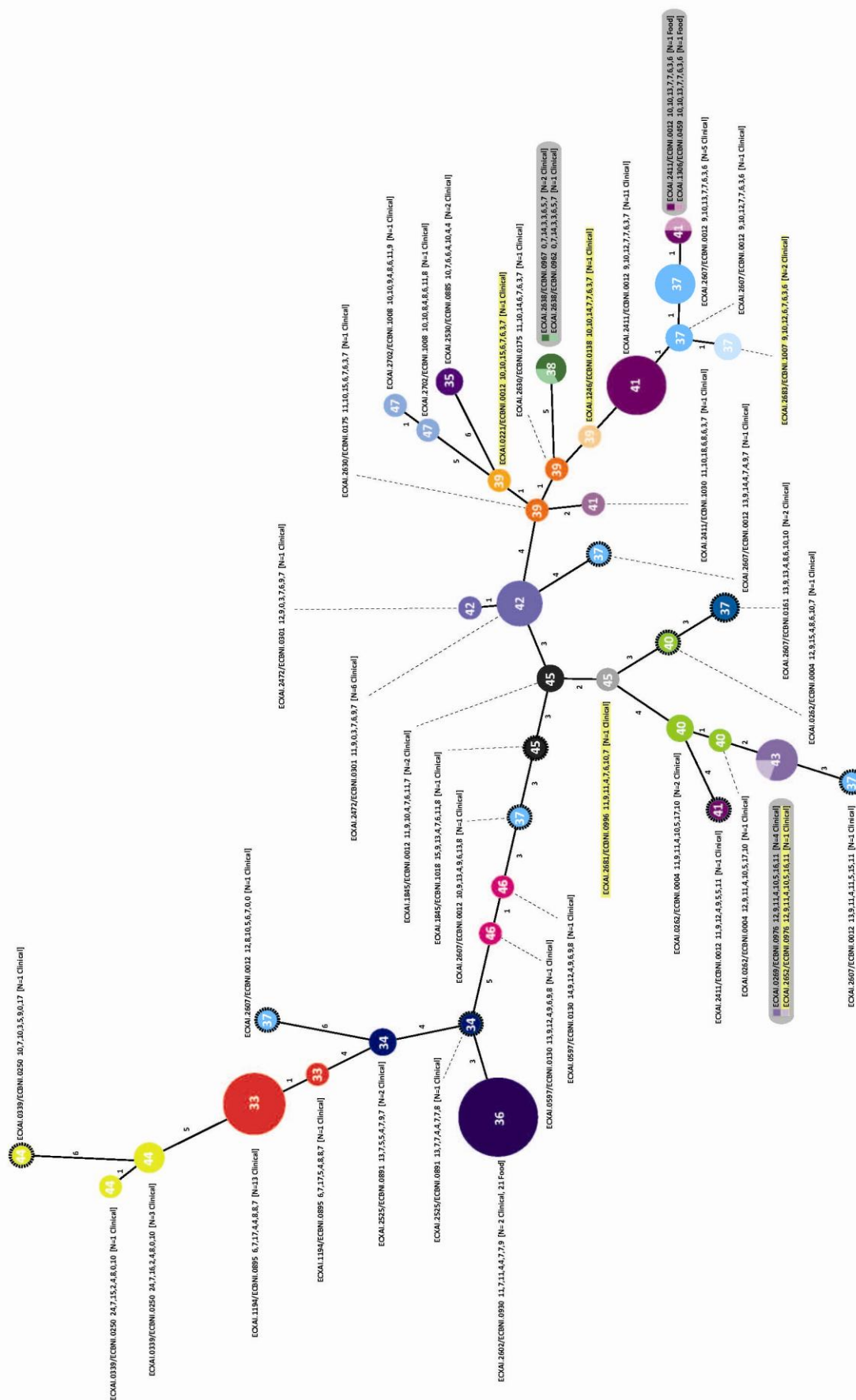


Table 10. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 15 *Escherichia coli* O157:H7 clusters that occurred in 2011 as reported by PulseNet Canada. Differences in repeat copy number in comparison to the main cluster profile are indicated in red.

Cluster (source)	Multi-Provincial	Isolates identified in original real-time findings		Isolates identified via MLVA-enhanced retrospective analysis			
		n	PFGE Pattern Combination	n	PFGE Pattern Combination	n	MLVA Profile Variability in MLVA
33 ^o (walnuts)	Yes	14	ECXAI.1194/ECBNI.0895	14	ECXAI.1194/ECBNI.0895	13	6,7,17,4,4,8,8,7 1 repeat at 1 locus
34 ^o (hazelnuts)	Yes	3	ECXAI.2525/ECBNI.0891	3	ECXAI.2525/ECBNI.0891	2	6,7,17,5,4,8,8,7 13,7,5,5,4,7,9,7 13,7,7,4,4,7,7,8 1-2 repeats at 4 loci*
35	Yes	2	ECXAI.2530/ECBNI.0885	2	ECXAI.2530/ECBNI.0885	2	10,7,6,6,4,10,4,4 No Variability
36 ^o (veal-liver)	No	24	ECXAI.2602/ECBNI.0930	23*	ECXAI.2602/ECBNI.0930	23	11,7,11,4,4,7,7,9 No Variability
37	Yes	12	ECXAI.2607/ECBNI.0012 ECXAI.2607/ECBNI.0161	11*	ECXAI.2607/ECBNI.0012	5	9,10,13,7,7,6,3,6 1 repeat at 1 locus
						2	9,10,12,7,7,6,3,6 1-12 repeats at 8 loci*
						1	13,9,11,4,11,5,15,11 1-6 repeats at 7 loci*
						1	13,9,14,4,7,4,9,7 1-6 repeats at 8 loci*
						1	12,8,10,5,6,7,0,0 1-10 repeats at 8 loci*
						1	10,9,13,4,9,6,13,8 1-7 repeats at 8 loci*
						2	13,9,13,4,8,6,10,10 1 repeat at 2 loci
						2	9,10,12,6,7,6,3,6 No Variability
38	Yes	2	ECXAI.2638/ECBNI.0967 ECXAI.2638/ECBNI.0962	2	ECXAI.2638/ECBNI.0967 ECXAI.2638/ECBNI.0962	2	0,7,14,3,3,6,5,7 1 repeat at 1 locus
						1	0,7,14,3,3,6,5,7 1 repeat at 3 loci
39	Yes	2	ECXAI.2630/ECBNI.0175	2	ECXAI.2630/ECBNI.0175	1	11,10,15,6,7,6,3,7 1 repeat at 1 locus
						1	11,10,14,6,7,6,3,7 1 repeat at 1 locus
						1	10,10,15,6,7,6,3,7 1 repeat at 3 loci
40	Yes	4	ECXAI.0262/ECBNI.0004	4	ECXAI.0262/ECBNI.0004	2	11,9,11,4,10,5,17,10 1 repeat at 1 locus
						1	12,9,11,4,10,5,17,10 1-7 repeats at 6 loci*
						1	12,9,15,4,8,6,10,7
41 ^o (beef products)	Yes	13	ECXAI.2411/ECBNI.0012 ECXAI.2411/ECBNI.1030 ECXAI.1306/ECBNI.0459	13	ECXAI.2411/ECBNI.0012	11	9,10,12,7,7,6,3,7 1-4 repeats at 7 loci*
						1	11,9,12,4,9,5,5,11 1 repeat at 3 loci
						1	10,10,13,7,7,6,3,6
						1	11,10,18,6,8,6,3,7
42	Yes	7	ECXAI.2472/ECBNI.0301	7	ECXAI.2472/ECBNI.0301	1	10,10,13,7,7,6,3,6
						1	11,9,0,3,7,6,9,7
43	Yes	4	ECXAI.0269/ECBNI.0976	4	ECXAI.0269/ECBNI.0976	6	12,9,0,3,7,6,9,7 1 repeat at 1 locus
						1	12,9,0,3,7,6,9,7
44	Yes	5	ECXAI.0339/ECBNI.0250	5	ECXAI.0339/ECBNI.0250	4	12,9,11,4,10,5,16,11 No Variability
						1	12,9,11,4,10,5,16,11
45	Yes	2	ECXAI.1845/ECBNI.0012 ECXAI.1845/ECBNI.1018	2	ECXAI.1845/ECBNI.0012	3	24,7,16,2,4,8,0,10 1 repeat at 1 locus
						1	24,7,15,2,4,8,0,10 1-14 repeats at 6 loci*
						1	10,7,10,3,5,9,0,17
						1	11,9,10,4,7,6,11,7
46	Yes	2	ECXAI.0597/ECBNI.0130	2	ECXAI.1845/ECBNI.1018	1	15,9,13,4,7,6,11,7 3-4 repeats at 2 loci*
						1	11,9,11,4,7,6,10,7 1 repeat at 2 loci
						1	14,9,12,4,9,6,9,8
47	Yes	2	ECXAI.2702/ECBNI.1008	2	ECXAI.2702/ECBNI.1008	1	13,9,12,4,9,6,9,8 1 repeat at 1 locus
						1	10,10,8,4,8,6,11,9 1 repeat at 1 locus

^o Epidemiologically linked to a common source and/or food recall.

• Not all isolates were available for further study.

† Retrospectively related isolate.

* Not likely to be related due to the presence of a highly variable MLVA profile.

retrospective analysis in comparison to those results obtained during the original real-time findings. In regards to cluster 33, which was linked to the consumption of raw in-shell walnuts and cluster 36, which was linked to the consumption of veal liver, no PFGE or MLVA variability was observed within each grouping except for 1 isolate in cluster 33, which varied by 1 repeat at 1 locus, indicating these isolates were indeed outbreak related. Of note, 1 isolate pertaining to cluster 36 was not available for retrospective analysis. Based on the original real-time findings, 3 isolates demonstrating indistinguishable PFGE patterns were included in cluster 34, which was associated with the consumption of in-shell raw hazelnuts. However, only 2 of these isolates clustered based on MLVA, and the remaining isolate varied by 1 to 2 repeats at 4 loci suggesting it was not related to the outbreak.

From epidemiological evidence and previously compiled laboratory data, cluster 41 comprised 3 variant PFGE patterns (> 96 percent similar) and was linked to the consumption of frozen beef burgers. For 2 of the PFGE pattern groupings, no MLVA variability was observed. Among the main PFGE pattern grouping, little MLVA variability was observed (1 repeat at 3 loci) with the exception of 1 isolate, which differed by 1 to 4 repeats at 7 loci, indicating it may not have been outbreak related. It is important to note that cluster 41 was epidemiologically linked to 2 additional clusters that occurred in 2012 (cluster 49 and cluster 50) based on the consumption of a common food source (frozen beef burgers). Of the remaining clusters investigated, no PFGE or MLVA variability was observed in clusters 35, 36, or 43. With respect to cluster 43, an additional isolate with a highly similar PFGE pattern (95 percent similar) and identical MLVA profile was identified and thus, considered to be retrospectively related. Two highly similar PFGE patterns (> 97 percent similar) were associated with cluster 37. Seven out

of the 13 isolates available for analysis (1 isolate could not be obtained) demonstrated highly comparable MLVA profiles (maximum of 1 repeat at 1 locus), while the remaining 6 isolates showed a high degree of variability (1 to 12 repeats at 7 to 8 loci) and were unlikely to be cluster related. An additional isolate with a 86 percent similar PFGE pattern combination and an MLVA profile varying by 1 repeat at 2 loci from the main cluster profile was retrospectively identified. Based on the low level of variability, this isolate was thought to be related to the cluster.

Isolates associated with cluster 39 demonstrated indistinguishable PFGE patterns and highly similar MLVA profiles (maximum of 1 repeat at 1 locus), and therefore, were considered to be related. Two additional isolates demonstrating comparable PFGE patterns (> 88 percent similar) and highly similar MLVA profiles to that of the main cluster profile were identified; one MLVA profile differed by 1 repeat at 1 locus and the other by 1 repeat at 3 loci. In retrospect, based on the low level of MLVA variability, these isolates appear to be related to the cluster. Although no PFGE variability was observed in cluster 40, only 3 of the 4 isolates demonstrated highly similar MLVA profiles (maximum of 1 repeat at 1 locus) indicating a high level of relatedness. Due to the high level of MLVA variability (1 to 7 repeats at 6 loci) observed for the remaining isolate, it was not likely to be related to the cluster. Interestingly, the MLVA profiles identified for cluster 40 and for cluster 43 were observed to be highly similar as they differed by only 1 repeat at 2 to 3 loci (Figure 7); this low level of variability suggests these clusters may have been related. No PFGE or MLVA variability was observed among the majority of isolates in cluster 42; even though 1 isolate differed by 1 repeat at 1 locus, the low level of MLVA variability suggests a high level of relatedness among the isolates included in the cluster.

Among isolates demonstrating indistinguishable PFGE patterns, a maximum of a 1 repeat difference at 1 locus was observed in the MLVA profiles associated with cluster 44, except for 1 isolate, which showed substantial variation (1 to 14 repeats at 6 loci), and therefore was unlikely to be cluster related. Within the main PFGE grouping (ECXAI.1845/ECBNI.0012) associated with cluster 45, no MLVA variability was observed. Despite only differing by a band shift in the *BlnI* restriction enzyme pattern, the PFGE variant demonstrated an MLVA profile that differed by 3 to 4 repeats at 2 loci from the main profile. This high level of variability suggests this particular isolate was not cluster related. An additional isolate identified via MLVA-enhanced retrospective analysis was presumed to be related to the cluster as the PFGE pattern was 88 percent similar and the MLVA profile was highly comparable (maximum of 1 repeat at 2 loci). Very little MLVA variability was observed in the MLVA profiles associated with the PFGE patterns included in clusters 46 and 47. Due to the low level of variability among isolates (1 repeat at 1 locus), these were likely to represent true clusters.

3.3.5 2012 Clusters

An MST of the MLVA profiles associated with 9 clusters and 7 outbreaks of *E. coli* O157:H7 that occurred in 2012 was constructed (Figure 7) and the main findings of the MLVA-enhanced retrospective analysis are detailed in Table 11. Of the 2 isolates available for analysis of cluster 48, which was linked to the consumption of ground beef, no PFGE or MLVA variability was observed indicating these isolates were highly related. Linked to the consumption of frozen beef burgers, isolates included in cluster 49 demonstrated indistinguishable PFGE

Figure 7. Minimum spanning tree (MST) of the multiple locus variable number tandem repeat analysis (MLVA) profiles for 183 *Escherichia coli* O157:H7 isolates (clinical and food) associated with 16 retrospective clusters that occurred in 2012 as reported by PulseNet Canada. Each node in the MST corresponds to a unique MLVA profile and is color coded based on the pulsed-field gel electrophoresis (PFGE) pattern. Node sizes are related to the number of isolates and the number indicated in the middle of each node corresponds to the arbitrarily assigned cluster code. The solid lines infer relatedness by stating the number of differing loci. Retrospectively related isolates not previously identified in original real-time findings are highlighted in yellow, and unrelated isolates demonstrating highly variable MLVA profiles are indicated by a dotted line outlining the node.

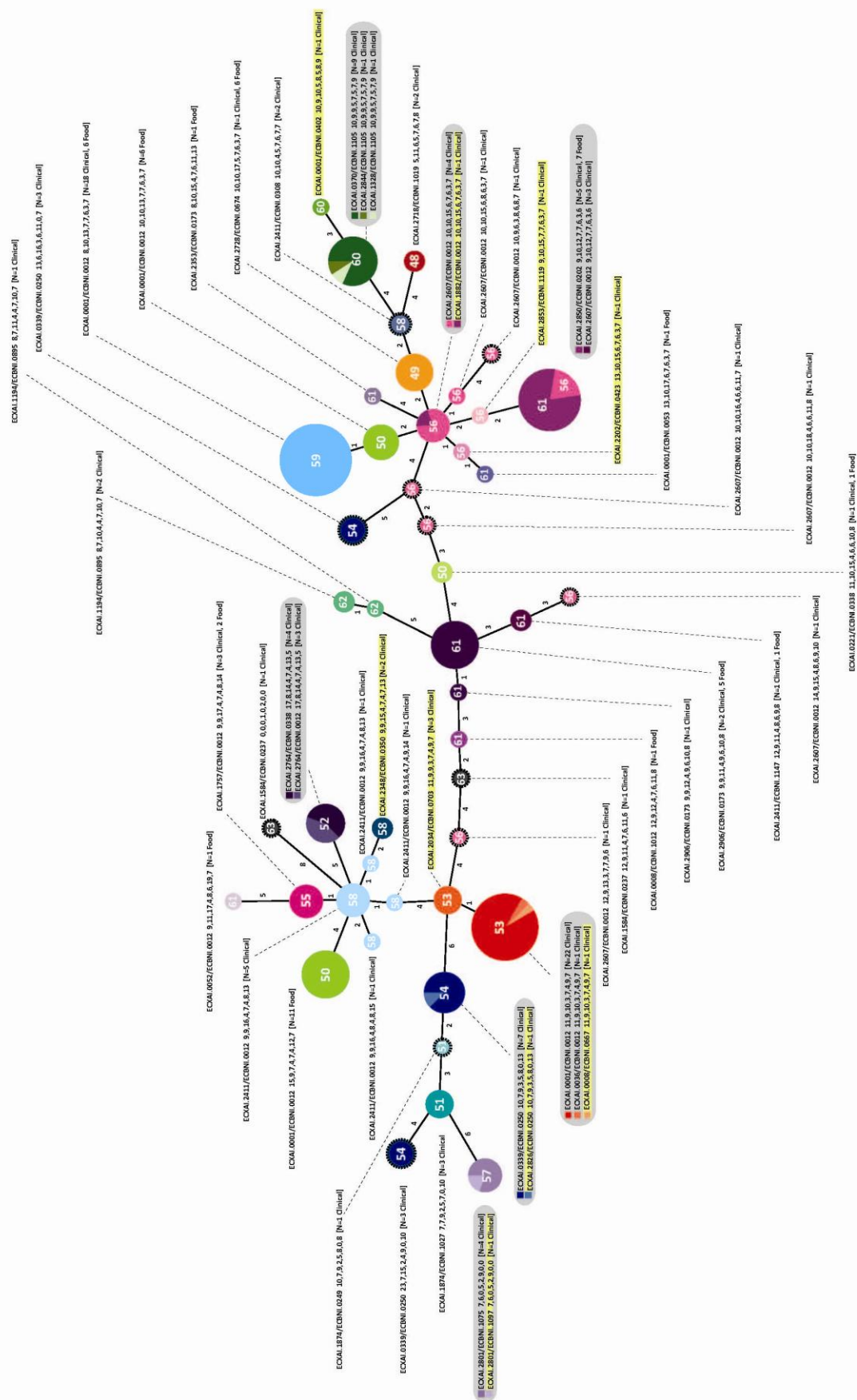


Table 11. Comparative analysis of isolates identified in original real-time findings with those identified via MLVA-enhanced retrospective analysis for 16 *Escherichia coli* O157:H7 clusters that occurred in 2012 as reported by PulseNet Canada. Differences in repeat copy number in comparison to the main cluster profile are indicated in red.

Cluster (source)	Multi-Provincial	Isolates identified in original real-time findings		Isolates identified via MLVA-enhanced retrospective analysis				
		n	PFGE Pattern Combination	n	PFGE Pattern Combination	n	MLVA Profile	Variability in MLVA
48 ^a (beef products)	No	3	ECXAI.2718/ECBNI.1019	2*	ECXAI.2718/ECBNI.1019	2	5,11,6,5,7,6,7,8	No Variability
49 ^a (beef products)	No	7	ECXAI.2728/ECBNI.0674	7	ECXAI.2728/ECBNI.0674	7	10,10,17,5,7,6,3,7	No Variability
50 ^a (beef products)	Yes	17	ECXAI.0001/ECBNI.0012	17	ECXAI.0001/ECBNI.0012	11	15,9,7,4,7,4,12,7	#
		2	ECXAI.0221/ECBNI.0338	2	ECXAI.0221/ECBNI.0338	6	10,10,13,7,7,6,3,7	
51	Yes	3	ECXAI.1874/ECBNI.1027	3	ECXAI.1874/ECBNI.1027	2	11,10,15,4,6,6,10,8	1-3 repeats at 3 loci
		1	ECXAI.1874/ECBNI.0249	1	ECXAI.1874/ECBNI.0249	3	7,7,9,2,5,7,0,10	
52	Yes	4	ECXAI.2764/ECBNI.0338	4	ECXAI.2764/ECBNI.0338	1	10,7,9,2,5,8,0,8	No Variability
		3	ECXAI.2764/ECBNI.0012	3	ECXAI.2764/ECBNI.0012	4	17,8,14,4,7,4,13,5	
53 ^a (lettuce)	Yes	22	ECXAI.0001/ECBNI.0012	22	ECXAI.0001/ECBNI.0012	3	17,8,14,4,7,4,13,5	1 repeat at 1 locus
		1	ECXAI.0036/ECBNI.0012	1	ECXAI.0036/ECBNI.0012	22	11,9,10,3,7,4,9,7	
		1	ECXAI.0008/ECBNI.0667	1	ECXAI.0008/ECBNI.0667†	1	11,9,10,3,7,4,9,7	
				3	ECXAI.2034/ECBNI.0703†	1	11,9,9,3,7,4,9,7	
54	Yes	14	ECXAI.0339/ECBNI.0250	13*	ECXAI.0339/ECBNI.0250	3	10,7,9,3,5,8,0,13	1-7 repeats at 6 loci* 1-13 repeats at 6 loci*
						3	13,6,16,3,6,11,0,7	
						3	23,7,15,2,4,9,0,10	
						1	10,7,9,3,5,8,0,13	
55 ^a (lettuce)	Yes	5	ECXAI.1757/ECBNI.0012	5	ECXAI.1757/ECBNI.0012	5	9,9,17,4,7,4,8,14	No Variability
56	Yes	14	ECXAI.2607/ECBNI.0012	13*	ECXAI.2607/ECBNI.0012	4	10,10,15,6,7,6,3,7	1-3 repeats at 4 loci* 1 repeat at 1 locus 1-8 repeats at 4 loci* 1-6 repeats at 7 loci* 1-6 repeats at 6 loci* 1-8 repeats at 5 loci* 1-9 repeats at 5 loci*
						3	9,10,12,7,7,6,3,6	
						1	10,10,15,6,8,6,3,7	
						1	10,10,16,4,6,6,11,7	
						1	12,9,13,3,7,7,9,6	
						1	14,9,15,4,8,6,9,10	
						1	10,10,18,4,6,6,11,8	
						1	10,9,6,3,8,6,8,7	
						1	10,10,15,6,7,6,3,7	
						1	13,10,15,6,7,6,3,7	
						1	9,10,15,7,7,6,3,7	
57	Yes	6	ECXAI.2801/ECBNI.1075	4*	ECXAI.2801/ECBNI.1075	4	7,6,0,5,2,9,0,0	No Variability
						1	7,6,0,5,2,9,0,0	
58	Yes	8	ECXAI.2411/ECBNI.0012 ECXAI.2411/ECBNI.0308	8	ECXAI.2411/ECBNI.0012	5	9,9,16,4,7,4,8,14	1 repeat at 2 loci 1 repeat at 1 locus 1 repeat at 1 locus 1-12 repeats at 7 loci* 1 repeat at 3 loci
						1	9,9,16,4,8,4,8,15	
						1	9,9,16,4,7,4,8,13	
						1	9,9,16,4,7,4,9,14	
						2	10,10,4,5,7,6,7,7	
						2	9,9,15,4,7,4,7,13	
59 ^a (beef products)	Yes	24	ECXAI.0001/ECBNI.0012	24	ECXAI.0001/ECBNI.0012	24	8,10,13,7,7,6,3,7	No Variability
60	Yes	9	ECXAI.0370/ECBNI.1105 ECXAI.2844/ECBNI.1105 ECXAI.1328/ECBNI.1105	9	ECXAI.0370/ECBNI.1105	9	10,9,9,5,7,5,7,9	1 repeat at 3 loci
						1	10,9,9,5,7,5,7,9	
						1	10,9,9,5,7,5,7,9	
						1	10,9,10,5,8,5,8,9	
61 ^a (beef products)	Yes	12	ECXAI.2850/ECBNI.0202 ECXAI.2906/ECBNI.0173 ECXAI.2411/ECBNI.1147 ECXAI.0008/ECBNI.1012 ECXAI.0052/ECBNI.0012 ECXAI.0001/ECBNI.0053 ECXAI.2353/ECBNI.0173	12	ECXAI.2850/ECBNI.0202	12	9,10,12,7,7,6,3,6	1 repeat at 1 locus
						7	9,9,11,4,9,6,10,8	
						1	9,9,12,4,9,6,10,8	
						2	12,9,11,4,8,6,9,8	
						1	12,9,12,4,7,6,11,8	
						1	9,11,17,4,8,6,19,7	
						1	13,10,17,6,7,6,3,7	
						1	8,10,15,4,7,6,11,13	
62	Yes	3	ECXAI.1194/ECBNI.0895	3	ECXAI.1194/ECBNI.0895	2	8,7,10,4,4,7,10,7	1 repeat at 1 locus
						1	8,7,11,4,4,7,10,7	
63	Yes	2	ECXAI.1584/ECBNI.0237	2	ECXAI.1584/ECBNI.0237	1	12,9,11,4,7,6,11,6	3-12 repeats at 8 loci*
						1	0,0,0,1,0,2,0,0	

^a Epidemiologically linked to a common source and/or food recall.

* Not all isolates were available for further study.

† Retrospectively related isolate.

Considered to be outbreak-related as it was isolated from recalled food products.

* Not likely to be related due to the presence of a highly variable MLVA profile.

patterns, 3 MLVA profiles differing by multiple repeats at multiple loci were observed. No variation was observed in the MLVA profiles for one of the PFGE variants, whereas a high level of variability was noted in the profiles associated with the other pattern variant (1 to 9 repeats at 6 loci). Despite the highly variable nature of the cluster associated MLVA profiles, all isolates were considered to be outbreak related based on evidence obtained from epidemiological summaries. No MLVA variability was observed for 3 of the isolates in cluster 51; however, for the PFGE variant ECXAI.1845/ECBNI.0249 (> 94 percent similar), a high degree of MLVA variability was observed (1 to 3 repeats at 3 loci), and therefore was not likely related.

All isolates associated with cluster 52 demonstrated a high level of relatedness as the PFGE patterns were 94 percent similar and no variability was observed in the MLVA profiles. As well, no MLVA variability was observed among the 2 PFGE variants (> 93 percent similar) included in cluster 53, which was linked to the consumption of lettuce. Three additional isolates demonstrating highly similar PFGE patterns (> 93 percent similar) and highly comparable MLVA profiles (maximum of 1 repeat at 1 locus) were retrospectively identified. These isolates were likely to be related due to the high degree of similarity and thus, may have been previously overlooked in original real-time findings. For cluster 54, only 13 of the 14 isolates were available for retrospective analysis. Despite the occurrence of a single PFGE pattern combination, 3 highly variable MLVA profiles were observed. Since 2 profiles differed by 1 to 13 repeats at 6 loci from the main MLVA profile, which was demonstrated by a majority of the isolates, they were not likely to be related. An additional isolate was identified with a highly similar restriction enzyme pattern (95 percent similar) and an MLVA profile indistinguishable from the main cluster profile. In retrospect, as indicated by the laboratory

evidence, this isolate may have been related to the cluster. Linked to the consumption of lettuce, no PFGE or MLVA variability was observed for cluster 55, and therefore, these isolates appear to be highly related. According to the original real-time findings, 14 isolates demonstrating a single PFGE pattern were included in cluster 56; however, only 13 of those were available for retrospective analysis. A total of 5 isolates were found to demonstrate comparable profiles (maximum of 1 repeat at 1 locus), and thus, were likely to be related. Due to the high level of MLVA variability observed in the remaining profiles (1 to 9 repeats at 4 to 7 loci), these isolates did not appear to be related to the main cluster profile. Three additional isolates demonstrating similar PFGE pattern combinations (> 74 percent similarity) and highly similar MLVA profiles differing by a minimum of 1 repeat at 2 loci and a maximum of 3 repeats at 1 locus were identified. Based on the low level of MLVA variability, these 3 isolates were presumed to be retrospectively related to the cluster.

Of the 6 isolates identified in original real-time findings for cluster 57, only 4 were available for this study. Since no PFGE or MLVA variability was observed among these isolates, they were regarded as highly likely to be related. An additional isolate demonstrating a highly comparable PFGE pattern (95 percent similar) and an indistinguishable MLVA profile was retrospectively identified as cluster related. Two highly similar PFGE variants (> 97 percent similar) were observed for cluster 58. Very little MLVA variability (a maximum difference of 1 repeat at 2 loci) was observed for one of the PFGE pattern variants, while conversely, a high level of MLVA variability (1 to 12 repeats at 7 loci) occurred for the other PFGE pattern combination. As a result of the high degree of variability, these isolates were unlikely to be related to the cluster. An additional isolate demonstrating a comparable PFGE pattern

combination (> 95 percent similar) and a highly similar MLVA profile was identified via MLVA-enhanced analysis. Since the MLVA profile varied by only 1 repeat at 3 loci from the main profile, this isolate appeared to be cluster related. For cluster 59, which was linked to the consumption of beef products, no PFGE or MLVA variability was observed, thereby indicating a high level of relatedness among these isolates. Despite the occurrence of 3 variant, yet highly similar PFGE patterns (> 95 percent similar), no MLVA variability was observed among cluster 60 isolates. Proceeding MLVA-enhanced analysis, an additional isolate demonstrating a comparable PFGE pattern (> 96 percent similar) and MLVA profile (maximum of 1 repeat at 3 loci) was identified. Due to the low level of MLVA variability, this isolate was considered to be retrospectively related to the cluster.

Cluster 61 was defined as a multi-strain outbreak and was linked to the consumption of frozen beef burgers. No MLVA variability was observed within each of the PFGE variants (> 88 percent similar) with the exception of 1 isolate, which differed by 1 repeat at 1 locus from the main profile for that particular pattern grouping. Due to this low level of MLVA variability and the fact that all isolates were either isolated from clinical specimens, recalled food products or from both, these isolates were all considered to be outbreak related. No PFGE or MLVA variability was observed for cluster 62 isolates with the exception of 1 isolate, which had a profile that differed by 1 repeat at 1 locus. Since the level of MLVA variability is extremely low, these isolates were indeed likely related. In accordance with the original real-time findings, 2 isolates were initially included in cluster 63. Although indistinguishable by PFGE, the MLVA data indicated a high level of variability among the isolates (3 to 12 repeats at 8 loci) and therefore,

this was not likely to be a true cluster. Furthermore, amplification was detected at only 2 loci for 1 of the cluster profiles suggesting this isolate was unlikely to be an *E. coli* O157:H7.

3.4 Variability in Related and Unrelated MLVA Profiles

Based on the retrospective analyses of the 63 clusters, it was determined that non-cluster/outbreak associated MLVA profiles were highly variable, ranging from 1-14 repeats at 2 to 8 loci, while cluster/outbreak related isolates demonstrated low levels of variability, differing by no more than 3 repeats at 1 locus or 1 repeat at 3 loci and no variation at VNTR 34.

3.5 Performance of MLVA as a Supplemental Method for Cluster and Outbreak Investigations

Based on the results presented in Table 12, the MLVA-enhanced retrospective analysis would have impacted the laboratory evidence for 38 of the 63 clusters analyzed. Twenty five clusters demonstrated no change in the number of isolates irrespective of the subtyping method used; MLVA-enhanced retrospective analysis excluded isolates formerly included in the original real-time findings for 18 clusters, and was able to identify additional isolates that may have been overlooked based on PFGE alone for 10 clusters. The use of MLVA as a supplemental subtyping method enabled both the exclusion of unrelated isolates with similar PFGE patterns as well as the inclusion of additional isolates demonstrating comparable PFGE patterns for 10 clusters. Although not all isolates could be obtained for 12 of the clusters in this study, the average number of isolates included in clusters based on original real-time findings was higher (8.98) in comparison to the average number of cluster-related isolates (7.23) based on MLVA-enhanced retrospective analysis. These data suggests the overall performance of MLVA (with PFGE) is superior to that of PFGE alone.

Table 12. Summary of the isolates included in 63 clusters of *Escherichia coli* O157:H7 based on the original real-time findings (PFGE alone) in comparison to those isolates identified as cluster related following MLVA-enhanced retrospective analysis (PFGE and MLVA).

Cluster	Isolates (n) included in the cluster based on original-real time findings	Isolates (n) included in the cluster based on MLVA-enhanced analysis	Adding MLVA resulted in: no change; excluding isolates that were likely not outbreak related; findings isolates that may have been overlooked originally, and both
1	4	4	no change
2	3	3	no change
3	2	6	added missed cases
4	11	13	added missed cases
5	4	4	no change
6	4	2	excluded unrelated isolates
7*	59	52	both
8*	32	31	no change
9*	7	7	added missed cases
10	13	14	added missed cases
11	17	4	excluded unrelated isolates
12	3	3	no change
13*	30	6	both
14*	9	3	excluded unrelated isolates
15	6	4	both
16	7	2	excluded unrelated isolates
17	2	2	no change
18	3	4	both
19	3	0*	excluded unrelated isolates
20	3	0*	excluded unrelated isolates
21	8	6	excluded unrelated isolates
22	4	6	added missed cases
23*	11	9	no change
24	3	3	no change
25	5	2	excluded unrelated isolates
26	2	0*	excluded unrelated isolates
27	2	0*	excluded unrelated isolates
28	2	0*	excluded unrelated isolates
29	10	10	no change
30	4	4	no change
31	5	4	excluded unrelated isolates
32*	6	4	both
33	14	14	no change
34	3	2	excluded unrelated isolates
35	2	2	no change
36*	24	23	no change
37*	14	9	both
38	3	3	no change
39	2	4	added missed cases
40	4	3	excluded unrelated isolates
41	15	14	excluded unrelated isolates
42	7	7	no change
43	4	5	added missed cases
44	5	4	excluded unrelated isolates
45	3	3	both
46	2	2	no change
47	2	2	no change
48*	3	2	no change
49	7	7	no change
50	19	19	no change
51	4	3	excluded unrelated isolates
52	7	7	no change
53	23	27	added missed cases
54*	14	8	both
55	5	5	no change
56*	14	8	both
57*	6	5	added missed cases
58	10	10	both
59	24	24	no change
60	11	12	added missed cases
61	26	26	no change
62	3	3	no change
63	2	0*	excluded unrelated isolates
Average	8.98	7.23	

• Not all isolates were available for further study.

* Not a true cluster.

Table 13. Comparative analysis of the sensitivity and specificity of MLVA when used as a supplemental method in combination with the current gold standard PFGE for *Escherichia coli* O157:H7 that occurred in Canada from 2008-2012. Table 13a depicts the comparison of PFGE + MLVA to PFGE data alone and Table 13b depicts the comparison of PFGE + MLVA with epidemiologically confirmed outbreak data + PFGE.

13a.

Subtyping Method Under Investigation (PFGE + MLVA)	Gold Standard (PFGE only)		
	Cluster Related	Not Related to the Cluster	
Cluster Related	360	34	Positive Predictive Value = 91%
Not Related to the Cluster	95	1087	Negative Predictive Value = 92%
	Sensitivity = 79%	Specificity = 97%	Accuracy = 92%

13b.

Subtyping Method Under Investigation (PFGE + MLVA)	Gold Standard (PFGE + Outbreak Epidemiological Data)		
	Outbreak Related	Not Related to the Outbreak	
Outbreak Related	181	7	Positive Predictive Value = 96%
Not Related to the Outbreak	10	788	Negative Predictive Value = 99%
	Sensitivity = 95%	Specificity = 99%	Accuracy = 98%

3.6 Sensitivity and Specificity of MLVA When Used in Combination with PFGE

When evaluating the utility of a clinical test, sensitivity and specificity are traditionally used and commonly defined as the ability of a test to correctly identify those individuals with disease and the ability of a test to correctly identify those individuals without disease, respectively (Lalkhen and McCluskey, 2008). However, in the context of cluster and outbreak investigations, sensitivity measures the ability of a typing method to correctly identify all cases that are truly cluster/outbreak related, while conversely, specificity relates to the ability of a typing method to correctly identify all cases that are truly unrelated to the cluster/outbreak, to rule them out. A total of 1,576 clinical isolates associated with the 63 clusters selected for retrospective analyses were used to calculate the sensitivity and specificity of MLVA when used in combination with PFGE, the current gold standard (Table 13a). Based on the results depicted in Table 11a, a total of 360 cases were identified as cluster-related by both PFGE and by PFGE + MLVA, indicating these cases are likely to be truly related to a cluster. As well, both typing methods agreed in ruling out 1087 cases, which suggest the cases are highly unlikely to be related to the clusters under study. Collectively, PFGE + MLVA ruled out 95 cases as cluster-related, whereas PFGE would have included them in the cluster.

In contrast, where PFGE + MLVA had identified 34 cases as cluster related, these cases were considered to be unrelated by PFGE. The sensitivity for PFGE + MLVA was calculated to be 79 percent, and the specificity was 97 percent compared to PFGE alone. The positive predictive value (PPV), which indicates the probability that a case will be truly cluster/outbreak related given the result obtained by the conjunctional use of PFGE + MLVA is positive (indicates relatedness among cases), was determined to be 91 percent. The negative predictive value

(NPV) indicates the probability of a case being truly unrelated to the cluster/outbreak given the result provided by PFGE + MLVA is negative (indicates the cases are unrelated) and was calculated to be 92 percent. The overall accuracy of the method, which can be described as the proportion of cases correctly classified by PFGE + MLVA out of the total number of cases included in the analysis, was calculated to be 92 percent.

Estimates of the ability of PFGE and MLVA combined to accurately categorize cases as “outbreak related” or “not outbreak related” was repeated for a subset of the clusters, representing the 16 well-characterized outbreaks (Table 13b). Both typing methods classified 181 cases as outbreak related and 788 cases and non-outbreak related, indicating these cases were most likely categorized correctly. A total of 10 cases were ruled out as outbreak related by PFGE + MLVA, but had been considered outbreak-related by PFGE; and vice versa, where PFGE + MLVA had included 7 cases that had been excluded by PFGE. The sensitivity and specificity of the 16 well-characterized outbreaks was 95 percent and 99 percent, respectively. The PPV (96 percent), NPV (99 percent), and accuracy (98 percent) were also observed to be greater than the statistical values acquired for the analysis of the 63 *E. coli* O157:H7 clusters. These results suggest that categorization of cases by PFGE + MLVA is superior to the categorization by PFGE alone and the greatest benefits of utilizing MLVA as a supplemental test may be realized during routine surveillance, when epidemiological information is not available.

Chapter 4: Discussion

4.1 Discriminatory Power of MLVA

An ideal subtyping method should provide a high level of discrimination among closely related isolates while detecting and quantifying the degree of relatedness among phylogenetically distant strains (Keys *et al.* 2005). The ability to rapidly and accurately differentiate among related and unrelated isolates, as well as to confirm the source(s) of transmission, are essential for laboratory surveillance and outbreak detection of *E. coli* O157:H7 and depend largely on the availability of reliable, high resolution molecular typing systems that provide enhanced discriminatory power. Discriminating power is a key characteristic of typing systems as it conditions the probability that isolates demonstrating identical or highly related types are truly clonal and linked to the same chain of transmission (Struelens, 1998). The SID was used to evaluate the overall discriminatory power of MLVA when used as a supplemental method in combination with PFGE and was calculated to be 0.998.

Ideally, the index should equal 1, however, a combination of typing systems displaying a discriminatory index greater than 0.95 is generally considered acceptable (Struelens, 1998). The discriminatory power of PFGE alone measured in this study was 0.993. Based on the statistical analysis, the conjunctional use of PFGE and MLVA provides increased discriminatory power for *E. coli* O157:H7 over PFGE alone with a less than a 1 percent probability of erroneously assigning independent isolates to the same clone. In order to effectively compare/combine studies performed using different subtyping methods, it is imperative to clarify the extension of agreement between the different clusterings in order to determine whether these techniques are identifying the same relationship among isolates (Severiano *et al.*, 2011).

4.2 Level of Congruency between PFGE and MLVA

In addition to discriminatory power, another indication of a subtyping method's performance is to estimate the congruency of its results with that of the gold standard. If the results are in complete agreement, it may be enough to collect information from a single method; while conversely, if the results disagree, combining the data obtained from the different subtyping methods may offer additional information and discriminatory power (Severiano *et al.*, 2011). Since the two typing systems are based on different approaches, perfect agreement among them was not expected. The statistical significance ($p < 0.001$) between the AW coefficients obtained for PFGE (0.227) and MLVA (0.696) indicated these techniques are unlikely to be congruent methods and therefore, combining their results may offer additional information and discriminatory power during analyses of clusters and outbreak investigations of *E. coli* O157:H7.

4.3 Impact of MLVA on Cluster Detection

The addition of MLVA changed the overall picture for 60 percent of the clusters analyzed in this study, which highlights the ability of this subtyping method to redefine the scope of a cluster/outbreak when used with traditional methodologies like PFGE. For approximately 16 percent of the clusters (10 out of the 63 clusters analyzed), adding MLVA resulted in the inclusion of isolates in the cluster that were previously missed by PFGE alone. A minimum of 1 additional isolate would have been ruled in for a majority of those clusters (cluster number 9, 10, 43, 57, and 60), and a maximum of 4 additional isolates would have been ruled in for two clusters (clusters 3, 53). For the remaining clusters (4, 22, 39), MLVA ruled in 2 additional isolates in each of those grouping that were likely overlooked by PFGE alone. The

PFGE pattern combinations of isolates that were ruled in by the addition of MLVA were similar to the main cluster patterns by a minimum of 74.5 percent and a maximum of 97.8 percent. This level of comparability would suggest that during cluster identification and outbreak investigations, PFGE patterns that are considered to be different from one another may in fact be related in some way.

Conversely, the use of MLVA also excluded isolates that were likely erroneously included by PFGE alone. A total of 29 percent of the clusters had at least 1 isolate identified as non-related to the cluster based on the MLVA-enhanced retrospective analysis. Based on a high level of MLVA variability observed at VNTR 34, an astonishing 13 isolates were excluded from cluster 11, all of which demonstrated an indistinguishable PFGE pattern. This finding highlights the importance of MLVA as a discriminating tool, as it is able to effectively discriminate between related and unrelated isolates that are otherwise indistinguishable by PFGE. Of note, since these isolates were all linked to the consumption of beef products, one could hypothesize the outbreak may have been caused by multiple strains, which would account for the variability in the MLVA profiles. Notwithstanding, the addition of MLVA also identified 6 clusters (19, 20, 26, 27, 28, and 63), which had been previously characterized by PFGE, that did not represent true clusters as the MLVA profiles were highly variable. MLVA effectively excluded unrelated isolates displaying indistinguishable/highly similar PFGE patterns, and included isolates with highly similar profiles for 10 out of the 63 clusters (clusters 7, 13, 15, 18, 32, 37, 45, 54, 56, and 58).

Although MLVA improved the overall picture for a large part of the clusters investigated, no change was observed in 25 out of the 63 clusters (approximately 40 percent). Since the

restriction enzyme patterns associated with these particular clusters were new/highly uncommon pattern combinations at the time of the investigation, finding any matches based on PFGE alone would have produced a strong enough signal that these isolates were likely related. In such situations, MLVA would be useful to confirm the results obtained by PFGE, but may not add further information or discriminatory power to the cluster investigation per se. Overall, the results acquired from the retrospective analyses illustrate the importance of MLVA as a supplemental typing system during cluster identification and outbreak investigations.

These findings have already impacted the operational procedures for laboratory surveillance in Canada. MLVA is now routinely used at PulseNet Canada as a supplemental subtyping tool in order to enhance surveillance and provide increased discriminatory power to outbreak detection and response. For example, the PFGE pattern combination ECXA1.0001/ECBNI.0012, which was linked to an outbreak associated with the consumption of beef products from XL Foods Inc., is the most frequently isolated restriction enzyme pattern for *E. coli* O157:H7 in Canada. The results provided by MLVA were integral in separating outbreak-related from temporally associated sporadic cases of *E. coli* O157:H7 during the investigation, as well as for confirming the source of contamination. Even though the addition of MLVA may not always change the number of isolates included in a cluster/outbreak, the subtyping tool has a significant impact on the accuracy of cluster detection and therefore, provides a clearer picture of the overall scope of the outbreak.

4.4 MLVA Offers Enhanced Specificity to Cluster Detection and Outbreak Investigations

The relationship between sensitivity and specificity exists in all laboratory tests. When considering the entire spectrum, with 100 percent sensitivity at one end, and 100 percent

specificity at the other, we need to situate the 'bar' somewhat in the middle in order to achieve a strong signal during cluster/outbreak investigations that is neither too sensitive nor too specific. For example, if we limit our focus to individuals who are more likely to have a common association, meaning they share an identical/highly similar fingerprint (PFGE pattern), we would be more likely to eliminate cases that are likely to be falsely associated than truly associated. However, if the PFGE pattern combination is relatively high in frequency, such as ECXAI.0001/ECBNI.0012, which is the most common restriction enzyme pattern for *E. coli* O157:H7 in Canada, temporally and geographically related sporadic cases often occur and therefore, based on the common DNA fingerprint, may be erroneously included in the investigation as cluster/outbreak related. While conversely, if we use criteria that are too stringent, essentially all non-related cases would be excluded, but however, a large proportion of truly associated cases may be excluded as well, thereby limiting the confidence in the results and substantially altering the scope of the outbreak.

Essentially, the main focus of the study was to calculate the performance of PFGE in relation to the epidemiological 'truth', and to calculate the performance of PFGE and MLVA collectively in relation to the 'truth' and then compare the results obtained from the two methods to determine whether the addition of MLVA changes the ability of researchers to classify isolates as cluster/outbreak related or non-related. However, since the epidemiological 'truth' was not always available, PFGE was used as the gold standard comparator. Hence, in operational reality, PFGE is the only 'gold standard' readily available as a reference when detecting clusters in real-time, so any retrospective truth will be biased to PFGE (obtaining epidemiological 'truth' that has been derived independently of PFGE or MLVA is not possible

retrospectively; nationally, epidemiological follow up for *E. coli* O157:H7 doesn't typically occur until after a case has been defined by PFGE). With epidemiological truth relying almost entirely on PFGE, comparing it in a 2 X 2 table would put PFGE at 100 percent agreement and PFGE + MLVA at some value less than that, which may not be informative.

In terms of the use of sensitivity and specificity, these metrics were used to describe how well PFGE + MLVA is able to sort cases into the correct category (or what is perceived to be correct) compared to PFGE alone. Table 13b, which contains the statistical results collected from the analysis of 16 well-characterized outbreaks, is a closer approximation of the epidemiological truth than Table 13a as it describes a subset of data comprising outbreak epidemiological categorization in the gold standard. Although the sensitivity and specificity of PFGE + MLVA for the 63 clusters shown in Table 13a was noticeably lower in comparison to PFGE alone, it was expected and when taken together with the rest of the MLVA analyses, reaffirms the contention that PFGE + MLVA is not only different from PFGE alone, but it is also better. While the actual 'truth' could only be approximated from readily available epidemiological outbreak data, the statistical results suggest that the categorization of cases by PFGE + MLVA is much better than the categorization by PFGE alone. When a closer approximation of epidemiological 'truth' is available (as displayed in Table 13b), differences between the data is still observed, however, these differences are likely attributable to improvements in case categorization when using PFGE + MLVA. Therefore, as the data set becomes more reflective of the epidemiological 'truth', both the sensitivity and specificity of PFGE + MLVA are expected to increase.

Based on the analyses, the use of MLVA as a secondary typing method in combination with PFGE would achieve the optimal balance required between sensitivity and specificity for the purposes of cluster/outbreak investigations of *E. coli* O157:H7 in Canada. Additionally, since the PPV and NPV obtained for both the analysis of clusters and outbreaks were relatively close to 1, it provides a strong indication that the performance of PFGE and MLVA is highly similar (if not superior) to PFGE alone. Although the sensitivity (79 percent) and specificity (97 percent) obtained for the analysis of clusters was less than the sensitivity (95 percent) and specificity (99 percent) obtained for the 16 outbreaks, these values still provide strong statistical support that MLVA (when used in combination with PFGE) offers enhanced specificity to cluster/outbreak investigations without significantly impacting the sensitivity. To our knowledge, this was a completely novel way to evaluate the performance of a subtyping method.

4.5 Conclusion

In conclusion, MLVA offers enhanced discriminatory power to cluster identification and outbreak investigations and since PFGE and MLVA are not likely to be congruent typing systems, combining their results offers additional information and discriminatory power. Based on the findings of the MLVA-enhanced retrospective analysis, MLVA is able to provide additional resolution over PFGE analysis and generally agrees with PFGE when isolates are identical and epidemiologically linked. Additionally, MLVA has the capacity to refine the scope of a cluster/outbreak as this typing system is able to identify isolates with variant PFGE pattern combinations demonstrating highly related profiles that may have been previously missed by PFGE alone. The MLVA assay shows great promise as a molecular epidemiological tool to

complement PFGE as it provides enhanced specificity, which improves case categorization during investigations of *E. coli* O157:H7 in Canada. As well, the work presented in the study will serve as a useful model system for evaluating the impact of a typing system on outbreak detection and response (i.e. whole genome sequencing).

References

- Ahmed, A., C. Bopp, A. Borczyk, and S. Kastiya.** 1987. Phage-typing scheme for *Escherichia coli* O157:H7. *J. Infect. Dis.* **155**:806-809.
- Barrett, T. J., H. Lior, J. H. Green, R. Khakhria, J. G. Wells, B. P. Bell, K. D. Greene, J. Lewis and P. M. Griffin.** 1994. Laboratory investigation of a multistate food-borne outbreak of *Escherichia coli* O157:H7 by using pulsed-field gel electrophoresis and phage typing. *J. Clin. Microbiol.* **32**:3013-3017.
- Barrett, T. J., P. Gerner-Smidt and B. Swaminathan.** 2006. Interpretation of pulsed-field gel electrophoresis patterns in foodborne disease investigations and surveillance. *Foodborne Pathog. Dis.* **3**:20-31.
- Belongia, E. A., M. T. Osterholm, J. T. Solder, D. A. Ammend, J. E. Braun, and K. L. MacDonald.** 1993. Transmission of *Escherichia coli* O157:H7 infection in Minnesota child day-care facilities. *J. Am. Med. Assoc.* **269**:883-888.
- Besser, J.** 2011. MLVA and the case definition. Proceedings from the 15th Annual PulseNet Update Meeting and 7th Annual OutbreakNet Conference, Long Beach, CA.
- Boerlin, P., S.A. McEwen, F. Boerlin-Petzold, J. B. Wilson, R. P. Johnson, and C. Gyles.** 1999. Associations between virulence factors of Shiga-toxin producing *Escherichia coli* and disease in humans. *J. Clin. Microbiol.* **37**:497-503.
- Bravaro, M. F.** 2012. *E. coli* O157:H7 and other toxigenic strains: the curse of global food distribution. *Curr. Gastroenterol. Rep.* **14**:317-323.
- Brooks, J. T., E. G. Sowers, J. G. Wells, K. D. Greene, P. M. Griffin, R. M. Hoekstra, and N. A. Strockbine.** 2005. Non-O157 Shiga toxin-producing *Escherichia coli* Infections in the United States, 1983-2002. *J. Infect. Dis.* **192**:1422-1429.
- Buchanan, R.L., and M. P. Doyle.** 1997. Foodborne disease significance of *Escherichia coli* O157:H7 and other enterohemorrhagic *E. coli*. *Food Technol.* **51**:69-76.
- Caprioli, A., S. Morabito, H. Brugere, and E. Oswald.** 2005. Enterohemorrhagic *Escherichia coli*: emerging issues on virulence and modes of transmission. *Vet. Res.* **36**:289-311.
- Carrico, J.A., C. Silva-Costa, J. Melo-Cristino, F. R. Pinto, H. de Lencastre, J. S. Almeida, and M. Ramirez.** 2006. Illustration of a common framework for relating multiple typing methods by application to macrolide-resistant *Streptococcus pyogenes*. *J. Clin. Microbiol.* **44**:2534-2532.

- Chapman, P. A.** 2000. Methods available for the detection of *Escherichia coli* O157:H7 in clinical, food and environmental samples. *World J. Microb. Biot.* **16**:733-740.
- Cizak, A., I. Literak, and P. Scheer.** 2000. Survival of *Escherichia coli* O157 in faeces of experimentally infected rats and domestic pigeons. *Lett. Appl. Microbiol.* **31**:349-352.
- Creel, J. S., J. R. Sharkey, A. McIntosh, J. Anding, and J. C. Huber Jr.** 2008. Availability of healthier options in traditional and non-traditional rural fast-food outlets. *BMC Public Health.* **8**:395.
- Deisingh, A. K., and M. Thompson.** 2004. Strategies for the detection of *Escherichia coli* O157:H7 in foods. *J. Appl. Microbiol.* **96**:419-429.
- Doyle, M. E., J. Archer, C. W. Kaspar, and R. Weiss.** 2006. Human illness caused by *E. coli* O157:H7 from food and non-food sources. *FRI Briefings.*
- Elder, R. O., J. E. Keen, G. R. Siragusa, G. A. Barkocy-Gallagher, M. Koohmaraie, and W. W. Laegreid.** 2000. Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. *PNAS* **97**:2999-3003.
- Erickson, M. C., and M. P. Doyle.** 2007. Food as a vehicle for transmission of Shiga toxin-producing *Escherichia coli*. *J. Food Protect.* **70**:2426-2449.
- Feng, P.** 1995. *Escherichia coli* serotype O157:H7: novel vehicles of infection and emergence of phenotypic variants. *Emerg. Infect. Dis.* **1**:47-52.
- Feng, P., R. C. Sandlin, C. H. Park, R. A. Wilson, and M. Nishibuchi.** 1998. Identification of a rough strain of *Escherichia coli* O157:H7 that produces no detectable O157 antigen. *J. Clin. Microbiol.* **36**:2239-2341.
- Fitzhenry, R. J., D. J. Pickard, E. L. Hartland, S. Reece, G. Dougan, A. D. Phillips, and G. Frankel.** 2002. Intimin type influences the site of human intestinal mucosal colonisation by enterohaemorrhagic *Escherichia coli* O157:H7. *Gut* **50**:180-185.
- Fraser, M. E., M. Fujinaga, M. M. Cherney, A. R. Melton-Celsa, E. M. Twiddy, A. D. O'Brien, and M. N. G. James.** 2004. Structure of Shiga toxin type 2 (Stx2) from *Escherichia coli* O157:H7. *J. Biol. Chem.* **279**:27511-27517.
- Galen R.S., and S. R. Gambino.** 1977. Beyond Normality: The Predictive Value and Efficiency of Medical Diagnoses. *Ann. Intern. Med.* **86**:513.

- Gansheroff, L. J., and A. D. O'Brien.** 2000. *Escherichia coli* O157:H7 in beef cattle presented for slaughter in the U.S.: higher prevalence rates than previously estimated. *Proc. Natl. Acad. Sci. USA* **97**:2959–2961.
- Garmendia, J., G. Frankel, and V. F. Crepin.** 2005. Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. *Infect. Immun.* **73**:2573-2585.
- Garmendia, J., A. D. Phillips, M. F. Carlier, Y. Chong, S. Schuller, O. Marches, S. Dahan, E. Oswald, R. K. Shaw, S. Knutton, and G. Frankel.** 2004. TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol.* **6**:1167-1183.
- Gerner-Smidt, P., J. Kincaid, K. Kubota, K. Hise, S. B. Hunter, M. Fair, D. Norton, A. Woo-Ming, T. Kurzynski, M. J. Sotir, M. Head, K. Holt and B. Swaminathan.** 2005. Molecular surveillance of Shiga toxigenic *Escherichia coli* O157 by PulseNet USA. *J. Food Protect.* **68**:1926-1931.
- Gerner-Smidt, P., K. Hise, J. Kincaid, S. Hunter, S. Rolando, E. Hyytiä-Trees, E. M. Ribot B. Swaminathan and the PulseNet Taskforce.** 2006. PulseNet USA: a five-year update. *Foodborne Pathog. Dis.* **3**:9-19.
- Goering, R. V.** 2010. Pulsed field gel electrophoresis: a review of application and interpretation in the molecular epidemiology of infectious disease. *Infect. Genet. Evol.* **10**:866-875.
- Goldwater, P. N., and K. A. Bettelheim.** 2012. Treatment of enterohemorrhagic *Escherichia coli* (EHEC) infection and hemolytic uremic syndrome (HUS). *BMC Med.* **10**:12
- Grundmann, H., S. Hori, and G. Tanner.** 2001. Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J. Clin. Microbiol.* **39**:4190-4192.
- Gu, J., Y. Liu, S. Yu, H. Wang, Y. Yi, F. Zhu, X. Yu, Q. Zou, and X. Mao.** 2009. Enterohemorrhagic *Escherichia coli* trivalent recombinant vaccine containing EspA, intimin and Stx2 induces strong humoral immune response and confers protection in mice. *Microbes Infect.* **11**:835-841.
- Haus-Cheymol, R., E. Espie, D. Che, V. Vaillant, H. DeValk, and J. C. Desenclos.** 2006. Association between indicators of cattle density and incidence of paediatric haemolytic-uraemic syndrome (HUS) in children under 15 years of age in France between 1996 and 2001: an ecological study. *Epidemiol. Infect.* **134**:712-718.

- Health Canada, Public Health Agency of Canada, and Canadian Food Inspection Agency.** 2011. Weight of evidence: factors to consider for appropriate and timely action in a foodborne illness outbreak investigation. Publically available at: http://www.hc-sc.gc.ca/fn-an/alt_formats/pdf/pubs/securit/2011-food-illness-outbreak-eclosion-malad-ailments-eng.pdf.
- Hrudley, S. E., P. Payment, P. M. Huck, R. W. Gillham, and E. J. Hruddy.** 2003. A fatal waterborne disease epidemic in Walkerton, Ontario: comparison with other waterborne outbreaks in the developed world. *Water Sci. Technol.* **47**:7-14.
- Hunter, P. R., and M. A. Gaston.** 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J. Clin. Microbiol.* **26**:2465-2466.
- Hyttia-Trees, E., S. C. Smole, P. A. Fields, B. Swaminathan and E. M. Ribot.** 2006. Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis for Shiga toxin-producing *Escherichia coli* O157:H7 (STEC O157). *Foodborne Pathog. Dis.* **3**:118-131.
- Hyttia-Trees, E.** 2009. Use of MLVA during outbreak investigations. Proceedings from the 15th Annual PulseNet Update Meeting, Snowbird, UT.
- Islam, M., M. P. Doyle, S. C. Phatak, P. Millner, and X. Jiang.** 2004. Persistence of enterohemorrhagic *Escherichia coli* O157:H7 in soil and on leaf lettuce and parsley grown in fields treated with contaminated manure composts or irrigation water. *J. Food Protect.* **67**:1365-1370.
- Jordan, K. N., and M. Maher.** 2006. Sensitive detection of *Escherichia coli* O157:H7 by conventional plating techniques. *J. Food Protect.* **69**:689-692.
- Karch, H., M. Bielaszewska, M. Bitzan, and H. Schmidt.** 1999. Epidemiology and diagnosis of Shiga toxin-producing *Escherichia coli* infections. *Diagn. Microbiol. Infect. Dis.* **34**:229-243.
- Keys, C., S. Kemper and P. Keim.** 2005. Highly diverse variable number tandem repeat loci in the *E. coli* O157:H7 and O55:H7 genomes for high-resolution molecular typing. *J. Appl. Microbiol.* **98**:928-940.
- Khakhria, R., D. Duck, and H. Lior.** 1990. Extended phage-typing scheme for *Escherichia coli* O157:H7. *Epidemiol. Infect.* **105**:511-520.
- 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.

- Lalkhen, A. G., and A. McCluskey.** 2008. Clinical tests: sensitivity and specificity. *Contin. Educ. Anaesth. Crit. Care Pain.* **8**:221-223
- Law, D.** 2000. Virulence factors of *Escherichia coli* O157 and other Shiga toxin-producing *E. coli*. *J. Appl. Microbiol.* **88**:729-745.
- Leclercq, A., B. Lambert, D. Pierard, and J. Mahillon.** 2001. Particular biochemical profiles for Enterohemorrhagic *Escherichia coli* O157:H7 isolates on the ID 32E system. *J. Clin. Microbiol.* **39**:1161-1164.
- Lim, J. Y., J. W. Yoon, and C. J. Hovde.** 2010. A brief overview of *Escherichia coli* O157:H7 and its plasmid O157. *J. Microbiol. Biotechnol.* **20**:1-10.
- Lindstedt, B., E. Heir, E. Gjernes, T. Vardund and G. Kapperud.** 2003. DNA fingerprinting of Shiga-toxin producing *Escherichia coli* O157 based on multiple-locus variable-number tandem-repeats analysis (MLVA). *Ann. Clin. Microbiol. Antimicrob.* **2**:12.
- Lynch, M. F., R. V. Tauxe, and C. W. Hedberg.** 2009. The growing burden of foodborne outbreaks due to contaminated fresh produce: risks and opportunities. *Epidemiol. Infect.* **137**:307-315.
- Majowicz, S. E., W. B. McNab, P. Sockett, S. Henson, K. Dore, V. L. Edge, M. C. Buffett, A. Fazil, S. Read, S. McEwen, D. Stacey, and J. B. Wilson.** 2006. Burden of gastroenteritis in a Canadian community. *J. Food Protect.* **69**:651-659.
- Maki, D. G.** 2009. Coming to grips with foodborne infection — peanut butter, peppers, and nationwide *Salmonella* outbreaks. *N. Engl. J. Med.* **360**:10.
- Manning, S. D., A. S. Motiwala, A. C. Springman, W. Qi, D. W. Lacher, L. M. Ouellette, J. M. Mladonicky, P. Somsel, J. T. Rudrik, S. E. Dietrich, W. Zhang, B. Swaminathan, D. Allard, and T. S. Whittam.** 2008. Variation in virulence among clades of *Escherichia coli* O157:H7 associated with disease outbreaks. *PNAS.* **105**:4868-4873.
- March, S. B., and S. Ratnam.** 1986. Sorbitol-MacConkey medium for detection of *Escherichia coli* O157:H7 associated with hemorrhagic colitis. *J. Clin. Microbiol.* **23**:869-872.
- Michino, H., K. Araki, S. Minami, S. Takaya, N. Sakai, M. Miyazaki, A. Ono, and H. Yanagawa.** 1999. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. *Am. J. Epidemiol.* **150**:787-796.
- Nadon, C., and J. Rumore.** 2013. PulseNet Canada use of MLVA for *E. coli* O157:H7 outbreak response. Presented at PulseNet Canada, Laboratory and Epidemiology experts meeting, Winnipeg, MB. Available upon request.

- Nadon C.A., E. Trees, L. K. Ng, E. Moller Nielsen, A. Reimer, N. Maxwell, K. A. Kubota, P. Gerner-Smidt P, and the MLVA Harmonization Working Group.** 2013. Development and application of MLVA methods as a tool for inter-laboratory surveillance. *Euro. Surveill.* **18**:20565.
- Naylor, S. W., J. C. Low, T. E. Besser, A. Mahajan, G. J. Gunn, M. C. Pearce, I. J. McKendrick, D. G. E. Smith, and D. L. Gally.** 2003. Lymphoid follicle-dense mucosa at the terminal rectum is the principal site of colonization of enterohemorrhagic *Escherichia coli* O157:H7 in the bovine host. *Infect. Immun.* **71**:1505-1512.
- Neil, K. P., G. Biggerstaff, J. K. MacDonald, E. Trees, C. Medus, K. A. Musser, S. G. Stroika, D. Zink, and M. J. Sotir.** 2011. A novel vehicle for transmission of *Escherichia coli* O157:H7 to humans: multistate outbreak of *E. coli* O157:H7 infections associated with consumption of ready-to-bake commercial prepackaged cookie dough – United States, 2009. *Clin. Infect. Dis.* **54**:511-518.
- Newell, D. G., M. Koopmans, L. Verhoef, E. Duizer, A. Aidara-Kane, H. Sprong, M. Opsteegh, M. Langelaar, J. Threfall, F. Scheutz, J. van der Giessen, and H. Kruse.** 2010. Food-borne diseases – the challenges of 20 years ago still persist while new ones continue to emerge. *Int. J. Food. Microbiol.* **139**:S3-S15.
- Noller, A. C., M. C. McEllistrem, A. G. F. Pacheco, D. J. Boxrud and L. H. Harrison.** 2003. Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic *Escherichia coli* O157:H7 isolates. *J. Clin. Microbiol.* **41**:5389-5397.
- 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.
- Noller, A. C., M. C. McEllistrem, K. A. Shutt and L. H. Harrison.** 2006. Locus-specific mutational events in a multilocus variable-number tandem repeat analysis of *Escherichia coli* O157:H7. *J. Clin. Microbiol.* **44**:374-377.
- Nyachuba, D.G.** 2010. Foodborne illness: is it on the rise. *Nutr. Rev.* **68**:257-296.
- Ochman, H., J. G. Lawrence, and E. A. Groisman.** 2000. Lateral gene transfer and the nature of bacterial innovation. *Nature.* **405**:299-304.
- Park, S., R. W. Worobo, and R. A. Durst.** 1999. *Escherichia coli* O157:H7 as an emerging foodborne pathogen: A literature review. *Crit. Rev. Food Sci.* **39**:481-502.

- Parma, Y. R., P. A. Chacana, A. Roge, A. Kahl, A. Cangelosi, P. Geoghegan, P. M. A. Lucchesi, and M. E. Fernandez-Miyakawa.** 2011. Antibodies anti-Shiga toxin 2 B subunit from chicken egg yolk: isolation, purification and neutralization efficacy. *Toxicon* **58**:380-388.
- Paton, A. W., and J. C. Paton.** 1998. Detection and characterization of Shiga toxigenic *Escherichia coli* by using multiplex PCR assays for stx1, stx2, eaeA, enterohemorrhagic *E. coli* hlyA, rfbO111, and rfbO157. *J. Clin. Microbiol.* **36**:598-602.
- Potter, A. A., S. Klashinsky, Y. Li, E. Frey, H. Townsend, D. Rogan, G. Erickson, S. Hinkley, T. Klopfenstein, R. A. Moxley, D. R. Smith, and B. B. Finlay.** (2004) decreased shedding of *Escherichia coli* O157:H7 by cattle following vaccination with type III secreted proteins. *Vaccine* **22**:362-369.
- Price, S. B., J. C. Wright, F. J. DeGraves, M. Castanie-Cornet, and J. W. Foster.** 2004. Acid resistance systems required for survival of *Escherichia coli* O157:H7 in the bovine gastrointestinal tract and in apple cider are different. *Appl. Environ. Microbiol.* **70**:4792-4799.
- Public Health Agency of Canada.** 2012. National Enteric Surveillance Program annual summary. Publically available at: <http://publications.gc.ca/site/eng/465160/publication.html>.
- Public Health Agency of Canada.** 2014. General Pulsed-field gel electrophoresis for *Escherichia coli*, *Salmonella* and *Shigella*. PulseNet Canada Internal SOP # PNC-PR-210. Publically available at: <http://www.pulsentinternational.org>.
- Public Health Agency of Canada.** 2013. Agarose gel electrophoresis. PulseNet Canada Internal SOP # E-TT-PR-012. Available upon request.
- Public Health Agency of Canada.** 2013. Vero cell assay for toxin detection. PulseNet Canada Internal SOP # E-TT-PR-004. Available upon request.
- Public Health Agency of Canada.** 2014. General PCR protocol for toxin detection. PulseNet Canada Internal SOP # E-TT-PR-011. Available upon request.
- Public Health Agency of Canada.** 2014. Multiple locus variable number tandem repeat analysis protocol for *E. coli* O157:H7, *S. Enteritidis*, and *S. Typhimurium*. PulseNet Canada Internal SOP # PNC-PR-404. Publically available at: <http://www.pulsentinternational.org>.
- Rahn, K., J. B. Wilson, K. A. McFadden, S. C. Read, A. G. Ellis, S. A. Renwick, R.C. Clarke and R. P. Johnson.** 1996. Comparison of Vero cell assay and PCR as indicators of the presence of verocytotoxigenic *Escherichia coli* in bovine and human fecal samples. *Appl. Environ. Microbiol.* **62**:4314-4317.

- Rangel, J. M., P. H. Sparling, C. Crowe, P. M. Griffin, and D. L. Swerdlow.** 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States 1982-2002. *Emerg. Infect. Dis.* **11**:604-609.
- Ribot, E. M., M. A. Fair, R. Gautom, D. N. Cameron, S. B. Hunter, B. Swaminathan and T. J. Barrett.** 2006. Standardization of pulsed-field gel electrophoresis protocols for the subtyping of *Escherichia coli* O157:H7, *Salmonella*, and *Shigella* for PulseNet. *Foodborne Pathog. Dis.* **3**:59-67.
- Rozand, C., and P. C. H. Feng.** 2009. Specificity analysis of a novel-phage-derived ligand in an enzyme-linked fluorescent assay for the detection of *Escherichia coli* O157:H7. *J. Food. Prot.* **72**:1078-1081.
- Sabat, A. J., A. Budmir, D. Nashev, R. Sa-Leao, J. M. van Dijk, F. Laurent, H. Grundmann, A. W. Friedrich, on behalf of the ESCMID Study Group of Epidemiological Markers (ESGEM).** 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. *Euro. Surveill.* **18**:pii=20380.
- Sanchez, S., R. Martinez, J. Rey, A. Garcia, J. Blanco, M. Blanco, J. E. Blanco, A. Mora, S. Herrera-Leon, A. Echeita, and J. M. Alonso.** 2010. Pheno-genotypic characterization of *Escherichia coli* O157:H7 isolates from domestic and wild ruminants. *Vet. Microbiol.* **142**:445-449.
- Scott, E.** 1996. A review of foodborne disease and other hygiene issues in the home. *J Appl Bacteriol.* **80**:5-9.
- Severiano, A., F. R. Pinto, M. Ramirez, and J. A. Carrico.** 2011. Adjusted Wallace coefficient as a measure of congruence between typing methods. *J. Clin. Microbiol.* **49**:3997-4000.
- Shelton, D. R., J. S. Karns, and C. H. Park.** 2008. A multiple protocol to improve diagnosis and isolation of Shiga toxin-producing *Escherichia coli* from human stool specimens. *Diagn. Microbiol. Infect. Dis.* **62**:7-10.
- Sheng, H., J. Y. Lim, H. J. Knecht, J. Li, and C. J. Hovde.** 2006. Role of *Escherichia coli* O157:H7 virulence factors in colonization at the bovine terminal rectal mucosa. *Infect. Immun.* **74**:4685-4693.
- Sowers, E. G., J. G. Wells, and N. A. Strockbine.** 1996. Evaluation of commercial latex reagents for identification of O157 and H7 antigens of *Escherichia coli*. *J. Clin. Microbiol.* **34**:1286-1289.

- Solomon, E. B., S. Yaron, and K. Matthews.** 2002. Transmission of *Escherichia coli* O157:H7 from contaminated manure and irrigation water to lettuce plant tissue and its subsequent internalization. **68**:397-400.
- Sproston, E. L. M. Macrae, I. D. Ogden, M. J. Wilson, and N. J. C. Strachan.** 2006. Slugs: potential novel vectors of *Escherichia coli* O157. *Appl. Environ. Microbiol.* **72**:144-149.
- Struelens, M. J.** 1998. Molecular epidemiologic typing systems of bacterial pathogens: current issues and perspectives. *Mem. Inst. Oswaldo Cruz.* **93**:581-585.
- Su, C, and L. J. Brandt.** 1995. *Escherichia coli* O157:H7 infection in humans. *Ann. Intern. Med.* **123**:698-714.
- Swaminathan, B., P. Gerner-Smidt, L. NG, S. Lukinmaa, K. Kam, S. Rolando, E. P. Gutiérrez and N. Binsztein.** 2006. Building PulseNet International: An interconnected system of laboratory networks to facilitate timely public health recognition and response to foodborne disease outbreaks and emerging foodborne diseases. *Foodborne Pathog. Dis.* **3**:36-50.
- Talley, J.L., A. C. Wayadande, L. P. Wasala, A. C. Gerry, J. Fletcher, U. DeSilva, and S. E. Gilliland.** 2009. Association of *Escherichia coli* O157:H7 with filth flies (muscidae and calliphoridae) captured in leafy greens fields and experimental transmission of *E. coli* O157:H7 to spinach leaves by house flies (diptera: muscidae). *J. Food Protect.* **72**:1547-1552.
- Tauxe, R. V.** 2002. Emerging foodborne pathogens. *Int. J. Food. Microbiol.* **78**:31-41.
- Tauxe, R. V.** 2006. Molecular subtyping and the transformation of public health. *Foodborne Pathog. Dis.* **3**:4-8.
- Thomas, M. K., S. Majowicz, P. N. Sockett, A. Fazil, F. Pollari, K. Dore, J. A. Flint, and V. L. Edge.** 2006. Estimated numbers of community cases of illness due to *Salmonella*, *Campylobacter*, and verotoxigenic *Escherichia coli*: pathogen-specific community rates. *Can. J. Infect. Dis. Med. Microbiol.* **17**:229-234.
- Thomas, M. K., S. E. Majowicz, F. Pollari, and P. N. Sockett.** 2008. Burden of acute gastrointestinal illness in Canada, 1999-2007: interim summary of NSAGI activities. *Can. Commun. Dis. Rep.* **34**:8-15.
- Thomas, M. K., R. Murray, L. Flockhart, K. Pintar, F. Pollari, A. Fazil, A. Nesbitt, and B. Marshall.** 2013. Estimates of the burden of foodborne illness in Canada for 30 specified pathogens and unspecified agents, circa 2006. *Foodborne Pathog. Dis.* **10**:639-648.

- Tree, J. J., E. B. Wolfson, D. Wang, A. J. Roe, and D. L. Gally.** 2009. Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. *Trends Microbiol.* **17**:361-370.
- Valcour, J. E., P. Michel, S. A. McEwen, and J. B. Wilson.** 2002. Associations between indicators of livestock farming intensity and incidence of human Shiga toxin-producing *Escherichia coli* infection. *Emerg. Infect. Dis.* **8**:252-257.
- van Belkum, A.** 2007. Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). *FEMS Immunol. Med. Microbiol.* **49**:22-27.
- van Belkum, A., P. T. Tassios, L. Dijkshoorn, S. Haeggman, B. Cookson, N. K. Fry, V. Fussing, J. Green, E. Feil, P. Gerner-Smidt, S. Brisse, and M. Struelens.** 2007. Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect. Dis.* **13**:1-46.
- Vernozy-Rozand, C.** 1997. Detection of *Escherichia coli* O157:H7 and other verocytotoxin-producing *E. coli* (VTEC) in food. *J. Appl. Microbiol.* **82**:537-551.
- Vogler, A. J., C. Keys, Y. Nemoto, R. E. Colman, Z. Jay and P. Keim.** 2006. Effect of repeat copy number on variable-number tandem repeat mutations in *Escherichia coli* O157:H7. *J. Bacteriol.* **188**:4253-4263.
- Wendel, A. M., D. H. Johnson, U. Sharapov, J. Grant, J. R. Archer, T. Monson, C. Koschmann, and J. P. Davis.** 2009. Multistate outbreak of *Escherichia coli* O157:H7 infection associated with consumption of packaged spinach, August-September 2006: the Wisconsin investigation. *Clin. Infect. Dis.* **48**:1079-1086.
- Williams, A. P., L. M. Avery, K. Killham, and D. L. Jones.** 2005. Persistence of *Escherichia coli* O157 on farm surfaces under different environmental conditions. *J. Appl. Microbiol.* **98**:1075-1083.
- Wong, C. S., S. Jelacic, R. L. Habeeb, S. L. Watkins, and P. I. Tarr.** 2000. The risk of the hemolytic-uremic syndrome after antibiotic treatment of *Escherichia coli* O157:H7 infections. *New Eng. J. Med.* **342**:1930-1936.
- World Health Organization.** 2008. Foodborne disease outbreaks: guidelines for investigation and control. p 1-52.
- Yang, J. R., F. T. Wu, J. L. Tsai, J. J. Mu, L. F. Lin, K. L. Chen, S. H. Kuo, C. S. Chiang, and H. S. Wu.** 2007. Comparison between O serotyping method and multiplex real-time PCR to identify diarrheagenic *Escherichia coli* in Taiwan. *J. Clin. Microbiol.* **45**:3620-3625.

Yetka, M. A., B. M. Goddeeris, D. Vanrompay, and E. Cox. 2011. Immunization of sheep with a combination of intimin, EspA and EspB decreases *Escherichia coli* O157:H7 shedding. *Vet. Immunol. Immunop.* **140**:42-46.